



UNIVERSITÀ  
DI PAVIA

UNIVERSITÀ DEGLI STUDI DI PAVIA  
DIPARTIMENTO DI STUDI UMANISTICI  
SCUOLA DI DOTTORATO IN SCIENZE LINGUISTICHE  
XXXIII CICLO

**The semantics of *sustainable development*:  
A corpus-assisted, ecological analysis of  
discourse across languages**

Tutor:  
Chiar.mo Prof. Gianguido Manzelli

Tesi di Dottorato di  
Elena Valvason

Anno Accademico 2019/2020



## Acknowledgements

The present work is the result of a three-years long journey, which has involved a network of people whose contribute to the growth of this research has been invaluable.

First and foremost, I would like to thank my supervisor, Gianguido Manzelli, who has taught me what it means to be a passionate linguist and who has always encouraged me with kind and wise guidance. I would also like to thank my co-advisors: Maria Freddi, for introducing me to the field of corpus linguistics and for constantly and keenly advising me during the years of my PhD course, and Nicoletta Puddu, for helping me frame this research in a clear-cut linguistic dimension.

I would like to thank the coordinators of the PhD program in Linguistics, Silvia Luraghi and Sonia Cristofaro, for their great assistance during the years of my PhD course, and the Academic staff of the PhD program, among which in particular Maria Pavesi and Elisa Roma, who have always been extremely supportive of my work and of my teaching experiences.

My deepest gratitude goes also to my reviewers: Paolo Driussi, for his everlasting support, for his enduring patience, for his enthusiasm when showing me the beauty of research, and Vito Pirrelli, whose kind, precise and insightful recommendations have been of paramount importance for the improvement of this work and for stimulating me to expand my research horizons.

I would like to heartily thank Arran Stibbe for his warmly welcoming me to the University of Gloucestershire and for his precious teachings and inestimable help with exploring the vast field of ecolinguistics. Our shared interest and our discussions on the linguistic shape of the United Nations' *2030 Agenda for Sustainable Development* have been essential for the development of this research.

I am also grateful to the lecturers and research fellows of the University of Pavia, among which Chiara Meluzzi, Erica Pinelli and Chiara Zanchi, for their confirming the importance and value of team-work, and to my colleagues at the PhD program, who have shared with me the steps of this journey.



# Table of Contents

List of Tables .....	ix
List of Figures.....	xiii
List of Graphs.....	xvii
Introduction .....	1
1. Corpus and Discourse approaches to ecology across languages.....	5
1.1. Cultural keywords.....	5
1.2. Meaning by collocation .....	9
1.3. Corpus approaches to ecological discourse studies across languages .....	26
1.3.1. Critical discourse studies .....	26
1.3.2. Ecological discourse analysis .....	31
1.3.3. Corpus linguistics.....	39
1.3.4. Corpus approaches to discourse.....	56
2. Data.....	75
2.1. The <i>2030 Agenda Corpus</i> .....	75
2.1.1. Corpus collection and annotation.....	75
2.1.2. Corpus features .....	76
2.2. The <i>Sustainable development Corpus</i> .....	79
2.2.1. Corpus design.....	80
2.2.2. Corpus collection and annotation.....	83
2.2.3. Corpus features .....	90
3. Methodology .....	97
3.1. The research questions .....	97
3.2. The identification of cultural keywords .....	99
3.2.1. The most frequent lexemes .....	100
3.2.2. The statistically significant lexemes.....	100
3.2.3. Economic, social and environmental lexemes.....	106
3.3. The reading of concordance lines.....	107
3.4. The analysis of meaning by collocation .....	108
3.4.1. Word sketches.....	109
3.4.2. Collocation networks.....	111
3.5. The cross-linguistic interpretation of the results .....	115
3.5.1. English .....	116
3.5.2. Hungarian.....	117

3.5.3. Italian .....	118
3.6. The ecological interpretation of the results .....	118
3.6.1. Sustainable development .....	119
3.6.2. Social ecology.....	121
3.6.3. The ecological framework of this study .....	123
3.7. The structure of the study .....	124
4. Analysis of the <i>2030 Agenda Corpus</i> .....	127
4.1. Cultural keywords of the <i>2030 Agenda Corpus</i> .....	127
4.1.1. The most frequent lexemes .....	127
4.1.2. The statistically significant lexemes.....	141
4.1.3. Economic, social and environmental lexemes.....	148
4.2. Meaning by collocation of <i>sustainable</i> in the <i>2030 Agenda Corpus</i> .....	156
4.2.1. The English <i>SUSTAINABLE</i> .....	159
4.2.2. The Hungarian <i>FENNTARTHATÓ</i> .....	182
4.2.3. The Italian <i>SOSTENIBILE</i> .....	193
4.3. Meaning by collocation of <i>sustainability</i> in the <i>2030 Agenda Corpus</i> .....	204
4.3.1. The English <i>SUSTAINABILITY</i> .....	205
4.3.2. The Hungarian <i>FENNTARTHATÓSÁG</i> .....	206
4.3.3. The Italian <i>SOSTENIBILITÀ</i> .....	207
4.4. Meaning by collocation of <i>sustainable development</i> in the <i>2030 Agenda Corpus</i> .....	208
4.4.1. The English <i>SUSTAINABLE DEVELOPMENT</i> .....	209
4.4.2. The Hungarian <i>FENNTARTHATÓ FEJLŐDÉS</i> .....	232
4.4.3. The Italian <i>SVILUPPO SOSTENIBILE</i> .....	245
4.5. Discussion .....	256
5. Analysis of the <i>Sustainable development Corpus</i> .....	259
5.1. Cultural keywords of the <i>SusCorp</i> .....	259
5.1.1. The most frequent lexemes .....	259
5.1.2. The statistically significant lexemes.....	271
5.2. Meaning by collocation of <i>sustainable</i> in the <i>SusCorp</i> .....	281
5.2.1. The English <i>SUSTAINABLE</i> .....	283
5.2.2. The Hungarian <i>FENNTARTHATÓ</i> .....	300
5.2.3. The Italian <i>SOSTENIBILE</i> .....	310
5.3. Discussion .....	321
6. Conclusion.....	325
6.1. Summary of the results.....	325
6.1.1. The <i>2030 Agenda Corpus</i> .....	325

6.1.2. The <i>Sustainable development Corpus</i> .....	329
6.1.3. The <i>2030 Agenda Corpus</i> vs. the <i>Sustainable development Corpus</i> .....	331
6.2. An ecological interpretation of the results.....	333
6.3. Discussion .....	336
6.4. Future developments.....	338
References.....	341
Sitography .....	369
Appendix.....	371
A1. Outline of the statistics for the computation of keywords and collocations...	371
A2. Analysis of the <i>2030 Agenda Corpus</i> .....	376
A3. Analysis of the <i>Sustainable development Corpus</i> .....	402



## List of Tables

<b>Table 1.</b> Number of word types, number of tokens, and MATTR of the English, Hungarian, and Italian versions of the 2030 Agenda.....	77
<b>Table 2.</b> Labels for the texts in the English, Hungarian, and Italian subcorpora. ....	87
<b>Table 3.</b> Number of word types, number of tokens, MATTR, average sentence length, and average text length of the English, Hungarian, and Italian sections of SusCorp.	91
<b>Table 4.</b> Number of texts, number of word types, and number of tokens of the English, Hungarian, and Italian sections of the EUROPARL7 corpus.....	102
<b>Table 5.</b> Number of texts, number of word types, and number of tokens of the English, Hungarian and Italian sections of the TenTen corpus.....	104
<b>Table 6.</b> Number of texts, number of word types, and number of tokens of the English, Hungarian and Italian sections of the Timestamped JSI webcorpus 2014-2020.....	105
<b>Table 7.</b> The seventeen Sustainable Development Goals of the United Nations' 2030 Agenda for Sustainable Development. ....	121
<b>Table 8.</b> The twenty most frequent content words of the 2030 Agenda Corpus (English). ....	129
<b>Table 9.</b> Noun and adjective phrases modifying development in the 2030 Agenda Corpus (English). ....	130
<b>Table 10.</b> Noun phrases modified by development in the 2030 Agenda Corpus (English). ....	131
<b>Table 11.</b> Noun phrases modified by economic in the 2030 Agenda Corpus (English). ....	134
<b>Table 12.</b> The twenty most frequent content words of the 2030 Agenda Corpus (Hungarian).....	136
<b>Table 13.</b> The twenty most frequent content words of the 2030 Agenda Corpus (Italian). ....	138
<b>Table 14.</b> The twenty most significant keywords of the 2030 Agenda Corpus (English). ....	142
<b>Table 15.</b> Noun phrases modified by inclusive in the 2030 Agenda Corpus (English). ....	143
<b>Table 16.</b> Noun and adjective phrases modifying technology in the 2030 Agenda Corpus (English). ....	144
<b>Table 17.</b> Noun phrases modified by technology (2030 Agenda Corpus-English). .	144
<b>Table 18.</b> The twenty most significant keywords of the 2030 Agenda Corpus (Hungarian).....	145

<b>Table 19.</b> The twenty most significant keywords of the 2030 Agenda Corpus (Italian). .....	147
<b>Table 20.</b> List of human beings of the 2030 Agenda Corpus (English). .....	152
<b>Table 21.</b> List of animals of the 2030 Agenda Corpus (English).....	153
<b>Table 22.</b> List of plants of the 2030 Agenda Corpus (English). .....	154
<b>Table 23.</b> List of natural elements of the 2030 Agenda Corpus (English). .....	154
<b>Table 24.</b> List of collocates of SUSTAINABLE in the 2030 Agenda Corpus (English)...	162
<b>Table 25.</b> List of collocates of FENNTARTHATÓ ‘sustainable’ in the 2030 Agenda Corpus (Hungarian).....	185
<b>Table 26.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the 2030 Agenda Corpus (Italian). .....	196
<b>Table 27.</b> List of collocates of SUSTAINABLE DEVELOPMENT in the 2030 Agenda Corpus (English). .....	211
<b>Table 28.</b> List of collocates of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian).....	234
<b>Table 29.</b> List of collocates of SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda Corpus (Italian).....	246
<b>Table 30.</b> The twenty most frequent content words of the SusCorp (English). .....	261
<b>Table 31.</b> Nouns and adjectives modifying development in the SusCorp (English).262	
<b>Table 32.</b> Nouns modified by development in the SusCorp (English). .....	263
<b>Table 33.</b> Nouns and noun phrases modified by global in the SusCorp (English). .	264
<b>Table 34.</b> The twenty most frequent content words of the SusCorp (Hungarian)..	267
<b>Table 35.</b> The twenty most frequent content words of the SusCorp (Italian).....	269
<b>Table 36.</b> The twenty most significant statistical keywords of the SusCorp (English) compared to the enTenTen 2012. ....	273
<b>Table 37.</b> The twenty most significant statistical keywords of the SusCorp (Hungarian) compared to huTenTen 2012.....	275
<b>Table 38.</b> The twenty most significant statistical keywords of the SusCorp (Italian) compared to itTenTen 2016. ....	279
<b>Table 39.</b> List of collocates of SUSTAINABLE in the SusCorp (English). .....	286
<b>Table 40.</b> List of collocates of FENNTARHATÓ ‘sustainable’ in the SusCorp (Hungarian).....	303
<b>Table 41.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – Z(10.0), 1L-1R/3L-3R/5L-5R, NC20-C20.....	313
<b>Table 42.</b> The ways of referring to the UN’s 2030 Agenda in the SusCorp (Italian).317	
<b>Table 43.</b> The ways of referring to the UN’s Sustainable Development Goals in the SusCorp (Italian).....	317
<b>Table 44.</b> Positive keywords from the 2030 Agenda Corpus (English).....	382
<b>Table 45.</b> Positive keywords of the 2030 Agenda (English, Hungarian, and Italian). .....	389

<b>Table 46.</b> List of collocates of SUSTAINABLE in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.....	389
<b>Table 47.</b> List of collocates of SUSTAINABLE in the 2030 Agenda Corpus (English) – Z(10.0), 3L-3R, NC5-C5.....	390
<b>Table 48.</b> List of collocates of SUSTAINABLE in the 2030 Agenda Corpus (English) – Z(10.0), 5L-5R, NC5-C5.....	391
<b>Table 49.</b> List of collocates of FENNTARTHATÓ ‘sustainable’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.....	391
<b>Table 50.</b> List of collocates of FENNTARTHATÓ ‘sustainable’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.....	392
<b>Table 51.</b> List of collocates of FENNTARTHATÓ ‘sustainable’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 5L-5R, NC5-C5.....	394
<b>Table 52.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.....	394
<b>Table 53.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 3L-3R, NC5-C5.....	395
<b>Table 54.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 5L-5R, NC5-C5.....	397
<b>Table 55.</b> List of collocates of SUSTAINABLE DEVELOPMENT in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.....	397
<b>Table 56.</b> List of collocates of SUSTAINABLE DEVELOPMENT in the 2030 Agenda Corpus (English) Z(10.0), 3L-3R, NC5-C5.....	398
<b>Table 57.</b> List of collocates of SUSTAINABLE DEVELOPMENT in the 2030 Agenda Corpus (English) – Z(10.0), 5L-5R, NC5-C5.....	398
<b>Table 58.</b> List of collocates of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.....	399
<b>Table 59.</b> List of collocates of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.....	399
<b>Table 60.</b> List of collocates of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 5L-5R, NC5-C5.....	400
<b>Table 61.</b> List of collocates of SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.....	401
<b>Table 62.</b> List of collocates of SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 3L-3R, NC5-C5.....	401
<b>Table 63.</b> List of collocates of SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 5L-5R, NC5-C5.....	402
<b>Table 64.</b> List of the collocates of SUSTAINABLE in the SusCorp (English) – Z(10.0), 1L-1R, NC20-C20.....	403
<b>Table 65.</b> List of the collocates of SUSTAINABLE in the SusCorp (English) – Z(10.0), 3L-3R, NC20-C20.....	403

<b>Table 66.</b> List of collocates of SUSTAINABLE in the SusCorp (English) – Z(10.0), 5L-5R, NC20-C20.....	404
<b>Table 67.</b> List of collocates of FENNTARHATÓ ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 1L-1R, NC20-C20.....	405
<b>Table 68.</b> List of collocates of FENNTARHATÓ ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 3L-3R, NC20-C20.....	406
<b>Table 69.</b> List of collocates of FENNTARHATÓ ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 5L-5R, NC20-C20.....	408
<b>Table 70.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – Z(10.0), 1L-1R, NC20-C20.....	408
<b>Table 71.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – Z(10.0), 3L-3R, NC20-C20.....	409
<b>Table 72.</b> List of collocates of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – Z(10.0), 5L-5R, NC20-C20.....	412

## List of Figures

<b>Figure 1.</b> The seventeen Sustainable Development Goals of the United Nations' 2030 Agenda for Sustainable Development. ....	120
<b>Figure 2.</b> The structure of the present research in a flowchart.....	124
<b>Figure 3.</b> The identification of cultural keywords in a flowchart.....	125
<b>Figure 4.</b> The retrieval of meanings by collocation in a flowchart.....	126
<b>Figure 5.</b> Word sketch of SUSTAINABLE in the 2030 Agenda Corpus (English) – LogDice(6.0), NC5-C5.....	160
<b>Figure 6.</b> Collocation network of SUSTAINABLE in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.....	163
<b>Figure 7.</b> Collocation network of SUSTAINABLE in the 2030 Agenda Corpus (English) – Z(10.0), 3L-3R, NC5-C5.....	164
<b>Figure 8.</b> Collocation network of SUSTAINABLE in the 2030 Agenda Corpus (English) – Z(10.0), 5L-5R, NC5-C5.....	165
<b>Figure 9.</b> Word sketch of FENNTARTHATÓ 'sustainable' in the 2030 Agenda Corpus (Hungarian) – LogDice(6.0), NC5-C5.....	183
<b>Figure 10.</b> Collocation network of FENNTARTHATÓ 'sustainable' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.....	186
<b>Figure 11.</b> Collocation network of FENNTARTHATÓ 'sustainable' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.....	187
<b>Figure 12.</b> Collocation network of FENNTARTHATÓ 'sustainable' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 5L-5R, NC5-C5.....	188
<b>Figure 13.</b> Word sketch of SOSTENIBILE 'sustainable' in the 2030 Agenda Corpus (Italian) – LogDice(6.0), NC5-C5.....	194
<b>Figure 14.</b> Collocation network of SOSTENIBILE 'sustainable' in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.....	197
<b>Figure 15.</b> Collocation network of SOSTENIBILE in the 2030 Agenda Corpus (Italian) – Z(10.0), 3L-3R, NC5-C5.....	198
<b>Figure 16.</b> Collocation network of SOSTENIBILE 'sustainable' in the 2030 Agenda Corpus (Italian) – Z(10.0), 5L-5R, NC5-C5.....	199
<b>Figure 17.</b> Word sketch of SUSTAINABILITY in the 2030 Agenda Corpus (English) – LogDice(6.0), NC5-C5.....	205
<b>Figure 18.</b> Word sketch of FENNTARTHATÓSÁG 'sustainability' in the 2030 Agenda Corpus (Hungarian) – LogDice(6.0), NC5-C5.....	206
<b>Figure 19.</b> Word sketch of SUSTAINABLEDEVELOPMENT in the 2030 Agenda Corpus (English) – LogDice(6.0), NC5-C5.....	210

<b>Figure 20.</b> Collocation network of SUSTAINABLE DEVELOPMENT in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.....	212
<b>Figure 21.</b> Collocation network of SUSTAINABLE DEVELOPMENT in the 2030 Agenda Corpus (English) – Z(10.0), 3L-3R, NC5-C5.....	213
<b>Figure 22.</b> Collocation network of SUSTAINABLE DEVELOPMENT in the 2030 Agenda Corpus (English) – Z(10.0), 5L-5R, NC5-C5.....	214
<b>Figure 23.</b> Word sketch of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – LogDice(6.0), NC5-C5.....	232
<b>Figure 24.</b> Collocation network of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.....	235
<b>Figure 25.</b> Collocation network of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.....	236
<b>Figure 26.</b> Collocation network of FENNTARTHATÓ FEJLŐDÉS ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 5L-5R, NC5-C5.....	237
<b>Figure 27.</b> Collocation network of SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.....	247
<b>Figure 28.</b> Collocation network of SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 3L-3R, NC5-C5.....	248
<b>Figure 29.</b> Collocation network of SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 5L-5R, NC5-C5.....	249
<b>Figure 30.</b> Word sketch of SUSTAINABLE in the SusCorp (English) – LogDice(6.0), NC20-C20.....	284
<b>Figure 31.</b> Collocation network of SUSTAINABLE in the SusCorp (English) – Z(10.0), 1L-1R, NC20-C20.....	287
<b>Figure 32.</b> Collocation network of SUSTAINABLE in the SusCorp (English) – Z(10.0), 3L-3R, NC20-C20.....	288
<b>Figure 33.</b> Collocation network of SUSTAINABLE in the SusCorp (English) – Z(10.0), 5L-5R, NC20-C20.....	289
<b>Figure 34.</b> Word sketch of FENNTARTHATÓ ‘sustainable’ in the SusCorp (Hungarian) – LogDice(6.0), NC20-C20.....	301
<b>Figure 35.</b> Collocation network of FENNTARTHATÓ ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 1L-1R, NC20-C20.....	304
<b>Figure 36.</b> Collocation network of FENNTARTHATÓ ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 3L-3R, NC20-C20.....	305
<b>Figure 37.</b> Collocation network of FENNTARTHATÓ ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 5L-5R, NC20-C20.....	306
<b>Figure 38.</b> Word sketch of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – LogDice(6.0), NC20-C20.....	311

<b>Figure 39.</b> Collocation network of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – Z(10.0), 1L-1R, NC5-C5.....	313
<b>Figure 40.</b> Collocation network of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – Z(10.0), 3L-3R, NC5-C5.....	314
<b>Figure 41.</b> Collocation network of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian) – Z(10.0), 5L-5R, NC5-C5.....	315
<b>Figure 42.</b> The combined collocation networks of SOSTENIBILE ‘sustainable’ and sviluppo ‘development’ in the SusCorp (Italian) – Z(10.0), 5L-5R, NC20-C20.....	316



## List of Graphs

<b>Graph 1.</b> Distribution of tokens across newspapers in SusCorp (English).....	92
<b>Graph 2.</b> Distribution of tokens across newspapers in SusCorp (Hungarian).....	93
<b>Graph 3.</b> Distribution of tokens across newspapers in SusCorp (Italian).....	94
<b>Graph 4.</b> Distribution of the word types belonging to the categories of humans, animals, plants and natural elements. ....	155
<b>Graph 5.</b> Distribution of the tokens belonging to the categories of humans, animals, plants and natural elements. ....	156
<b>Graph 6.</b> The colligational patterns of SUSTAINABLE and nouns in the 2030 Agenda (English). .....	169
<b>Graph 7.</b> The colligational patterns of SUSTAINABLE and adjectives in the 2030 Agenda (English).....	171
<b>Graph 8.</b> The colligational patterns of SUSTAINABLE and verbs in the 2030 Agenda (English). .....	175
<b>Graph 9.</b> The colligational patterns of SUSTAINABLE and prepositions in the 2030 Agenda (English).....	178
<b>Graph 10.</b> The colligational patterns of FENNTARTHATÓ ‘sustainable’ in the 2030 Agenda (Hungarian).....	193
<b>Graph 11.</b> The colligational patterns of SOSTENIBILE ‘sustainable’ in the 2030 Agenda (Italian). .....	203
<b>Graph 12.</b> The colligational patterns of SUSTAINABLE DEVELOPMENT and nouns in the 2030 Agenda (English).....	218
<b>Graph 13.</b> The colligational patterns of SUSTAINABLE DEVELOPMENT and adjectives in the 2030 Agenda (English).....	220
<b>Graph 14.</b> The colligational patterns of SUSTAINABLE DEVELOPMENT and lexical verbs in the 2030 Agenda (English).....	222
<b>Graph 15.</b> The colligational patterns of SUSTAINABLE DEVELOPMENT and prepositions in the 2030 Agenda (English).....	227
<b>Graph 16.</b> The colligational patterns of the nominative fenntartható fejlődés ‘sustainable development’ in the 2030 Agenda (Hungarian).....	243
<b>Graph 17.</b> The colligational patterns of the adjectival form fenntartható fejlődési ‘sustainable development’ in the 2030 Agenda (Hungarian).....	244
<b>Graph 18.</b> The colligational patterns of noun phrases including SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda (Italian).....	254
<b>Graph 19.</b> The colligational patterns of prepositional phrases including SVILUPPO SOSTENIBILE ‘sustainable development’ in the 2030 Agenda (Italian).....	255
<b>Graph 20.</b> The colligational patterns of SUSTAINABLE and nouns in the SusCorp (English). 292	

**Graph 21.** The colligational patterns of SUSTAINABLE and adjectives in the SusCorp (English).  
..... 294

**Graph 22.** The colligational patterns of SUSTAINABLE and lexical verbs in the SusCorp  
(English)..... 296

**Graph 23.** The colligational patterns of FENNTARTHATÓ ‘sustainable’ in the SusCorp  
(Hungarian)..... 309

**Graph 24.** The colligational patterns of SOSTENIBILE ‘sustainable’ in the SusCorp (Italian).. 320

## Introduction

In our Western society, reference to green issues is made on a daily basis. European politicians and media often mention sustainable development, green deals, extreme weather events due to climate change. But what do politicians and media mean when they speak about “sustainable development”? What do they really believe “sustainable development” to be? And what political and cultural issues do they tackle in relation to sustainable development?

These questions can be addressed by several disciplines, among which linguistics plays a paramount role. Within linguistics, for instance, Antelmi (2018) notices that the use of the lexical item *sustainable development* in discourse has become pervasive. Alexander (2002) observes that *sustainable development* is frequently surrounded by an aura of vagueness. Mahlberg (2007) studies the semantics of the lexeme by analysing the words that it co-occurs with most significantly and she notices that in news discourse the expression *sustainable development* is meaningfully constructed as a goal to be achieved. What seems to characterise the meaning of *sustainable development* in the news is not an inherent quality of the concept but the attitude that those who are engaged with sustainability have in relation to it (Mahlberg 2007). In addition, and from an ecological perspective, according to Naeem et al. (2016) the discursive construction of *sustainable development* is once again one of achievement but with an anthropocentric attitude to the matter, which does not always consider the remaining elements of the natural environment.

These studies contribute to defining the meaning of *sustainable development* and to identifying its most typical use in language. However, as Halliday (1992: 65) states,

Semogenic processes – processes of constructing meaning – cannot be understood *outside* of their historical contexts; but neither can they be *derived from* these contexts by any simple relation. Let us put it this way: language is at the same time a part of reality, a shaper of reality, and a metaphor for reality. Once any form of language – any grammar – has come into being, it participates itself in the shaping of historical processes, including those which constitute the means and relations of production.

The existence of a bidirectional relationship between language and reality reinforces the importance and utility of studying the way issues of utmost importance like sustainable development are linguistically constructed in discourse. Linguistic

research on politically, socially or culturally meaningful issues allows to describe how these matters are shaped by reality and how they can shape it back, contributing or impeding the well-being of those human communities and natural environments that are touched by these matters.

Based on these assumptions, the present research aims at investigating the linguistic construction of sustainable development in English, Hungarian and Italian discourse, focusing on political discourse represented by the United Nations' *2030 Agenda for Sustainable Development* and on news discourse appeared after the release of the UN's 2030 Agenda. More precisely, the research intends to answer the following research questions. First, in relation to the 2030 Agenda:

1. What are the politically, socially and culturally most salient lexemes emerging in the English, Hungarian and Italian versions of the 2030 Agenda? And how are the economic, social and environmental dimensions of sustainable development lexically represented in the document?
2. What meaning do *sustainable development*, *sustainable*, and *sustainability* and their Hungarian and Italian translational equivalents acquire in the UN's resolution?

Second, regarding news discourse

3. What are the politically, socially and culturally most significant words emerging in the English, Hungarian and Italian news discourse on sustainable development? And how are the economic, social and environmental dimensions of sustainable development lexically represented in this kind of discourse?
4. What meaning do *sustainable* and its Hungarian and Italian corresponding lexical items acquire in news discourse?

The identification of politically, socially and culturally significant words is carried out under the light of the notion of *cultural keyword* (Williams 1983), whereas the study of meaning is conducted thanks to the notion of *meaning by collocation* (Firth 1957a, 1957b). Cultural keywords and meaning by collocation are analysed within the theoretical framework of *cross-linguistic, corpus-assisted discourse studies* (Taylor and Marchi 2018), and the results of the analysis are interpreted under the light of the theoretical framework offered by *ecological discourse analysis* (Alexander and Stibbe 2014). While research on cultural keywords and meaning by collocation has a long tradition, and corpus-assisted approaches to the ecological analysis of discourse have been carried out for a decade, it seems that the combination of these notions and

theoretical frameworks for the cross-linguistic study of languages like English, Hungarian and Italian is still unprecedented.

The dissertation is organised into six chapters. After this introduction, in the first chapter I discuss in depth the theoretical background of the research. I start with an account of the notions of cultural keyword and of meaning by collocation. Then, I summarise the main tenets of cross-linguistic, corpus-assisted discourse analysis and I connect it to ecological discourse analysis. I highlight the value of this approach for the cross-linguistic study of the discursive construction of sustainable development. In the second chapter, I present the data used for the analysis, namely the *2030 Agenda Corpus* (the original UN's document) and the *Sustainable development Corpus* (a reference corpus of news discourse built on purpose for this study). In the third chapter, I account for the methodology adopted during the analysis of the corpus. First, I focus on the methodological aspects of the identification of cultural keywords and of meaning by collocation. Then, I outline some of the main tenets necessary for the cross-linguistic and ecological interpretation of the results. In the fourth and in the fifth chapters of the thesis I report the results of the analysis of the discursive construction of sustainable development in the *2030 Agenda Corpus* and of the *Sustainable development Corpus* in turn. In the sixth chapter, I conclude by summarizing the results of the analysis and by opening them to an interdisciplinary interpretation. First, I compare the discursive construction of sustainable development in the *2030 Agenda Corpus* and in the *Sustainable development Corpus*. Then, I interpret these findings in light of the current ecological theories considered for the current research. Finally, I end by listing some of the pitfalls of this approach for a cross-linguistic, discursive investigation of sustainable development, and I introduce potential future developments of the present work.



# 1. Corpus and Discourse approaches to ecology across languages

The first chapter of the dissertation presents the theoretical framework that shapes the research, namely a corpus approach to the study of cultural keywords and meaning by collocation within an ecological framework and with a cross-linguistic perspective. First, I briefly describe the notion of cultural keyword and its development in linguistic and cultural studies. I match cultural keywords and *statistical keywords*. Then, I outline the evolution of the concept of meaning by collocation in linguistics, focusing on the way it can be employed to study the semantics of lexical items. I concentrate on the notions of *collocation*, *colligation*, *semantic preference*, and *evaluative prosody*. Second, I describe corpus approaches to the cross-linguistic study of discourse from an ecological perspective. I start with an overview of corpus-assisted discourse studies, and I detail the main characteristics of two of its founding disciplines, namely *critical discourse studies* and *corpus linguistics*. I continue by presenting ecological discourse analysis and by summarising the way it can be aided by corpus approaches, and what kind of assistance can be provided by corpora when carrying out cross-linguistic analyses of discourse.

## 1.1. Cultural keywords

The first concept employed in the current research for the study of the discursive construction of sustainable development is that of cultural keyword. The concept of cultural keyword dates back to German- and French-language work of the early Twentieth century on “dictionaries of words which are important in social and intellectual history” (Stubbs 2010: 23). It establishes then in the work of Raymond Williams and it develops further in the studies of Anna Wierzbicka on the cross-cultural distribution of keywords and in the corpus-aided research on lexical semantics of Michael Stubbs.

*Keywords as “the vocabulary we share with others” (Raymond Williams).* Variably rooted in the fields of “cultural history, historical semantics, history of ideas, social criticism, literary history and sociology”, Raymond Williams’s cultural keywords consist in the “vocabulary we share with others, often imperfectly, when we wish to discuss many of the central processes of our common life” (Williams 1983: 13-14). Williams records

more than one hundred keywords that he believes to stand out and to shape the society he is part of (e.g. *capitalism, image, sociology, violence*; Williams 1983).

These cultural keywords are philologically illustrated in terms of the linguistic context they are included in and according to their historical development. For example, referring to *ecology*, Williams (1983) states that the word is a compound of the Greek *oikos* ('household') and *logos* ('discourse') and that it shares its *eco-*morpheme with the word *economics*. *Ecology* is introduced in English in the late 19<sup>th</sup> Century meaning "the study of plants and animals with each other and with their habitat" (Williams 1983: 111). Later it acquires a connotation of social concern and it is used as a synonym of *environmentalism* (interpreted first as the influence of the environment on development and then as the more specific human attitude towards the preservation of the Earth against pollution). It is only in the 1960s that *ecology* starts substituting *environmentalism* and all the compounds of *environment*. When Williams writes his book, *ecology* is concerned with the relationship between social, political, economic interests on the one side and the physical surroundings on the other. Moreover, it seems to be stimulating the rise of neologisms such as *ecocrisis* and *ecocatastrophe* (Williams 1983).

The list and meaning of Williams's cultural keywords and their function for the interpretation of society and institutions is reached by means of introspection and research on the ideas of popular philosophers and thinkers. Although the meaning of the selected cultural keywords is said by Williams (1958: xvii) to be found "in language", the writer's approach to the study of these lexemes is cultural and philosophical.

*Cultural keywords across languages (Anna Wierzbicka)*. According to Anna Wierzbicka (1997: 15-16), cultural keywords "are words which are particularly important and revealing in a given culture". They are not limited to a finite-size set of words and their subjective collection procedure might follow the following question:

How can one justify the claim that a particular word is one of a culture's "key words"? To begin with, one may want to establish (with or without the help of a frequency dictionary) that the word in question is a common word, not a marginal word. One may also want to establish that the word in question (whatever its overall frequency) is very frequently used in one particular semantic domain, for example, in the domain of emotions, or in the domain of moral judgements. Furthermore, one may want to show that this word is at the center of

a whole phraseological cluster [...] One may also be able to show that the proposed “key word” occurs frequently in proverbs, in sayings, in popular songs, in book titles, and so on. (Wierzbicka 1997: 16)

Wierzbicka’s cultural keywords are analysed as linguistic items. Their cultural content is assessed through the study of their collocational and idiomatic behaviour, namely through the study of the words and multi-word units they co-occur with in a statistically significant way (for a thorough account of collocation see § 1.2 and § 1.3.2. in this chapter). In-depth knowledge of this cultural content allows “to show the general organizing principles which lend structure and coherence to a cultural domain as a whole, and which often have an explanatory power extending across a number of domains” (Wierzbicka 1997: 16-17). This potentiality enables cultural keywords to be researched also cross-culturally to identify similarities and differences in the conceptualisation of words and cultures (Wierzbicka 1997; see also Wierzbicka 1996).

*Cultural keywords and their meaning by collocation (Michael Stubbs)*. Also Michael Stubbs’s cultural keywords are “words and phrases which have particular cultural significance” (Stubbs 2001: 145). They are sprinkled with well-established features of the previous tradition (especially Williams’s) and with essential elements of novelty: the notion of cultural keyword remains ontologically intact, maintaining its status of culturally meaningful term. Nevertheless, the retrieval techniques and the approach adopted for their analysis changes significantly, involving empirical tests and the perusal of the highlighted words’ *concordances*. Concordances consist in all occurrences of a word inserted in the textual environment that the word belongs to. Historical semantics and the study of words in historical and social context leave the ground to corpus-assisted analyses of discourse and to a computational interest for the common behaviour of words in ideologically charged contexts (Stubbs 2001; for more on ideology in discourse see § 1.3.1).

In search for cultural keywords, Stubbs selects his terms with a combination of introspection and empirical observation. He integrates his intuitive reflection on the world with commentaries previously produced by cultural analysts on significant terms. Stubbs (2001: 167) also weighs the frequency of occurrence of the same terms in books, magazines, mass media or “prominent statements by prominent speakers”. He does not focus on politically or ideologically loaded words (like Williams’s *imperialism*), but he does leave space for apparently innocuous items such as *little* or *proper*, in the belief that culture is encoded through a combination of common words (Stubbs 2001).

Apart from divergences in mode for the retrieval of cultural keywords and the type of words to be picked, the real clash between Williams's and Stubbs's approaches to cultural keywords lies in the relation that they assign to culture and language: while for the former it is culture to influence the panorama of politically, socially or culturally meaningful lexical items that could be possibly signalled in discourse, the latter judges culture to be fashioned by linguistic habits, "the repertoire of shared meanings which circulate in a community" (Stubbs 2001: 146). In other words, according to Williams, culture produces keywords whereas for Stubbs keywords and their semantic networks are at the same time products and producers of culture (Stubbs 2001).

Stubbs (1996) believes that cultural keywords could not be studied only in their historical and social dimensions: he assumes that an initial exploration of meanings and occurrences in dictionaries needs to be complemented by a thorough empirical analysis of the selected word in a large computer *corpus* (i.e. in an electronic collection of language data serving the purposes of linguistic analyses) so as not to miss relevant linguistic and cultural patterns.

For example, in exploring meaning and use of the lemma WORK (present also in Williams's list of cultural keywords), Stubbs (1996: 177) starts from Williams's first considerations: work is part of a process whereby "people who work have a job which they are paid to do", as in *My sister is working as a shop assistant*. In a second stage of the analysis, Stubbs strengthens his first thoughts with data computationally extracted from corpora. In particular, he notices that the various word-forms belonging to the lemma WORK are surrounded by different linguistic patterns, so that every word-form has a specific textual behaviour and contributes to building a wide range of compound words and fixed expressions. *Working*, for example, is frequently part of phrases like *working class*, *working conditions* or *working mother*; *work* features in compounds such as *workaholic*, *workforce*, *workplace* and *workstation* whose more or less recent birth proves the compounding process productive; *worker* seems to co-occur with terms like *factory*, *social*, *airport* and *bank* (in *factory worker*, *social worker*, *airport worker* and *bank worker*) and some of these expressions are progressively substituting other combinations (for instance, in *bank worker* used instead of *bank clerk*). Corpus data show that the compounds of the word *worker* cover a large set of jobs, from blue- to white-collar positions, and that the identification and characterisation of working people occupied writers and researchers in 1950s' and 1960s' United Kingdom (Stubbs 1996).

## 1.2. Meaning by collocation

The second concept utilised in the current research for the description of the discursive construction of sustainable development is meaning by collocation. Collocation is “certainly one of the most controversial notions in linguistics, even though it is based on a compelling, widely shared intuition that certain words have a tendency to occur near each other in natural language” (Evert 2008: 1212).

The notion is elaborated in the most renowned formulation by John Rupert Firth in the late 1950s (1957a, 1957b) but it seems to have been already in use in the first half of the 20<sup>th</sup> Century, as Doyle (2003) notices. Firth’s seminal notion of collocation is further strengthened in the studies of the so called “neo-Firthian” linguists, as McEnery and Wilson (2001: 23) call them. The group of the neo-Firthian linguists includes most notably Michael Halliday, Michael Hoey and John Sinclair. In addition, the concept of collocation is developed further also by Michael Stubbs.

*Collocation as “mode of meaning” (John Rupert Firth).* Collocation is defined by Firth (1957a) as a phenomenon whereby “actual words” are found “in habitual company”, namely when a specific lexical item habitually occurs next to another (Firth 1957a: 14). As Firth notices, for instance, the word *time* can collocate with *saved, spent, wasted, frittered away, etc.*: in other terms, when the former appears in discourse it is frequently found in company of the latter (Firth 1957b).

Collocation in Firth (1957a, 1957b) is considered as a component of the semantics of a word and, by extension, of a sentence. Firth (1957b: 196) believes that the habitual company that words keep functions as a “mode of meaning” in that it provides the word with a certain meaning gathered from the terms it collocates with most frequently; this process is called “meaning by collocation”. This meaning potential of a word is expressed not only thanks to the lexical company it is kept by other words but also through its habitual occurring together with grammatical items of a regular kind (namely through *colligation*; Firth 1957a).

According to Firth (1957b: 195-196), collocational meaning “is not at all the same thing as contextual meaning, which is the functional relation of the sentence to the processes of a context of situation in the context of culture”; on the contrary, “meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words”. This entails that a word does not obtain meaning *per se* but thanks to its habitually appearing close to another word. For instance, *time* acquires a specific meaning in a sentence or discourse

thanks to the word that it is found in company of: it means differently when it is associated with *saved* then when it is co-textually uttered with *frittered away*.

Firth deals with collocation within the framework of semantics: he deems it as a phenomenon that contributes to the meaning of a word in actual context, as it summarised by Saeed (2015). His seminal works (Firth 1957a, 1957b) pave the way for a deeper understanding of the role of collocation in semantics. Halliday takes on Firth's approach to collocation as a mode of meaning and he extends it to a textual perspective.

*Collocation as lexical cohesion* (Michael Halliday). Halliday and Hasan (1976: 287) write that collocation "results from the co-occurrence of lexical items that are in some way or other typically associated with one another, because they tend to occur in similar environments". For example, *fork* often appears together with *knife*, *lend* is frequently combined with *money* (Halliday 2004: 11).

As it is reported in Kress (1976), Halliday's (1966) seminal work states that collocation consists in the tendency of two words to consistently appear together within a sentence or across sentence boundaries regardless of their grammatical relation. Grammar, in fact, functions as a founding structure that welcomes lexical choices bound one to the other by means of collocation (Halliday 1966).

Collocation can be both a purely lexical relationship between words (e.g. *lend-money*) or it can bear a semantic value (as in *fork-knife*). It generates a semantic relationship whenever the meaning of the collocates is related in some way; for instance, two co-occurring words could be synonyms or near-synonyms or they could be tied by hyponymy or meronymy. Collocation is purely a lexical phenomenon if it keeps together semantically unrelated words (Halliday and Matthiessen 2004).

In general, however, the tendency of words to co-occur can be interpreted as "a significant feature of the meaning" of the lexeme in the analysed text (Halliday and Matthiessen 2004: 38). This implies that the meaning of a word is constructed in text thanks to the word's regular co-occurring with semantically related or unrelated lexemes.

This meaning and the collocational patterns it is associated with vary according to the "particular register, or functional variety of the language" considered for exploration; for example, the meaning of *hunting* will be constructed through the collocating *quarry* and *hound* in the discourse of English aristocracy, while it will be shaped by words like *gathering*, *agricultural* and *pastoral* in an anthropological writing (Halliday and Matthiessen 2004: 577-578).

Furthermore, not only does Halliday (1978) believe that collocation can impact on the stabilised meaning of a word, but he does also retain it as an element characterising the development of meaning in language across time. When discussing about the raise of new words in technical fields (such as mathematics), for example, he states that

only gradually as the new words become familiar, and come to be used in broader and broader ranges of verbal environments - in new structures, and particularly in new collocations, and new arguments - will they be fully domesticated in the language, and acquire a wide range of associations. (Halliday 1978: 203)

Collocation's role in generating meaning impacts also on a socio-semiotic perspective: the contextual meaning acquired or displayed by a word in text and the consequent meaning of the text are mirrored also in the Malinovskian "context of situation" that produces and receives the word's and text's meaning (Halliday 1975: 65; see also Halliday 1978). In this way, collocation links a meaningful constellation of intertwined words to the complexity of a culture.

In addition to shaping the meaning of a word, in fact, collocational patterns do also "contribute significantly to the unfolding meaning of a text" (Halliday and Matthiessen 2004: 39). Among other things, the meaning of a specific text is accordingly built thanks to the accumulation and interaction of frequently co-occurring words. In Halliday and Hasan (1976: 286), collocation, in fact, can stretch both "within the same sentence or across sentence boundaries" and it might participate in providing lexical cohesion to a text.

Nonetheless, when collocation functions as a cohesive device, its cohesive effect "depends not so much on any systematic semantic relationship" but on the word's "tendency to share the same lexical environment"; cohesion can thus be "achieved through the association of lexical items that regularly co-occur" (Halliday and Hasan 1976: 284-286). According to Halliday and Hasan (1976: 286), collocation, therefore, contributes to textuality by knitting together parts of a sentence or by bridging different sentences through the regular co-occurrence of word pairs or word chains. It can match a pair of words (e.g. *door-window* or *king-crown*) but it can additionally involve a longer list of items as *candle-flame-flicker* or *sky-sunshine-cloud-rain*.

These patterns of collocational co-occurrence are identified by Halliday and Hasan (1976) only within text boundaries and they do not apply to an entire corpus. Later redefining collocation from a corpus linguistics perspective, Halliday and

Matthiessen (2004) state that the estimate of a collocational pattern can be regarded as “the degree to which the probability of a word (lexical item) increases given the presence of a certain other word (the **node**) within a specified range (the **span**)” (Halliday and Matthiessen 2004: 38).

*Collocation and the lexical item (John Sinclair)*. John Sinclair’s treatment of collocation slightly varies from Halliday’s for two main reasons: first, it is deeply rooted in the field of corpus linguistics since the very beginning; second, it starts as the starting point of a different theory of grammar, namely a theory of grammar which is based on lexis through the *idiom principle* and the concept of *lexical items*.

Introduced by Sinclair (1991: 110), the idiom principle is a principle whereby “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments”. These idioms can be interpreted as fossilizations of collocational patterns: collocation, in fact, implies that “the choice of one word conditions the choice of the next, and of the next again” until the choice stabilises into an idiom (Sinclair 2004: 19).

However, Sinclair (2004: 20) states that the “precise identification of idioms is by no means clear-cut”. The idiom principle and the idea of co-selection is much broader than what is traditionally classified as idiom and applies to many linguistic choices where word co-selection is observable. The idiom principle can thus be implied by the lexical item (Sinclair 1998).

A lexical item is “a unit of language representing a particular area of meaning which has a unique pattern of co-occurrence with other lexical items” (Sinclair et al. 2004: 9; for a development of the lexical item into a *linear unit* see Sinclair and Mauranen 2006). According to Sinclair et al. (2004: 9), the lexical item is opposed to the grammatical item (i.e. “a unit of language whose presence in the text is due to its grammatical function rather than to any “meaning” it may represent”) and its boundaries are set by collocational patterns. A lexical item can be identified within the span of a collocational pattern and its structure can be formalized through patterns of increasing levels of abstractness (collocation, colligation, semantic preference and semantic prosody as it is explained in Sinclair 1996).

Since the first works, Sinclair (1991: 170) defines collocation as “the occurrence of two or more words within a short space of each other in a text”. The co-occurring words are respectively called *node* and *collocate*: the node is “the word that is being studied” and the collocate is “any word that occurs in the specified environment of a

node” (Sinclair 1991: 115). This specified environment is called *span* and it consists in “the amount of text within which collocation between items is said to occur” (Sinclair et al. 2004: 10).

Sinclair (1991) classifies collocation as *downward collocation* and *upward collocation*: downward collocation is found whenever the collocate is less frequent than the node in a corpus; upward collocation is identified in the opposite case, namely when a node is less frequent than the collocate in a corpus.

Sinclair et al. (2004: 10) further distinguish collocation between *significant collocation* and *casual collocation*: significant collocation “is regular collocation between two items, such that they co-occur more often than their respective frequencies, and the length of text in which they appear, would predict” while casual collocation is identified whenever collocation is not significant. They add that “from the linguistic point of view, there is no hard and fast distinction between a casual and regular collocation, simply different degrees of probability” (Sinclair et al. 2004: 72).

Sinclair elaborates on his theory of collocation and he affirms that the syntactic and semantic relations linking a node and its collocates can be described in terms of colligation, semantic preference and semantic prosody. Colligation is meant to be “the co-occurrence of grammatical choices” within a short space from each other in a text (Sinclair 1996: 84). Semantic preference is the tendency of a word to co-occur with lexemes belonging to specific semantic areas whereas semantic prosody is the word’s tendency to be found together with lexemes triggering a positive or negative connotation for the meaning of the word (Sinclair 1991).

For example, Sinclair (1996) writes that in The Bank of English<sup>1</sup> the strongest collocational pattern for the word *brook* (meaning ‘tolerate’) involves negatives both on the left-hand side of the word (e.g. *not, cannot*) and on its right-hand side (e.g. *no, not*). In addition, on the left-hand side, the verb colligates with modal verbs (as *will* and *would*) while on the right-hand side its most prominent colligational pattern is filled by nouns. The verb’s most peculiar semantic preference is that of intrusion and its prosody “partly concerns the absence of something (in this case first and second person subjects), and partly includes words like *said, make/made clear, shows, indication*” (Sinclair 1996: 89). The collocational and colligational networks of *brook* together with

---

<sup>1</sup> The Bank of English is a monitor corpus of written English (mainly British, with some texts produced in North America, Australia, New Zealand, South Africa, etc.) developed from the 1980s as a part of the Cobuild project (Sinclair 1987). Monitor corpora are corpora constantly updated with new language data. In the mid-1995, when Sinclair’s (1996) study was conducted, The Bank of English consisted of roughly 200 million words (Sinclair 1996).

its semantic preference and prosody shape for it a meaning of intolerance for “intrusive behaviour by another” and they exemplify the way in which lexical units mean in language thanks to their being inherently and extensively bound to the concept of collocation (Sinclair 1996: 89).

Sinclair’s idea of collocation is deeply grounded in a Firthian view of collocation and it developed by means of the corpus linguistic perspective. As Sinclair et al. (2004) suggest, collocational patterns can be extracted from a corpus within a pre-defined span by exploiting the frequencies of node and collocate for the calculation of the statistical significance of the collocation. Statistical significance assures that “there is but a small chance of it [a co-occurrence of words] being accidental” (Sinclair et al. 2004: xxi; collocation is exhaustively described in § 1.3 in this chapter).

The aforementioned Halliday and Matthiessen (2004) corpus-based definition of collocation owes much to John Sinclair’s work on the topic. The definition of the concept is similar in the three authors. However, the starting point of their reflection is different. Moreover, Sinclair pushes forward the corpus approach that was present only in essence in Halliday’s work and he advances in the then rising field of corpus semantics, as it is recalled by Stubbs (2015).

*Collocation and lexical priming (Michael Hoey).* John Sinclair’s discussion on collocational patterns is the starting point for Michael Hoey’s theory of collocation and lexical priming in text and discourse.

According to Hoey (2005: 2), collocation is a “property of language whereby two or more words seem to appear frequently in each other’s company (e.g. *inevitable + consequence*)”. This frequent co-occurrence of words in text or discourse is called *lexical priming* and it can revolve around single words or multi-word expressions. Lexical priming is at the basis of a grammar in which “lexis is complexly and systematically structured” and “grammar is an outcome of this lexical structure” (Hoey 2005: 1). Lexical priming triggers pragmatic, semantic and syntactic patterns revolving around the lexeme whose collocation network has been expanded on.

Pragmatic patterns are named *pragmatic associations* and they occur “when a word or word sequence is associated with a set of features that all serve the same or similar pragmatic functions (e.g. indicating vagueness, uncertainty)” (Hoey 2005: 26). For instance, in a specialised corpus used by Hoey (2005: 28) for the sake of explication, *reason* is often found in company of words or word sequences expressing denial, such as *But there was no, Really I see no*, etc. The notion of pragmatic association is similar to Sinclair’s notion of semantic prosody

These pragmatic associations glide towards semantic patterns called *semantic associations*. Semantic associations are the tendency of words or word sequences to appear next to items belonging to the same “semantic set” (Hoey 2005: 18). In the aforementioned corpus, for example, the noun *hour* tends to co-occur with numbers and with words belonging to the semantic set of *journey* (e.g. *half-hour drive* or *two-hour trip*; Hoey 2005: 16). Hoey’s notion of semantic association is tantamount to Sinclair’s notion of semantic preference.

These pragmatic and semantic associations display a syntactic counterpart too. This syntactic counterpart is called colligation. Similarly to Firth (1957a) and Sinclair (1991), Hoey (2004: 389) defines colligation as

- a) the grammatical company a word keeps, or avoids keeping, either within its own group or at a higher rank;
- (b) the grammatical functions that a word’s group prefers or avoids;
- (c) the place in a sequence that a word prefers or avoids.

Patterns of this kind rely on the concept of collocation, which might be psychologically interpreted as a manifestation of *priming*. Priming has been dealt with first by Hoey (2004, 2005, 2017), but also by Durrant and Doherty (2010), Pace-Sigge (2018), Pace-Sigge and Patterson (2017). Priming is a psychological process that can be used to account for collocation

if we assume that every word is mentally **primed** for collocational use. As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context. (Hoey 2005: 8)

Priming is thus context-specific and personal. It is context-specific in that the range of collocates that a word attracts depends on the specific social and textual context in which they are encountered and produced. It is personal because it is intertwined with the individual’s experience of a language and it cannot be generalised for a whole population: words are differently primed for different people according to the different contextual and textual experiences they have lived (Hoey 2005).

Because of its being context-specific and personal, priming might feature a transitory or (semi)permanent behaviour. When a word or word sequence is repeatedly associated with another, its collocational patterns might stabilize and

contribute to the cohesion of a text. In line with Halliday and Hasan (1976), Hoey believes that collocational priming can be interpreted as a hallmark of textual cohesion (Hoey 2004, 2005, 2017).

On the other hand, collocational priming “may accordingly shift in the course of an individual’s lifetime, and if it does so, and to the extent that it does so, the lexical item shifts slightly in meaning and/or function” (Hoey 2004: 386). In this case, it is possible to witness the phenomenon of *drift in collocation*, that can lead to language change when it is generalized from a single language user to the whole community (Hoey 2005).

What distinguishes Hoey’s theory of collocation from Sinclair’s is Hoey’s focus on the psychological nature of the phenomenon, both from a theoretical point of view and from a textual and corpus perspective.

The relationship of the concept of collocation with corpus linguistics, in fact, is one in which corpora can highlight the individual tendencies of priming although all “that a corpus can do is indicate that certain primings are likely to be shared by a large number of speakers, and only in that sense is priming independent of the individual” (Hoey 2005: 15). According to Hoey (2005: 5), in fact, collocates are “evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution”.

Together with Hoey, Durrant and Doherty (2010) affirm that corpora star as important tools in listing a word’s or multi-word unit’s collocational patterns and that they can assist in further researching the psychological validity of collocates. Usual co-occurrence is measured in collections of texts by sheer frequency counts or with association measures, which may link simple linguistic data to the force of the psychological environment where these data are produced or received (Durrant and Doherty 2010).

In order to test the applicability of these ideas both from a psycholinguistic and from a corpus linguistics perspective, Durrant and Doherty (2010) compare the impact of *associative priming* with that of collocational priming, with associative priming meaning a form of linguistic priming that resembles collocational priming but that focuses especially on related words as a result of *association*. Association is the “relationship between a word and other words which it ‘brings to mind’” and it favours high-occurrence word-combinations (Durrant and Doherty 2010: 132). Durrant and Doherty (2010) show that collocational priming is more effective than associative priming in determining combinations of words, and that the psychological

strength of a node-collocate pair in collocational priming does not rely only on its commonality but also on the peculiarity of it.

In this panorama, while clustering to form collocation networks, words are treated exclusively by the mental lexicon, without requiring any effort from higher-order abilities to process the linguistic data and recognise common patterns. Durrant and Doherty (2010: 128), in fact, think that “speeded reaction to primed words is a result of neurological activation ‘spreading’ from the context word to related words” and that it does not exploit other mental capacities. According to Durrant and Doherty (2010), the speed and strength with which primed words are activated in the mind depends on the frequency and on the statistical relevance of their combined occurrence.

The idea of testing the psychological validity of collocation extracted from corpora is shared by Hughes (2017, 2018) and Hughes and Hardie (2020). Hughes (2018) explores the psychological validity of collocation with experiments directly measuring neural activity (i.e. electroencephalography, EEG) on native and non-native speakers of English. She observes that collocational pairs of adjective-noun bigrams demand less cognitive effort than non-collocational pairs to be processed both by native and non-native speakers of English. Furthermore, Hughes (2018) adds that the measure of load and confidence in processing corpus-derived collocation depends also on the association measure used when extracting collocates from a corpus. From a grammatical point of view, Hughes (2017: 2) concludes that collocation could be better understood within a “network model of language processing, whereby collocations are represented in the brain as transitions across a network” (for a recent network model of grammar see Diessel 2019).

*Collocation and semantics (Michael Stubbs)*. Although the idea of collocation found in Michael Stubbs’s work is similar to Sinclair’s and Hoey’s, their conceptualisation diverges in focus: Sinclair studies collocation as a means towards a description of language; Hoey brings in a psychological interpretation for the grammar of collocational patterns; Stubbs introduces collocation in a broad exploration of the linguistic, cultural and social features of discourse. Like Sinclair and Hoey, also Stubbs adopts a corpus approach to collocation.

Stubbs (1995a:1) defines collocation as “a relationship of habitual co-occurrence between words (lemmas or word-forms)”: in the LOB corpus<sup>2</sup>, for instance, the lemma CAUSE tends to co-occur with negative collocates belonging to a generic semantic field of *accident* (e.g. *accident, damage, harm, etc.*; see also Stubbs 1995b).

Collocation links a node word to a collocate, namely it connects a “word-form or lemma being investigated” with a “word-form or lemma which co-occurs with a node in a corpus” (Stubbs 2001: 29). The co-occurrence of node and collocate is generally located within a predefined span (i.e. “the number of word-forms, before and/or after the node”; Stubbs 2001: 29) and it is usually measured both in terms of frequency and with association measures like Mutual Information and t-score (Stubbs 1995a). The Mutual Information association measure “picks out lexical collocates (which are relatively infrequent)” while the t-score association measure “picks out both lexical and grammatical collocates” (Stubbs 1995a: 12).

Span, frequency, and association measures participate in the collection of a word’s collocates and, consequently, in the outlining of a word’s meaning. Stubbs (1995a: 19), in fact, retains that meaning “can be analysed empirically by methods of text and corpus analysis”. So, the meaning of a word can be read through its collocational patterns as “each word is represented by a set of values which comprise a list of the most significant collocates with associated statistics” (Stubbs 1995a: 15). Besides, meaning “is not constant across the inflected forms of a lemma”: the set of collocates that a lemma is endowed with varies according to the investigated word-form and so does meaning (Stubbs 1996: 40).

In addition, the study of collocational patterns functions as a preliminary step towards the identification of a word’s semantic preference and semantic prosody, namely to the recognition of the semantic areas and of the connotation that most frequently associate to the word (Stubbs 1995a, 1995b, 1996). Collocational patterns also open to the retrieval of a word’s discourse prosody, which corresponds to the distribution of the same connotational value as semantic prosody “over more than one unit in a linear string” (Stubbs 2001: 65).

---

<sup>2</sup> The LOB corpus (Lancaster-Bergen/Oslo corpus) is the British counterpart of the Brown corpus, a 1-million-word corpus created at the Brown University (Rhode Island, USA) in the early 1960s by sampling 500 texts of 2,000 words each, that belonged to 15 different textual genres (Francis and Kučera 1979). The LOB corpus was developed at Lancaster (United Kingdom), Bergen (Norway), and Oslo (Norway) in the same period and with the same sampling pattern. It contains British texts from 1961 totalling 1,020,445 running words (Johansson et al. 1978).

According to Stubbs (2001: 108), semantic preference, semantic prosody, and discourse prosody “contribute to textual cohesion and to intertextual relations between texts and corpus”. On the same matter, Stubbs (1995a) states that

It has doubtless often been noted that a series of words which are negative, unpleasant or pejorative (or positive, pleasant, etc) will contribute to cohesive texture. But I have here discussed a more specific cohesive mechanism. A “semantic prosody” stretches across a span of words, and therefore contributes to the cohesion of a text. An occurrence of CAUSE sets up an expectation of some unpleasant, probably abstract, word(s). If/when this occurs, then a little bit of textual cohesion results. (Stubbs 1995a: 21)

Collocational patterns can be deemed cohesive devices if they are to be found within a single text. If they are distributed across texts, they can be rather said to be intertextual devices. According to Stubbs (2001: 101), in fact, “intertextual relations between individual texts and routine language use are expressed largely in collocations”.

In comparison with Sinclair and Hoey, the peculiarity of Stubbs’s approach to the study of collocational patterns is the overwhelming value that he assigns to the relationship between collocation and discourse and between collocation and society, reviving Halliday’s idea of language as social semiotics. Stubbs (1996: 194) claims that by “searching out frequent collocations, we can glimpse the recurrent wordings which circulate in the social world, and glimpse how linguistic categories become social categories”. Hence, patterns of collocation in text and discourse come from and become patterns of meaning in society.

*More on semantic preference and semantic prosody.* The relationship between collocations and society can be better understood through the notions of semantic preference and semantic prosody, whose development is accounted in Forest (2007), Hunston (2007), Morley and Partington (2009), Partington (2004), Steward (2010), and Whitsitt (2005).

Semantic preference is described by Sinclair (1998) as “the restriction of regular co-occurrence to items which share a semantic feature”. It consists in the tendency of a word to attract collocates belonging to a specific semantic area and it varies according to the form of a lemma (Sinclair 1998; republished in Sinclair 2004: 142).

According to Sinclair (1998), in the analysis of a word’s recurrent textual environment, semantic preference follows a path of abstraction that starts from collocation and that culminates in semantic prosody. Collocation, in fact, traces the co-

occurrence of lexemes within a fixed textual window and it roots a study to the concreteness of a text's words. Colligation groups these co-occurring words according to their word class ("where there is a preponderance of one particular word class, this is colligation") and it forces thus to move from the concreteness of the word to the abstractness of its grammatical features. Semantic preference "requires us to notice similarity of meaning regardless of word class", although some colligational patterns might be tied to precise semantic preferences. In other words, it compels a researcher to delve into the more abstract field of word meaning (Sinclair 1998, republished in Sinclair 2004: 142).

Semantic preference is both a syntagmatic and a paradigmatic property of words (Sinclair 1998). From a syntagmatic point of view, it follows the linear distribution of linguistic material that is demanded by a word and its collocation network. From a paradigmatic point of view, it guides the distribution of semantically related words in specific slots of the collocational pattern (Sinclair 1998).

At the same time, also Stubbs (1995a: 2) notices that "words may habitually collocate with other words from a definable semantic set". Stubbs (2001: 65) defines thus semantic preference as "the relation, not between individual words, but between a lemma or word-form and a set of semantically related words". In other terms, semantic preference connects a lemma or a word-form with "a class of words which share some semantic feature"; this class "will have frequent and typical members, but will be open-ended" (Stubbs 2001: 88). Stubbs (2001: 65), for instance, mentions the tendency of *large* to be found together with words expressing quantities and size (e.g. *number(s), amounts, quantities...* as in *large amounts of money* or *large quantities of rice*).

Semantic prosody, on the other hand, is described by Louw (1993: 157) as the "consistent aura of meaning with which a form is imbued by its collocates". This "aura of meaning" encompasses both semantic and pragmatic features: it is best recognised as the positive, negative, or neutral connotation that a range of semantically related lexical items assigns to the word they co-occur with. The semantic prosody of *utterly*, for instance, is clearly negative as the right-collocates of the word usually include terms like *meaningless* (Louw 1993: 160).

Semantic prosody, however, might need to be established not simply through a word's collocates but also via the reading of larger chunks of text, which challenges sometimes the computational identification of this feature. A further challenge resides in the importance of gathering as much linguistic material as possible: semantic

prosody, in fact, is established when it is shared by many instances of language (Louw 1993).

Nevertheless, apart from these challenges, semantic prosody retains an essential diagnostic power. According to Louw (1993), in fact, breaches in the use of semantic prosody can either help spotlight irony or insincerity of a writer or a speaker or they can be signposts of attitudinal stances. In the first case, for instance, if a writer or a speaker violates the common semantic prosody of a word on purpose, then they could make their sentences ironic or insincere. Louw (1993: 167-168) produces an example with *fine*: having a straightforward positive prosody when associated with professions (like *actress*) or to objects (such as *china*), it generates a sense of strangeness if it is combined with people or family members (as in the quotation from David Lodge's *Small World* "Will I shame you in front of your *fine* friends?"). The diagnostic power of semantic prosody in identifying irony is discussed also by Partington (1998, 2007, 2017).

In the second case, semantic prosody can reflect the attitude and evaluation that a writer or a speaker assigns to the lexical items they are relying on and it can function as a means of persuasion (Louw 1993). This point is further stressed, among others, by Sinclair and Stubbs.

According to Sinclair (1996: 87), in fact, this collocational phenomenon is "used to express the speaker's approval (good prosody) or disapproval (bad prosody) of whatever topic is momentarily the object of discourse". Sinclair (1991: 112) states that semantic prosody tracks a word's or a phrase's "tendency to occur in a certain semantic environment", as is the case with the verb *happen* that "is associated with unpleasant things" (see also Sinclair 1987).

Hunston (2007) comments on Sinclair's idea of semantic prosody. She points out that, in terms of the number of words to be considered when sketching a word's or phrase's semantic prosody, Sinclair starts by researching on a single term and then he enlarges the horizon of his studies to longer patterns clustering around the node and forming units of meaning whose identity is filled by semantic prosody. In terms of the confidence with which semantic prosody can be associated with a word, Hunston (2007) observes that Sinclair finds it difficult to clearly characterise prosody as positive or negative because of the many variables co-occurring in context; nevertheless, once semantic prosody has been agreed upon, it will hold steady in that precise context, being clearly define as positive or negative (see also Hunston 2004, 2010).

Stubbs (1996: 173) adds that semantic prosody is a “prosody in its Firthian sense of a feature which extends over more than one unit” and which covers a span of words. This feature consists in the tendency of words to attract collocates characterised by a precise polarity value. In other terms, a word can fall into a collocational pattern characterised by positive connotative content (e.g. *provide*, which collocates with terms like *care, money, opportunities, etc.*; Stubbs 1996: 174) or by negative prosody (as with *cause*, co-occurring with the negatively connotated *accident, concern, damage, death, trouble, etc.*; Stubbs 1995a: 3); a term can also display neutral prosody, whenever the word-forms it collocates with do not place themselves clearly either on the positive or on the negative side of attitudinal stances (Stubbs 1996).

Also in Stubbs (1995a), semantic prosody can aid expressing and identifying authorial stance. According to Partington (2015: 299), however, Stubbs’s definition of semantic prosody is “narrower than that of Sinclair”. Stubbs believes that semantic prosody resides in the subtle connotational meaning of words. Sinclair, on the contrary, thinks that prosody can be drawn from items holding an obvious and a less obvious evaluative strength; in other terms, it can be found in most of the lexicon (Partington 2015).

The breadth that semantic prosody can be assigned in linguistic research raises the issue of the relationship between semantic preference and semantic prosody. According to Partington (2004; but see also Morley and Partington 2009), the relation between them is two-fold when it comes to the identity of their constituents and to the nature of their interaction. In terms of identity of their constituents, the phenomena seem to be knitted together by the same collocational items. However, while the former is more narrowly tied to the idea of closely co-occurring semantic sets, the latter stretches over a larger section of text to reach different types of elements. Secondly, as far as interaction is concerned, semantic preference powerfully acts in structuring semantic prosody, which, in response, binds the linguistic environment of the former by imposing privileged semantic connections (Partington 2004).

Bringing the discussion one step further, Hunston (2007) reflects on the stability of semantic preference and semantic prosody. She states that “a word which is used in a certain way in most contexts is not necessarily used in that way in all contexts” and she assumes that the semantic preference and the semantic prosody of a lexical item need to be contextualised to every type of discourse (Hunston 2007: 252-253). She also adds that, whenever a word can hold multiple collocational patterns, a speaker or a writer immediately singles the suitable one out by relying on the discursive

environment that the collocation is employed in. Hunston (2007: 263), in fact, believes that every discourse comes along with specific lexical habits.

Hunston (2007) also deals with the confidence with which a specific semantic prosody can be associated with a word and she claims that while for Sinclair the recognition of semantic prosody is clear-cut, Partington (2004) uses a gradable scale that assigns prosody some place within a negative-positive continuum: a word's semantic prosody will find a place in the positive-negative continuum depending on the frequency with which it occurs in positive, negative, or neutral contexts (Hunston 2007).

Because of semantic prosody's extending "over more than one unit in a linear string", Stubbs (2001: 65) renames it as *discourse prosody*. Discourse prosody mirrors the tendency of a positive, neutral, or negative semantic connotation to stretch over a larger textual unit, beyond the boundaries of a collocational pattern or also of an entire sentence. Discourse prosody expresses a writer's or a speaker's attitude in a larger, discursive context.

Semantic prosody and discourse prosody are only two of the several ways in which this phenomenon of connotation stretching over units of a sentence has been called in literature. Partington (2015), for instance, introduces the expression *evaluative prosody* as a valid substitute for the almost equal concepts of semantic prosody (Louw 1993 and Sinclair 1987), discourse prosody (Stubbs 2001), and *emotive prosody* (Bublitz 2003). Partington (2015: 279) justifies the choice of the label in the following way: evaluative prosody can be regarded as a form of prosody since the term, "borrowed from phonology, is used to describe a language phenomenon expressed over more than a single linguistic unit"; this prosody is evaluative in that it refers to the "indication of whether the speaker thinks that something (a person, thing, action, event, situation, idea, etc.) is good or bad" (after Thompson 1996: 65). In the present research, the expression *evaluative prosody* is preferred over the others.

According to Partington (2015), one of the main goals of evaluative prosody is persuasion. Through it, in fact,

speakers/writers endeavour to convince an audience of what should be seen as right and proper and what not, and in this way persuade the audience to think and conduct itself in an appropriate manner. Thus speakers/writers constantly both communicate their own evaluative attitudes but can also seek to impose, overtly or covertly, particular values and stances. (Partington 2015: 280)

Evaluative prosody functions thus to persuade, but also to express opinion, to construct and maintain relations, and to organize a discourse. It expresses the opinion of a speaker or of a writer when it reflects “the value system of that person and their community”; it constructs and maintains relations “by negotiating both ideological *alignment* of viewpoint regarding the various topics being evaluated [...] and also interpersonal *affiliation* between the participants themselves”; it persuades “by convincing an interlocutor that one’s evaluations are correct”, and it organizes discourse by being a form of textual cohesion (Partington 2015: 280-281).

Evaluative prosody can easily be traced by studying a word’s collocates, although sometimes it is not confined to a word’s collocational patterns but it extends to a whole stretch of discourse. In the first case, it can be regarded as a lexical form of evaluative prosody, in the second as a textual kind of prosody. The lexical description of evaluative prosody “is used to describe a node’s inherent potential to participate in evaluative interaction with other items of similar polarity” while the textual description of evaluative prosody “is used to describe the interaction of the item with others of similar polarity as witnessed within a specific text” (Partington 2015: 283). Evaluative prosody’s description is psychological “if it can refer to a community of speakers’ shared intuitive knowledge of how to use a particular item in conjunction with others of similar polarity in order to maintain evaluative harmony when speaking or writing” and it is statistical when “the community of speakers has acquired this shared knowledge by repeatedly encountering an item in co-occurrence with other items of a certain evaluative polarity” (Partington 2015: 284).

Evaluative prosody is generally consciously or unconsciously chosen with consistency throughout a discourse in order to guarantee evaluative harmony and it is characterised by nine properties (after Forest 2007; in Partington 2015).

1. First, it is generally the result of collocational patterns: this means that evaluative prosody can be studied through the analysis of collocational patterns and that it shares with collocation the repetition of peculiar lexical patterns in specific contexts.
2. Second, evaluative prosody is generally beheld by speakers and writers as a form of implicit knowledge of language, or, in other terms, as a form of priming.
3. Third, it can be understood as a property of the word or it can be gathered from the spreading of a text throughout a discourse. When it is read in a discursive way, it can be merely characterised as a property of a single text

or it can be viewed in more statistical terms, functioning thus well within a corpus-assisted analytical framework (on corpus approaches to discourse see § 1.3). Such a corpus-assisted definition of evaluative prosody

entails analysing, via a concordance, how a node is actually instantiated many times in many texts. We are thus able to perceive both the various patterns of co-occurrence, that is, the items the speakers/writers have generally chosen to co-select, and also which semantic prosody polarity predominates, is simply numerically more frequent (in the discourse type under scrutiny) and is therefore the bedrock evaluating function the item in question performs in these texts in *general*. (Partington 2015: 293)

4. Fourth, evaluative prosody is strongly intertwined with semantic preference: semantic preference posits the basis for evaluative prosody by offering a sequence of semantically related collocates; at the same time, evaluative prosody constructs a general evaluative environment that constrains the choice of the lexical items that will collocate with the node.
5. Fifth, evaluative prosody can be spotted synchronically: it can take shape through the repetition of evaluatively loaded collocational patterns in a specific time span.
6. Sixth, it can also be observed diachronically. While this kind of evaluative prosody is more difficult to be gathered empirically, when it appears, it might contribute to semantic change.
7. Seventh, sometimes evaluative prosody is very commonly found in the unfolding of discourse.
8. Eight, some other times it is limited to very few lexical items.
9. Finally, evaluative prosody is not stable in language, but it can change according to syntactic variants, after discourse types, or when adopted for a metaphorical use.

In sum, like Stubbs's (2001) discourse prosody, also evaluative prosody covers a wider text span in comparison to semantic prosody. It can be said to originate from the repeated co-occurrence of lexical items (collocation), to assume a consistent grammatical pattern via the grammatical relation binding the co-occurring lexical items (colligation), to keep together lexical items belonging to specific but open-ended semantic areas (semantic preference), to draw from their content and to constrain their evaluative power (semantic prosody). Evaluative prosody, however, spans

throughout a whole discourse and it imbues it with connotative value. It functions to highlight stance, to persuade, and to provide discourse with evaluative harmony. In addition, evaluative prosody can be schematized via different theories of evaluation (Partington 2015).

Summing up, evaluative prosody encompasses all levels of analysis of collocational patterns, gathering its essence from collocation proper, colligation and semantic preference. Thus, together with collocation, colligation and semantic preference, it contributes to the nuanced meaning by collocation of a lexical item.

### **1.3. Corpus approaches to ecological discourse studies across languages**

In the present research, cultural keywords and meaning by collocation are studied within the theoretical framework of corpus approaches to the critical study of discourse. Corpus approaches to the critical study of discourse develop since the mid-1990s as a reaction to the limits and flaws of traditional discourse analysis, which are highlighted among others by Beaugrande (1999, 2001), Hardt-Mautner (1995) and Stubbs (1997). Corpus approaches to the critical study of discourse commit to match the strengths of critical discourse studies with the potentiality of corpus linguistics in a “fruitful synergy” (Baker et al. 2008: 273). From critical discourse studies they inherit a critical attitude to the study of discourse by recommending a qualitative, fine-grained exploration of a corpus so that ideological stances can be spot across the lines of a corpus. From corpus linguistics they adopt the use of electronically stored collections of data to add quantitative empirical evidence to qualitative analyses of the texts.

In order to better understand the principles of corpus approaches to the critical study of discourse, the tenets of critical discourse studies and corpus linguistics are now described.

#### **1.3.1. Critical discourse studies**

Critical discourse studies originate in the mid-1980s as an analytical approach for critically researching on discourse; at the beginning, Norman Fairclough labels this approach *critical discourse analysis* (Fairclough 1985). Critical discourse analysis, however, officially begins only in 1992, when Teun van Dijk, Norman Fairclough, Gunther Kress, Theo van Leeuwen, and Ruth Wodak meet in Amsterdam to discuss

about their critical studies. On that occasion, they come up with an agenda that sets the ground for the following critical discourse analytical research (Leeuwen 2006; see also Breeze 2013 and Wodak and Meyer 2016 for an account of the history of critical discourse studies).

The roots of critical discourse analysis can be traced back to *critical linguistics*, a movement founded in the 1970s at the University of East Anglia and represented, among others, by scholars like Roger Fowler (e.g. Fowler et al. 1979, Fowler 1981, 1996a, 1996b), Robert Hodge and Gunter Kress (Hodge and Kress 1979). According to Leeuwen (2006: 291), critical linguistics is a linguistic theory that stems from Michael Halliday's systemic-functional linguistics but that wishes "to move linguistic analysts beyond formal description and use it as basis for social critique".

This critical approach to the analysis of language is inherited by critical discourse analysis (quite recently renamed critical discourse studies; Dijk 2009). Critical discourse studies detach from critical linguistics because of their expanding the scope of their research from language to discourse, and from discourse to society. In fact, according to Fairclough (1992: 62), critical discourse studies aim at bringing together "linguistically-oriented discourse analysis and social and political thought relevant to discourse and language, in the form of a framework which will be suitable for use in social scientific research". Therefore, while critical linguistics is more language-oriented, critical discourse studies assign social theory a pivotal role and they focus on the way discourse relates to society (Fairclough 1992). In relation to this, van Dijk (2009: 111) states that "critical discourse analysis is problem-oriented" because it "does not primarily focus on discourse and its properties, but on social issues and problems".

This focus on social issues and problems results from the strong political commitment of critical discourse analysts. According to van Leeuwen (2006: 293), critical discourse analysts "make their position explicit and feel they do not need to apologize for the critical stance of their work". Such a political commitment engages scholars in critically working on the discursive construction of highly debated and socially meaningful topics like racism (e.g. Dijk 1991, 2005), national identity (e.g. Kovács and Wodak 2003), feminism (e.g. Lazar 2005; see, among others, Caldas-Coulthard and Coulthard 1996 for a classic collection of other topics touched upon by critical discourse studies), etc.

According to Wodak and Meyer (2016: 7), this politically committed work can be deemed critical to two extents. The first comes from the Frankfurt School and Jürgen Habermas: according to them, the social theory that guides critical discourse studies

“should be oriented toward critiquing and changing society as a whole, in contrast to traditional theory oriented solely to understanding or explaining it”. The second extent is more proactive: critical discourse studies want to “produce and convey critical knowledge that enables human beings to emancipate themselves from forms of domination through self-reflection” (Wodak and Meyer 2016: 7).

This critical knowledge needs to be obtained with an interdisciplinary attitude whereby the analyst becomes aware of the interconnections tying society, politics, and economy. Because of this, it requires “specific ethical standards” from the part of the researcher (Wodak and Meyer 2016: 7).

As it has already been pointed out, critical discourse studies operate their critical enterprise on discourse. The notion of discourse is a highly debated one in critical discourse studies. Some of the most influential practitioners regard it as “a form of ‘social practice’” that comes out of “a dialectical relationship between a particular discursive event and the situation(s), institution(s) and social structure(s), which frame it” (Fairclough and Wodak 1997: 258). According to Fairclough and Wodak (1997), the discursive events are influenced and moulded by situations, institutions, and social structures but at the same time they influence and mould them back. This definition of discourse is used also in the present study.

In critical discourse studies, the dialectical relationship that links discourse on the one hand and situations, institutions, and social structures on the other can be described and interpreted under the light of various social and political theories, among which most famously Michel Foucault’s (e.g. Foucault 1966, 1981; about this see, among others, Dijk 2009). Through the critical study of the dialectical relationship lying between discourse and situations, institutions, and social structures, researchers wish to uncover the way *ideology, power, and hegemony* are concealed through language (e.g. Fairclough 1989, Wodak 1989).

The definitions and the theories of ideology, power, and hegemony vary considerably throughout critical discourse studies. Ideology is usually intended as

the basis of the social representations shared by members of a group.  
[...] ideologies allow people, as group members, to organize the multitude of social beliefs about what is the case, good or bad, right or wrong, for them, and to act accordingly (Dijk 1998: 8).

Ideologies are intertwined with power

because the nature of the ideological assumptions embedded in particular conventions, and so the nature of those conventions themselves, depends on the power relations which underlie the conventions; and because they are a means of legitimizing existing social relations and differences of power, simply through the recurrence of ordinary, familiar ways of behaving which take these relations and power differences for granted. (Fairclough 1989: 2)

According to critical discourse analysts, the distribution of power in society is usually unfair and this disparity results in the hegemony of a group over another. Ideology, power, and hegemony are reflected and reproduced through language (Fairclough 1989).

Because of this, Wodak (2001: 2) points out that critical discourse studies are interested in “analysing opaque as well as transparent structural relationships of dominance, discrimination, power and control as manifested in language”. In other terms, they aim “to investigate critically social inequality as it is expressed, signalled, constituted, legitimized and so on by language use (or in discourse)”. To this extent, discourse plays “a key role in maintaining and legitimating inequality, injustice, and oppression in society” and critical discourse analysts wish to “spread awareness of this aspect of language use in society, and to argue explicitly for change on the basis of its findings” (Leeuwen 2006: 290).

According to Fairclough (1995), the opaque or transparent reproduction of ideology, power, and hegemony through language is often the result of a process of *naturalisation* (Fairclough 1995). Naturalised topics and naturalised linguistic patterns are themes and forms that have established as “common sense”; however, they conceal an ideological stance of power and hegemony (Fairclough 1995: 27). As a consequence, critical discourse studies need to identify naturalisation and expose it so that it can be challenged by the very analysts and by those who read their work (Fairclough 1995).

This scholarly work is often carried out under the light of Antonio Gramsci’s reflections on the way power and hegemony can be established and perpetuated in discourse (e.g. Gramsci 2009), through Pierre Bourdieu’s theory of the system of meaning that dominators and dominated accept when interacting linguistically (e.g. Bourdieu 1983, 1988), and through Habermas’s idea of the influence of language in creating social systems (e.g. Habermas 1971; in Breeze 2013).

Even though critical discourse studies share these theoretical premises, they frequently diverge on the approach they adopt for the analysis of discourse, on the data they scrutinize, and on the methodological choices they rely on.

As far as the data used for the studies is concerned, Meyer (2001) observes that critical discourse analysts usually collect texts that serve their *ad hoc* purposes. They do not analyse randomly selected data, but they ponder on their research needs and then they look for texts that could satisfy them in the best possible way. First, analysts collect a provisional set of texts, they peruse them listing the thorny issues that transpire, and then they pick new texts that might corroborate their initial hypothesis. As Meyer (2001) notes, they face little consultation on the statistical and theoretical representativeness of the textual collections they employ for their analyses. They prefer distinctiveness of a discourse to representativeness of the same discourse.

In terms of methodology, critical discourse studies employ a wide range of methodological choices that suit their different approaches. Some practitioners, for example, draw their methodology from *social actors* theory (e.g. Leeuwen 2008) while others adopt a more grammatically oriented approach (e.g. Fairclough 2003). This diversification implies that no uniform methodological framework and no unitary list of analytical tools can be produced for critical discourse studies (Meyer 2001).

Furthermore, and reaching back to the theoretical background of the discipline, general methodological trends suggest that text and discourse should not be treated as a monolithic whole but that they ought to be explored in search for the connections arising between texts, discourses, and society. These links can be studied, for example, under the light of the concept of *intertextuality* as it is meant by Fairclough (1992, 1995).

According to Fairclough (1992: 270), intertextuality “points to the productivity of texts, to how texts can transform prior texts and restructure existing conventions (genres, discourses) to generate new ones”. The transformation and restructuring of texts, however “is socially limited and constrained, and conditional upon relations of power” (Fairclough 1992: 270). Intertextuality helps to understand the peculiarities of discourse *per se* but it also expands on discourse change (Chouliaraki and Fairclough 1999).

Studies of this kind require “interdisciplinary work in order to gain a proper understanding of how language functions in constituting and transmitting knowledge, in organizing social institutions or in exercising power” (Wodak and Meyer 2016: 7). The analysis of language needs thus to be enriched with an overview of the complementary semiotic resources that assist language in producing meaning. Interdisciplinarity can characterise the “theoretical framework”, in case the theory that shapes a work draws from different research areas or fields, but it can also regard the “team research and [...] the collection and analysis of data” (Wodak and Meyer 2016: 19).

Research conducted within the framework of critical discourse studies is usually qualitative (Meyer 2001). This means that its quality needs to be preserved and assessed more carefully than in the case of quantitative studies of discourse (on qualitative research see, for instance, Silverman 2004, 2014, Silverman and Marvasti 2008, and Wodak and Krzyżanowski 2008).

As far as approaches are concerned, some of the most influential are the *discourse-historical approach*, the *social actors approach*, *dispositive analysis*, the *socio-cognitive approach*, and the *dialectical-relational approach* (in Wodak and Meyer 2016). In addition, Wodak and Meyer (2016) hint at approaches that started within the realm of critical discourse studies but that later developed independently. One of these is *multimodal analysis* (Kress and Leeuwen 1996, 2001; Machin 2016; Machin and Leeuwen 2007). Although the final goal of multimodal analysis is similar to critical discourse studies', multimodal analysis considers both text and non-textual semiotic devices as systems of meaning in discourse. It complements the study of the textual component of discourse with an exploration of the audio-visual material that comes with text. Practitioners of multimodal analysis believe, in fact, that also multimodal semiotic resources are governed by grammatical patterns.

Also Flowerdew and Richardson (2017a) provide a snapshot of the multiplicity of approaches to critical discourse studies: one of these is the ecological discourse analysis approach (Stibbe 2017; Flowerdew and Richardson 2017a).

### **1.3.2. Ecological discourse analysis**

Ecological analysis of discourse aims at carrying out a form of critical discourse study whereby “discourses are analysed within an ecological framework which considers the impact of the discourses on the systems which support life” (Alexander and Stibbe 2014: 109). The relationship between ecological analysis of discourse and critical discourse studies can be described in the following way:

since the rise of Critical Discourse Analysis (CDA), eco-criticism has been seen by many as an extension and further development of this movement. However, there is a fundamental difference: while CDA looks mainly at discourse, i.e. ‘parole’, and thus criticizes the words, the syntax and the pragmatics of spoken and written texts, ecolinguists now critically explore the language *system* as well (‘langue’), which in many instances favours an unecological fragmentation and a

separation of humans for the rest of animate and inanimate beings. (Fill and Mühlhäusler 2001: 6)

Ecological discourse analysis belongs to the field of *ecolinguistics*. Ecolinguistics can be defined as a discipline that “combines ecology and linguistics” (Alexander and Stibbe 2014: 104), with a wide variety of aims and through a wide range of methods (Steffensen and Fill 2014). Broadly speaking, it can be intended either as a study of “the relation between language and environment” or as research on “linguistic diversity” (Fill and Mühlhäusler 2001: 3). The current study focuses on ecolinguistics as research on the relation between language and environment.

As far as the relationship between language and environment is concerned, ecolinguistics is based on the idea that language impacts “on the life-sustaining relationships among humans, other organisms and the physical environment” and that consequently an ecological form of linguistics “is normatively orientated towards preserving relationships which sustain life” (Alexander and Stibbe 2014: 105). This approach originates from the Firthian and then Hallidayan belief that if languages are meant as social systems, they portray and stimulate the way people perceive and act in the world (Alexander and Stibbe 2014).

To this extent, ecolinguistics focuses on the impact of grammar for the representation and construction of beneficial and detrimental relationships among beings and between beings and the environment. Halliday (1992: 86), for instance, observes that a grammar “is a theory of experience; a theory that is born of action, and therefore serves as a guide to action, as a metalanguage by which we live by”; in English, the “grammar makes it hard for us to accept the planet earth as a living entity that not only breathes but feels and even thinks”. This idea is stressed also by Goatly (1996: 537), who believes that “ordinary language, especially the transitive clause, is inadequate to the representation of the world demanded by modern scientific theory, especially ecological theory” and that grammatical resources should be changed and used in a way that satisfies these theories.

Thus, according to Fill and Mühlhäusler (2001: 6), ecological studies of discourse criticise “language use from the ecological and environmental point of view” by focusing on discourse. This ecological and environmental point of view involves the main tenets of various ecological frameworks such as *sustainable development* (e.g. Baker 2006, Meadowcroft 2005), *deep ecology* (e.g. Devall and Sessions 1985, Naess 1995, 2005), *social ecology* (e.g. Bookchin 1989, 1995, 2005), etc. (see also Peterson del Mar 2012 for more on ecological frameworks).

In ecological analyses of discourse, the main tenets of ecological frameworks are shaped in the form of personal ecological frameworks labelled *ecosophies* by Naess (1995). An ecosophy is “a philosophy of ecological harmony” that “contains norms, rules, postulates, value priority announcements *and* hypotheses concerning the state of affairs” (Naess 1995: 8; in Stibbe 2015: 12). Ecosophies are personal and they draw from the most common ecological theories. Thanks to them, ecological discourse analysis is not simply a form of discourse analysis. In fact, it encourages to interpret the results of a research under the light of a precise and personal ecological framework (Alexander and Stibbe 2014).

Ecological discourse analysis operates in linguistics but it grounds its interpretive power in ecological frameworks. From ecology it gathers an interest in the study of “the relationships of humans with other humans, other organisms, and the physical environment” (Alexander and Stibbe 2014: 104). From linguistics it inherits a devotion to the study of language, which “is relevant to the extent that it plays a role in how humans relate to each other, to other organisms and to the environment” (Alexander and Stibbe 2014: 104). In fact, according to Mühlhäusler (2003)

meaning arises through the involvement of speakers with other speakers within a shared context of situation, and is shaped by their expectations, and their understandings of the world. Very importantly, meaning needs to be understood as part of ongoing discourses, not as located in decontextualised chunks of language. [...] different groups can mean very different things when using apparently identical linguistic materials or when talking about aspects of the same environment and [...] within the same language community, different norms for generating and interpreting meaning exist. (Mühlhäusler 2003: 9)

Ecological discourse analysis is a field of linguistics that tries to keep together the expertise of several disciplines (such as linguistics, anthropology, semiotics, ecology, etc.) in order to describe the way that the relations occurring within and between the natural and the animal worlds are linguistically represented (Fill and Mühlhäusler 2001, Stibbe 2015).

These representations can be identified through linguistic means in discourse. Discourses can be classified as *destructive*, *ambivalent*, or *beneficial* according to the ecological framework that is accepted and adopted by the analyst (Stibbe 2015). In destructive discourses “the ideologies they convey oppose the principles of the ecosophy”; ambivalent discourses “contain some aspects which align with the

analyst's ecosophy and some which oppose it"; beneficial discourses "convey ideologies which can actively encourage people to protect the systems that support life" (Stibbe 2015: 28-30).

These discourses can be analysed through a wide variety of linguistic features. According to Stibbe (2015: 9), ecological discourse analysis subsumes elements of several linguistic theories:

These include Critical Discourse Analysis (Fairclough 2003); frame theory (Lakoff and Wehling 2012); metaphor theory (Müller 2008); appraisal theory (Martin and White 2005); identity theory (Benwell and Stokoe 2006); fact construction (Potter 1996) and theories of erasure and salience (drawing on van Leeuwen 2008).

All theories offer theoretical and methodological tools for the study of language from an ecological point of view. In the present study, the critical discourse studies approach is preferred over the others.

In ecological discourse analysis, the combination of ecological and linguistic study of discourse turns out in "the use of techniques of linguistic analysis to reveal the stories-we-live-by, opening them up to question and challenge from an ecological perspective" (Stibbe 2015: 9). The *stories-we-live-by* are defined by Stibbe's (2015: 6) as "stories in the minds of multiple individuals across a culture", with *stories* meaning "cognitive structures in the minds of individuals which influence how they perceive the world". Stibbe (2015) identifies eight kinds of stories: *ideology, framing, metaphor, evaluation, identity, conviction, erasure, and salience*.

According to Stibbe (2015: 17), framing consists in "a story that uses a frame (a packet of knowledge about an area of life) to structure another area of life", metaphor is a kind of framing and it can be described as "a story that uses a frame to structure a distinct and clearly different area of life", identity is "a story about what it means to be a particular kind of person", conviction labels "a story about whether a particular description of the world is true, uncertain or false".

Ideology is "a story of how the world is and should be which is shared by members of a group" (Stibbe 2015: 17). Ideology is both individual and social and it is conveyed in language through specific linguistic features. These linguistic features have been listed by critical discourse studies (e.g. Fairclough 2003, Martin and Rose 2003) and they encompass patterns that range from syntax (e.g. argument structure and transitivity, relationships within and between clauses, etc.), through semantics (e.g. the choice of lexis, relationships between words) to pragmatics and text linguistics

(e.g. presuppositions and assumptions, intertextuality, figures of speech, etc.). They can be identified in discourse by selecting textual material that suits the purposes of one's research goals and by analysing it with the methods suggested by critical discourse analysts. The interpretation of the way these linguistic features contribute to the discursive construction of ideology depends on the ecosophy of the researcher.

Evaluation is meant as "a story about whether an area of life is good or bad" (Stibbe 2015: 17). Evaluation characterises destructive, ambivalent, and beneficial discourses by positively or negatively connoting environmental dynamics that the researcher's ecosophy judges as good or bad. In ecological discourse analysis, evaluation is usually looked for in a sample of meaningful texts through the theoretical and methodological tools offered by appraisal theory (e.g. Martin and Rose 2003, Martin and White 2005, Salvi and Turnbull 2010). According to Martin and White (2005), appraisal is performed with the expression of engagement, affect, judgement, appreciation, and graduation, which are linguistically encoded in evaluative lexis, modal verbs, modal adjuncts, polarity, prenumeration, intensification, repetition, manner or extent, logico-semantics, and vocation. In the present work, I argue that also patterns of positively or negatively connoted and co-occurring lexemes can be adopted for the study of evaluation (i.e. evaluative prosody through collocational patterns; for more on collocates and evaluative prosody see § 1.2 in this chapter).

Erasure consists in "a story that an area of life is unimportant or unworthy of consideration" (Stibbe 2015: 17). According to Stibbe (2015: 146), erasure can be treated as an appraisal pattern, "except rather than appraising something as *bad*, it appraises it as *unimportant* and generally unworthy of consideration". This phenomenon happens on a scale and its gradual nature determines a more articulated characterisation of erasure in the forms of *void*, *mask*, and *trace*. The void is recognised "where 'something important' is completely excluded from a text"; the mask appears where an important item "is erased but replaced by a distorted version of itself"; the trace is found "where something is partially erased but still present" (Stibbe 2015: 149; after Baudrillard 1994). These forms of erasure are typically spot first in single sentences and then searched through wider sets of data. They are performed in discourse with linguistic features like passives, metonymy, nominalisations, and hyponyms, and they are frequently associated with theories like social actor theory (Leeuwen 2008). Erasure is also to be read in transitivity patterns, where the nature of the participants in a process tells much about their role in the event (Halliday 2004). In the present work, erasure is identified among economic, social and environmental lexemes to observe what areas of life are deemed less worthy of consideration.

Salience is “a story that an area of life is important and worthy of consideration” (Stibbe 2015: 17). Like erasure patterns, also salience patterns operate as appraisal tools: salience patterns connote something as important and worthy of consideration whereas evaluation would merely present it as good. Erasure and salience can be spot in discourse through symmetrical linguistic patterns. If erasure is performed through *passivation*, salience is ensured by *activation*; if erased elements appear in transitivity patterns that suppress them, salient elements feature in transitivity patterns that foreground them. In addition, salience can also be obtained through *individualisation*, another concept drawn from social actor theory. In social actor theory, individualisation is described as the process whereby a social actor is introduced through their identity, by “representing individuals by name” (Fairclough 2003: 150; see also Leeuwen 2008). It is opposed to *impersonalisation*, that is obtained whenever a social actor is deprived of their specific identity (Fairclough 2003). Moreover, actors can be endowed with or deprived of salience also through abstraction or basic-level representations (Stibbe 2015). In this work, I maintain that salience can be traced also among cultural keywords (see § 1.1 in this chapter).

Being ecological discourse analysis strongly intertwined with critical discourse studies, it has suffered from the same criticisms that have been moved to critical discourse studies. Since the mid-1990s critical discourse studies have been criticised by some scholars working in the field of discourse analysis (e.g. Hardt-Mautner 1995, Stubbs 1997, Widdowson 1998, 2004). Among these, Stubbs (1997) negatively judges critical discourse studies from three points of view: the relationship they assign to discourse and society, data collection, and methodology (Stubbs 1997).

As far as the relationship between discourse and social practice is concerned, Stubbs (1997) wonders how discourse can be interpreted in relation to its producers, its receivers, and the social practices that they participate in. He finds it impossible for analysts to reach such an interpretation without positing any clear-cut criteria as to how the participants of discourse production and reception socially engage in discourse (Stubbs 1997).

As regards data collection, Stubbs (1997) criticises research in critical discourse studies for handling small amounts of data. Critical analysts do not look for much evidence, but they restrict their researches to a limited selection of texts. These texts are not always considered in their complete form and they are frequently excerpts of longer communicative exchanges. Furthermore, critical discourse analysts seem to dismiss the issue of text sampling, i.e. the way fragments should be isolated to save the representativeness of the discourse under inquiry. They do not seem to treat

representativeness in general either and they only collect those texts that seem useful for their research purposes (Stubbs 1997).

As far as methodology is concerned, Stubbs (1997) notices that critical discourse studies heavily suffer from circularity. They are frequently conducted with a political rather than with a linguistic focus. As a consequence, many practitioners strive to find data that could confirm their political ideas rather than challenge it. This results in confusion also when it comes to the analytical tools to be employed to study discourse. In fact, different authors trust diverse concepts, that are encoded in different linguistic patterns. What is more, they do not clarify the relationship existing between these linguistic patterns and the way power and ideology are reproduced in discourse. In other terms, they do not yield precise guidelines for the critical interpretation of linguistic structures (Stubbs 1997).

In addition to Stubbs's (1997) observations, Breeze (2013) comments on other aspects of critical discourse studies. The first is the political commitment of critical discourse studies, which Breeze (2013) does not always find robust and clearly stated. Thus, she recommends that the political stance used for the interpretation of textual observations should be systematised and explicitly dealt with by the analyst when analysing discourse so that the reader of the research can assess the validity of the scholar's interpretations.

Then, Breeze (2013) notices that critical discourse studies rely on a wealth of theoretical and methodological approaches that sometimes can be incompatible. Furthermore, the critical discourse analysts' methodological choices are sometimes blurred and applied without systematicity to the study of discourse (Breeze 2013). Thus, studies might result in a mere and shallow interpretation of text and discourse without any real linguistic analysis. Moreover, this "lack of interest in the epistemological and hermeneutic dimensions of textual analysis is matched by a corresponding over-emphasis on the theoretical dimension of explanation" (Breeze 2013: 207; see also Verschueren 2001 on this). Breeze (2013) suggests clarifying the theoretical and methodological framework adopted and being consistent in formalising and using it when analysing discourse.

Other issues raised by Breeze (2013) relate to micro- and macro-contexts of discursive analysis. At a micro-contextual level of discursive analysis, scholars tend to hint solely at the description of linguistic patterns, providing few data to the reader, and jumping quickly to the explanation and interpretation of the specimen of discourse that the linguistic patterns were extracted from. From a macro-contextual level of analysis, social context is not always taken into account when explaining and

interpreting discourse. Discourse, in fact, is frequently extrapolated from its original sources and it is studied in isolation. Moreover, once again in relation to the macro-context, Breeze (2013) notes that the study of the reception of discourse by readers or listeners is frequently neglected. Although critical discourse analysts affirm that they study the relationship between discourse, society and cognition, they actually disregard the way discourse is perceived and the way it affects its recipients. This is included, on the contrary, in research in media studies and ethnography of communication (e.g. Nightingale 2011, Ross and Nightingale 2003; in Breeze 2013).

Finally, according to Breeze (2013: 521) critical discourse studies have “mainly researched the way ideology works through discourse to maintain unequal power structures” and they have acted so with a negative attitude. Breeze (2013) suggests turning to a more positive attitude to society. According to her, researchers should consider discourse that purports beneficial dynamics for social systems in addition to discriminatory discourse. Put it another way, Breeze (2013) supports the backing of critical discourse analysis by positive discourse analysis (on this see, among others, Bartlett 2012, Macgilchrist 2007, Martin 2004; in Breeze 2013).

This encouragement to mould critical discourse studies into positive discourse analyses is supported also by ecological analyses of discourse (Stibbe 2018). Positive discourse analysis, in fact, is welcomed within ecolinguistics to stimulate research on linguistic choices that prove beneficial for the environment according to the ecosophy of the analyst, and that should therefore be encouraged. Within ecological discourse analysis, positive discourse analysis tries to expose linguistic patterns that reflect positive relations between the members of a society and, more broadly, between the living or natural inhabitants of the world (Stibbe 2018).

Both critical discourse studies and ecological discourse analysis can profit from the same advice that critics like Stubbs (1997) and Breeze (2013) propose for critical discourse studies. The most noteworthy recommendation made by Stubbs (1997) is that critical discourse studies should adopt corpus approaches for their analyses (Stubbs 1997). The same is claimed immediately after also by Beaugrande (1999, 2001) and later on by a manifold of researchers including most notably Breeze (2013) but also Baker (2006), Baker et al. (2008), Gabrielatos and Baker (2008), Hardt-Mautner (1995), Mautner (2007, 2009, 2016), Partington et al. (2013), Stubbs (2001), and Marchi and Taylor (2018).

Corpus approaches to the critical and ecological study of discourse, in fact, offer methodologies and tools that strengthen the outcome of critical and ecological

discourse studies by allowing the collection of more solid data and results. To do so, they rely on the methodologies and tools of corpus linguistics.

### 1.3.3. Corpus linguistics

Corpus linguistics is a computer-assisted “set of methods and procedures for the exploration of language” (McEnery and Hardie 2012: 1) that makes repeated linguistic patterns “visible by a combination of quantitative and categorial devices” (Teubert 2007: 131).

According to Leech (1992: 105) corpus linguistics can be regarded not as “a domain of study” but rather as “a methodological basis for pursuing linguistic research”. It “should be considered as a methodology with a wide range of applications across many areas and theories of linguistics” and not as

an independent branch of linguistics in the same sense as phonetics, syntax, semantics or pragmatics. These latter areas of linguistics describe, or explain, a certain aspect of language use. Corpus linguistics, in contrast, is not restricted to a particular aspect of language. Rather, it can be employed to explore almost any area of linguistic research (McEnery et al. 2006: 7-8).

Corpus linguistics is based on “bodies of text as the domain of study and as the source of evidence for linguistic description and argumentation” (Kennedy 1998: 7). These bodies of text are called corpora.

A corpus is usually defined as a “a body of text made available in computer-readable form for purposes of linguistic analysis” (Meyer 2002: xii). These computer-readable texts can be both “written or spoken” (McEnery et al. 2006: 4) and they are chosen so as to be “naturally occurring” (Hunston 2002: 2; but see also Tognini-Bonelli 2001: 2), “representative of a given language” (Tognini-Bonelli 2001: 2) and potentially “annotated with various forms of linguistic information” (McEnery et al. 2006: 4). According to McEnery et al. (2006: 4) a corpus is thus basically characterised as a set of texts which are “machine-readable”, “authentic”, “sampled”, and “representative”.

The texts included in corpora are chosen in a way that makes them an “appropriate basis on which to study a specific set of research questions” (McEnery and Hardie: 1). In other terms, the nature of the texts selected for a corpus depends on the aims of the research they are used for. This influences “the design, size and nature

of the individual corpus” and it introduces a high variability in the most peculiar features of the text collection (Kennedy 1998: 3).

### 1.3.3.1. Corpus classification

This high variability results in the rise of a wide range of corpus types, which serve various kinds of purposes. Some of these corpus types are described in the following paragraphs by grouping them according to their peculiarities.

*Written vs. spoken corpora.* Corpora can include examples of *written* or *spoken* language (McEnery and Hardie 2012). In addition, they can collect *multimodal* texts (e.g. Knight et al. 2009).

*General vs. specialised corpora.* Corpora can be either *general* or *specialised*. General corpora are also called *reference corpora*, and they are collections “of texts of many types” (Hunston 2002: 14) that “can be taken as representative of the language as a whole” (Tognini-Bonelli 2001: 9) and therefore they “make available a text base for unspecified linguistic research” (Kennedy 1998: 19). Specialised corpora, on the other hand, “are designed with particular research projects in mind” (Kennedy 1998: 20) and consequently they consist “of texts of a particular type” which they try to be representative of (Hunston 2002: 14).

Specialised corpora are collections that aim “to be representative of a given type of text”, in contrast with general corpora that try to reproduce the general characteristics of a language (Hunston 2002: 14). As a consequence, their texts “tend to be domain (e.g. medicine or law) or genre (e.g. newspaper text or academic prose) specific” (McEnery et al. 2006: 15). According to Hunston (2002), specialisation can be very sophisticated in this kind of corpora on condition that the texts comprised in the collection are chosen according to a carefully weighed set of parameters (Hunston 2002).

General or specialised corpora can be either close-ended or open-ended. In other terms, they can be either *static*, permanently consisting of the textual data collected when building the corpus, or they can allow for expansion. In this case they are called *monitor corpora* or *dynamic corpora* and they “develop a dataset which grows in size over time and which contains a variety of materials” (McEnery and Hardie 2012: 6). The monitor corpus “is added to annually, monthly or even daily” and “the proportion of text types in the corpus remains constant, so that each year (or month or

day) is directly comparable with every other” (Hunston 2002: 16). This pondered growing size is planned “to track current changes in a language” (Hunston 2002: 16).

*Synchronic vs. diachronic corpora.* The same function is also shared by *diachronic corpora*. Diachronic corpora are collections of texts that represent “a language over a period of time” and they are used to follow the development of linguistic features over that period of time (Kennedy 1998: 22). Diachronic corpora are opposed to *synchronic corpora*, namely collections that “contain texts created within a relatively narrow time-frame” (Meyer 2002: 45) and that try “to represent a language or a text type at a particular time” (Kennedy 1998: 22).

*Monolingual vs. multilingual corpora.* Another opposition distinguishes *monolingual corpora*, i.e. corpora that “contain texts in only one particular language”, from *multilingual corpora*, i.e. corpora that “contain texts in several different languages” (McEnery and Wilson 1996: 57).

Multilingual corpora are described by Johansson (2007: 9) as collections of texts “in two or more languages put together in a principled way for the purpose of comparative linguistic studies and prepared in electronic form for search and analysis by computer”. Corpora that include texts in more than a language are generally labelled multilingual corpora although a distinction can be suggested relative to the actual number of languages represented: corpora that feature only two languages can be properly said bilingual while collections of more than two languages are truly multilingual (Altenberg and Granger 2002b).

*Comparable vs. parallel corpora.* Multilingual corpora can be further classified in *parallel corpora* and *comparable corpora* according to the nature of their texts. Comparable corpora

consist of original texts in each language, matched by such criteria as time of composition, domain, genre, intended audience, etc. They represent natural language use in each language and should allow safe conclusions to be drawn on similarities and differences between the languages compared. (Johansson 1998: 5)

Comparable corpora gather texts of a similar kind but produced in different languages so that they can be contrasted in search for equivalences or correspondences. In particular, Aijmer and Altenberg (1996) state that comparable corpora can be of use to

spotlight features of a language that could go unnoticed when exploring monolingual corpora or to shed light on typological and cultural peculiarities of the varieties under inquiry.

According to Aijmer (2009), comparable corpora are relatively easy to build, especially when vast repositories of texts are already available. In addition, if the corpus is big enough or balanced enough to a language's genres, results drawn from its analysis can be extended to the whole language. However, since genres and other external criteria on which corpus builders rely when designing their collections are not always widespread cross-linguistically, true comparability of corpora is frequently at stake. Comparability is threatened further by legal issues: copyright, in fact, infringes sometimes the scholar's desire to collect all the material that might advance the corpus towards better comparability (Aijmer 2009). As a consequence, legal issues are seriously taken into account when building a corpus (McEnery and Hardie 2012; but see also Allora and Barbera 2007).

Comparability of the corpus's sections can be dealt with thanks to well-pondered and well-documented corpus design though: if a corpus developer reports the insightful criteria that they have exploited for the selection of the corpus's texts and if they try to stick to them for all languages involved in the collection, comparability is not at stake. This holds true especially when the corpus can be regarded of a special kind, namely if it is a specialised corpus. Comparable corpora, in fact, can be both general and specialised (McEnery et al. 2006). The same holds true also for parallel corpora.

Parallel corpora contain "original texts in one language and their translations into one or several other languages"; if the translations go from one language to the others, corpora are called unidirectional whereas if they also go the other way round they are labelled bidirectional (Altenberg and Granger 2002b: 8). Parallel corpora are "composed by real-world texts from which data on in-context translation equivalents can be extracted"; on the other hand, comparable corpora "can be processed to obtain information on cross-linguistic natural language lexically equivalent contexts within a given domain" (Peters et al. 2000: 74). At the same time, while corpus design is fundamental for comparable corpora to be useful for contrastive research, for parallel corpora "the sampling frame is irrelevant, because all of the corpus components are exact translations of each other" (McEnery et al. 2006: 48). What is extremely important for parallel corpora, however, is alignment. Alignment is obtained when "corresponding" segments of texts are matched and it can occur "at the level of individual sentences, paragraphs or text samples" (Mikhailov and Cooper 2016: 7). Its

goal is that of obtaining “translation equivalents of sentences, phrases or words between the source and translated texts in a parallel corpus” (McEnery et al. 2006: 50).

*Raw vs. annotated corpora.* Corpora can be stored in raw format or they can be annotated linguistically. Leech (1997: 2) defines linguistic annotation as “the practice of adding interpretative, linguistic information to an electronic corpus”. Linguistic information can be added on various levels of linguistic analysis, starting from a morphological layer and reaching to the discursive construction of a text. On the morphological level, tokenisation consists in the identification of separate words within texts, being tokens all the words of the collection; part-of-speech tagging (also called POS tagging) labels words according to the part of speech they belong to (e.g. nouns, verbs, etc.), while lemmatisation searches and groups all forms of a lemma across the corpus. Syntactic annotation signals the constituents or the dependencies that make up a sentence while semantic annotation encloses the semantic features of a corpus’s tokens into a tag. Finally, discursive and pragmatic annotation add interpretative information about the distribution of textual or pragmatic features respectively (Leech 1997).

According to Leech (1997), linguistic annotation enhances the quality of a corpus to three extents. First, it eases the extraction of grammatical information from the corpus: if a collection were not annotated, the plotting of grammatical patterns would be more complicated. Second, annotation increases the re-usability of the corpus by allowing all corpus users to profit from the linguistic analysis applied by the developer of the collection; apart from being helpful for many scholars, re-usability does also reward the creator of the corpus of the time expense that they went through when annotating the collection. Finally, Leech observes that not only is an annotated corpus re-usable, but it proves also multi-functional. In fact, thanks to linguistic annotation, the collection could be employed for purposes other than the one it was originally constructed for (Leech 1997).

### **1.3.3.2. Corpus design**

*Representativeness.* Corpora are affected by the thorny issue of representativeness. Representativeness is defined by Leech (1991) as the property of a corpus to reproduce the characteristics of a language or of a linguistic variety so that the results drawn from the analysis of the very collection can be comfortably generalised to the whole population. Biber (1993: 243), at the same time, considers it as “the extent to which a sample includes the full range of variability in a population”.

According to Biber (1993), a corpus's representativeness is bound to situational and linguistic variables (also called external and internal variables in McEnery et al. 2006). Situational variables are all those contextual characteristics that cover "the range of text types in a language": primary communication channel (spoken, written, or transcribed), format (published or unpublished), setting, addressee and addressor, factuality (distinguishing between informative and imaginative discourse), purposes and topic of the linguistic interchange. Linguistic variables, on the other hand, consist in the range of linguistic features in a language (Biber 1993: 243).

While the linguistic variables of a text or of a corpus can be determined only after the collection of the textual material, situational variables can be traced and decided beforehand. According to Biber (1993; but also in Atkins et al. 1992 and in Sinclair 1995), external variables should be prioritised when outlining the profile of the candidate texts to be included in the corpus. This could preserve linguistic variability within the corpus without a loss in the specificity of the collection.

In these cases, situational variables operate in determining the target population, i.e. the sum of the members that represent the language under inquiry. The target population can be operationalised in the form of a list of members (sampling frame) from which a representative sample can be extracted. In case feasibility does not enable a researcher to collect all the texts of a target population, a reduced set can be randomly sampled on the complete list of documents (Biber 1993).

Random sampling can be either simple or stratified. It is simple whenever all the texts are randomly chosen among a whole, irrespective of the group's internal variability. It is stratified if, prior to sampling, the whole is divided into sub-groups characterised by distinguishing features and then different sets of texts are selected in turn within the sub-groups by using simple sampling (Biber 1993).

However, not only does the process of sampling involve the isolation of candidate texts within a sampling frame, but it does also compel to decide on the length of the texts (i.e. whether to include full texts or chunks of texts). As far as sampling for specialised corpora is concerned, Baker (2006) suggests that corpora constructed for discourse analysis should contain whole texts in order to be representative of all the components of a writing (e.g. introduction, main body, conclusion, etc.; Baker 2006).

Also linguistic parameters should be pondered for the selection of the texts to insert in the corpus. In fact, once a pilot corpus is compiled with a certain sampling technique and by considering certain situational criteria, it should be analysed to ascertain whether the design features do cooperate towards representativeness. If this

is not the case, those features need to be changed, starting a cyclical process of corpus construction (Biber 1993). Tognini-Bonelli (2001: 61) specifies this by saying that after the construction of a pilot corpus, a scholar should measure the “coincidence of external and internal parameters” and, if they diverge, the researcher ought to reshape their configuration repeatedly until the corpus is representative enough.

McEnery et al. (2006: 21), however, believe that representativeness is extremely hard to achieve. It is a sort of mirage, “a statement of belief rather than a fact”: a corpus can never be perfectly representative of a language or of a linguistic variety. It can only aim for representativeness if it is constructed with pondered, suitable, and well-documented design criteria. Documentation, in fact, is essential because it enables the users of the corpus to judge whether the collection will suit their research purposes, being it satisfactorily representative or not (McEnery et al. 2006).

Apart from sampling, another feature that contributes towards representativeness is balance (Biber 1993). According to Kennedy, a corpus is balanced for certain features whenever the proportion of these characteristics resembles the one of the target population (Kennedy 1998). While balance is paramount for general corpora, specialised corpora demand not to be balanced because of their serving a different purpose in comparison with the former. So, on the one hand, a corpus that wishes to represent a standard variety of a language should be designed so that the proportion between written and spoken material (and, within these two categories, the proportion across sources, participants, etc.) resembles the one in the actual language use. On the other, a specialised corpus that reflects only a very limited type of discourse ought not to bother on the issue of balance (McEnery et al. 2006).

*Size.* Another thorny issue in designing a corpus is size: the literature is divided between those who state that textual collections ought to be as big as possible (Sinclair 1991) and others who believe that size is less important than other features and that corpora may be valuable in spite of their dimension (McCarthy and Carter 2001). Sinclair (1991), as well as other detractors of small corpora, believes that big corpora are essential to reproduce the characteristics of language. Recalling Zipf’s law (Zipf 1935, 1949; cited in Sinclair 1991), he observes that in a language (and, by extension, in a corpus) word frequency is not evenly distributed: lexicon and grammatical words follow an asymptotic pattern, with the most frequent items occurring exponentially more often than the least frequent ones. Some words (generally function words like *the*, *of*, and *at*) appear much more frequently than others, which may even occur only once or twice in the whole communication (Sinclair 1991).

Granath (2007) backs the benefit of using large corpora, but she believes that small corpora can boost the confirmation of someone's hypothesis on language. This observation is in tune with McCarthy and Carter's (2001) idea that it is not size but corpus structure what really matters in corpus linguistics. A similar approach is Leech's (1991: 10), who thinks that "size is not all-important" and that it must be coupled with corpus design and with the ease in collecting, processing and distributing the data. However, although standpoints about corpus size vary, authors agree on the fact that size for specialised corpora "is not-all important", as McEnery et al. (2006: 9) affirm.

*Design.* Specialised corpora built with web resources or from databases require to be well planned also in relation to the query words that are employed for the retrieval of the texts that have to be incorporated in the collection. In these cases, query term extraction can result from a blend of the techniques employed in three different circumstances: when building specialised corpora from the Web (Baroni and Bernardini 2004, Baroni and Ueyama 2004, Gatto 2014); when searching for material through Information Retrieval (Baeza-Yates and Ribeiro-Neto 1999, Chowdhury 2004); when collecting texts from a database (Gabrielatos 2007).

In all these cases, in fact, the choice of the terms (technically called *seeds*) used to query Web repositories or databases is crucial (Gatto 2014). The search words, in fact, will perform in the texts determining the content of the corpus. In case the Web repository or database is indexed the result of the query will be slightly different. Indexed material, in fact, can be queried not only for the words included in it but also for the labels that have been attached to its texts (Gatto 2014). In Stubbs's (2001: 7) words, they will contribute to the creation of "a mini-world or universe of discourse". Consequently, while brainstorming seeds that relate to the subject one wishes to investigate, the researcher ought to decide very carefully which to include and which to exclude from the search: too specialised seeds might result in an overspecialised corpus, that tackles exactly the researcher's hypothesis without leaving any space for a broader study of the matter; at the same time, too general search terms could cause an overgeneralization of the topic with a consequent loss of specialisation of the corpus (Sharoff 2006). In the latter case, this might result in an almost complete loss of information on the topic to be investigated, especially when the collection is small- or medium-sized. In big corpora, in fact, the amount of texts, and therefore of tokens, could make up for the generality of the corpus's content (Baker 2006).

### 1.3.3.3. Corpus analysis

Despite their different design, corpora in general tell us “what language is like” and they end up being “a more reliable guide to language use than native speaker intuition is” (Hunston 2002: 20) in so much that “some of the most revealing insights on language and language use have come from a blend of manual and computer analysis” (Kennedy 1998: 2-3). The opposition between manual and computer analysis introduces the opposition between qualitative and quantitative corpus analysis.

The difference between qualitative and quantitative corpus analysis, as the terms themselves imply, is that in qualitative research no attempt is made to assign frequencies to the linguistic features which are identified in the data. Whereas in quantitative research we classify features, count them, and even construct more complex statistical models in an attempt to explain what is observed, in qualitative research the data are used only as a basis for identifying and describing aspects of usage in the language and to provide ‘real-life’ examples of particular phenomena. (McEnery and Wilson 1996: 62)

Qualitative corpus analysis offers a precise and in-depth description of linguistic features that would be hardly identified with a quantitative approach. On the contrary, quantitative corpus analysis enables to track and generalise on frequent linguistic patterns that would be discovered with difficulty if a researcher adopted only a qualitative approach (McEnery and Wilson 1996).

These resources of qualitative and quantitative corpus analysis can be combined both “to count categories in traditional approaches to language” and “to observe categories and phenomena that have not been noticed before” (Hunston 2002: 1). The empirical testing of traditional approaches to language and the discovery of new categories and phenomena is in a nutshell the main goal of *corpus-based* and *corpus-driven* approaches to the use of corpora.

As Tognini-Bonelli (2001: 65) states, corpus-based approaches to the study of language employ corpora “to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study”. In other terms, corpus-based approaches exploit the rich linguistic material contained in corpora to document or eventually adjust their pre-conceived theories on language. This implies that scholars do isolate only those data that corroborate their hypotheses and they discount or treat as exceptions all those examples that confute their theory. From a methodological point of view, they utilise concepts like frequency of

occurrence, which allow them to assess whether a selected phenomenon is worth studying (Tognini-Bonelli 2001).

Corpus-driven approaches to the study of language develop the other way round in comparison with corpus-based approaches: they focus on the corpus and they let it spotlight meaningful patterns and features worth studying. Theoretical frameworks do not inform research but they spring from findings collected in corpora. These findings regard in many cases phenomena already known to the linguist but they can also touch on unexpected or unknown features of the language. This seems to be the most promising aspect of corpus-driven linguistics: it may supply a scholar with totally new research material. Research within corpus-driven linguistics employs methodologies like collocation analysis, i.e. the study of the words that co-occur most commonly with a certain term (Tognini-Bonelli 2001).

Quantitative corpus analysis relies on frequency and statistical counts that software operate on corpora. These frequency and statistical counts involve concepts like *frequency lists*, *keyword lists*, *collocational patterns*, *clusters*, and *n-grams*.

*Frequency lists.* Frequency lists, or *wordlists*, are lists of all the *types* in a corpus, namely of all the distinct word-forms included in the corpus, matched with the arithmetic count of their occurrence (on wordlists see, among others, Archer 2009 and Baker et al. 2006). In a wordlist types are registered with their *raw frequency*, i.e. the sheer count of the word-forms' occurrences, and with their *relative frequency*, i.e. the ratio between a type's raw frequency and the total number of words in the text or corpus. The items in a wordlist are generally ranked according to their frequency, from the commonest to the rarest ones, but they can also be ordered following a direct or inverse alphabetical order. In addition, frequency lists are sometimes organised according to the first appearance of a word, in order to reconstruct the distribution of contents and grammatical patterns across a text or a corpus. Corpora and their wordlists can also be lemmatised, so that all the various forms of a lemma are labelled under a single tag (Freddi 2014).

*Keyword lists.* Keyword lists are lists of the statistical keywords of a corpus, namely of the words that occur in a statistically significant way in a corpus in comparison with another (Scott 1997). Statistical keywords are frequently called simply keywords in the scientific literature on the matter. In the present research they are usually termed statistical keywords in order to avoid confusion with cultural keywords (on cultural keywords see § 1.1).

Pioneering work on statistical keywords has been carried out by Mike Scott since the beginning of 1990s in connection with the development of WordSmith Tools (Scott 1997, 1999, 2015, 2016).

In Scott's (1997: 136) terms, and more broadly in corpus linguistic terms, statistical keywords are words that occur "with unusual frequency in a given text". They can also be defined as those words "whose frequency (or infrequency) in a text or corpus is statistically significant, when compared to the standards set by a reference corpus" (Bondi 2010: 3).

A statistical keyword is thus a word that stands out in a text or in a corpus for its being statistically more or less frequent than the same words in another text or corpus. Statistical keywords are computed by comparing frequency distributions extracted from different texts or corpora. Keyword lists result from the comparison of two texts or corpora: a study and a reference corpus (Scott and Tribble 2006: 58). The "study corpus" (Gabrielatos 2018: 227; also termed "node corpus" in Milizia 2010: 132) is a text or a corpus whose prominent linguistic features a researcher aims at unveiling in comparison with a "reference corpus", which might be considered as a sample of the language population that the node corpus is compared to (Scott and Tribble 2006: 58).

Once study and reference corpora are chosen, statistical keywords are generated by comparing the frequency list of the study corpus with that of a reference corpus. This comparison is statistical and it produces indexes of statistical keywords ranked according to frequency or statistical score (Scott 1997; for more on this see § 3).

The tests used for measuring the statistical significance of a frequency distribution might be different depending on the software used and on the purpose of the study. Let us consider, for instance, the WordSmith Tools software and the AntConc software. WordSmith Tools is an integrated suite of programs developed by Mike Scott and devoted to the in-depth computational analysis of texts (Scott 2015). Similarly, AntConc is described by its developer Laurence Anthony as a "corpus analysis toolkit for concordancing and text analysis" (Anthony 2020). WordSmith Tools and AntConc have been using the *chi-square* or *Log Likelihood* statistical significance measures. Statistical keywords can be extracted using both statistical significance tests and effect size tests (a brief description of these statistical measures can be found in § A1 in the Appendix). Statistical significance and effect size can be combined, though: the results of keyword extraction with effect-size metrics can be boosted by evidence obtained with the use of statistical significance measures (Gabrielatos 2018).

The characteristics of the reference corpus significantly influence the outcome of keywords' extraction: according to Scott (2009), size, genre, topic, and age of the comparative collection must be pondered in order for statistical keyword lists to provide results which are suitable for one's research goals.

As far as size is concerned, it is generally believed that the bigger the reference corpus, the more effective the comparison (among others, Baker et al. 2006). Berber Sardinha (2004), for instance, recommends that the minimum dimension of a reference corpus should be between 2.3 and 5 times the size of the study corpus (Berber Sardinha 2004). Scott (2009) downplays the impact of the reference corpus's dimension by testing the precision<sup>3</sup> of keywords extraction with a reference corpus of growing size. Scott (2009: 86) holds that when one uses a "mixed bag" reference corpus, the larger the reference corpus the more precise results can be obtained; nevertheless, when the reference corpus is genre- or topic-specific, a "moderate sized" reference corpus can suffice. In fact, according to him, the differences in keyword lists are frequently due to factors other than size.

Two of these factors are genre or topic of the reference corpus (Scott 2009). To this extent, Berber Sardinha (2004) advocates that reference corpora should consist of texts belonging to a variety of genres and that they should cover a vast range of topics in order to enable keyword lists to be as ample as possible (Berber Sardinha 2004). In other terms, study corpora ought to be compared to general corpora. The choice between general purpose corpora and specialised corpora for the extraction of statistical keywords depends on the purposes of the research: general purpose corpora elicit differences and similarities between the language variety of the study corpus and an approximation of the general language; specialised corpora contrast the linguistic uses of the study corpus with the specificity of the texts included in the specialised corpus (Baker 2006).

In addition, in case of using a specialised corpus as reference corpus, both inter- and intra-corpus comparisons are possible. This means that a specialised corpus can be divided into subcorpora and these subcorpora can be compared in order to highlight their specificity in terms of keywords (Culpeper 2001, Baker 2010b).

Once size, genre and topic of the reference corpus are decided on, another factor that has to be taken into account for the choice of the appropriate *tertium comparationis* is age or "date" of production of the reference corpus (Scott 2009). In brief, if the

---

<sup>3</sup> Precision is "the proportion of retrieved items that are in fact relevant (the number of relevant items obtained divided by the total number of retrieved items)" (Oakes 1998:176).

reference corpus contains texts produced at the same time of the study corpus's, the comparison of the two corpora can highlight synchronic differences in discourse; if the texts of the reference corpus date from earlier or later than those of the study corpus, the comparison might produce, among others, results of diachronic variation in discourse (see, for instance, Baker 2010a).

This synchronic and diachronic variability of the results can involve a range of linguistic features. From a lexical point of view, in the diachronic study of newspaper discourse, for example, "keyword comparisons with a reference corpus that is not contemporary with the pilot corpus are bound to favour words referring to entities, concepts, etc. which were not current in the period represented by the reference corpus" (Gabrielatos 2007: 13). On other levels of linguistic analysis (e.g. morphological, syntactic, etc.), the same diachronic comparison might turn out statistical keywords that possibly highlight the development of new grammatical patterns (especially in a broad diachronic perspective as the one adopted in historical linguistics, e.g. Rissanen 2008, Rissanen et al. 1993).

A diachronic perspective in the study of statistical keywords raises the issue of time, as it compels the researcher to settle the segmentation of the time-span covered by their study and reference corpora (treated, among others by, Baker et al. 2013a, Duguid 2010, Gabrielatos 2007, Marchi 2018, Taylor and Marchi 2018b). According to Taylor and Marchi (2018b: 52), "there is no ultimately "right" unit (there certainly is no single natural one) and defining the appropriate level of aggregation, and the parameter on which to establish it, depends on establishing "appropriate for what purpose"". Notwithstanding, three criteria might help: "extra-linguistic criteria" centred upon "contextual historical knowledge or convention"; "text-lifecycle" criteria, dependent on publisher's choices; "bottom-up" criteria that set certain pieces of information as cornerstones (as in Gabrielatos and Baker 2008; Taylor and Marchi 2018b: 52).

Statistical keywords can be classified as *global* when they are "dispersed more or less evenly throughout the text" and as *local* if they are localised in a text fragment (Scott and Tribble 2006: 66). In other words, global statistical keywords track linguistic items that appear to be salient throughout a whole text or corpus; local statistical keywords draw attention to items that are key in a limited section of the studied text or corpus. While the former can be said to highlight some of the most meaningful linguistic items of the language or linguistic variety represented by the text or corpus under enquiry, the latter point to peculiar items of a specific component of that language or linguistic variety (Scott and Tribble 2006).

Statistical keywords can also be distinguished into *positive* and *negative*. Positive statistical keywords are words that occur “*more* often than would be expected by chance in comparison with the reference corpus”; negative statistical keywords are words that appear “*less* often than would be expected by chance in comparison with the reference corpus” (Scott 1999: 251). In other words, the former show what items are statistically peculiar in a study text or corpus, whereas the latter focus on those that are statistically poorly represented in the text or corpus.

*Collocational patterns.* Collocational patterns showcase the collocates of a word, namely the lexemes that co-occur with the searched type in a statistically significant way within a certain textual horizon (Sinclair 1991).

According to Brezina et al. (2015), collocation can be described according to the features of *distance*, *frequency* and *exclusivity*, and then *directionality*, *dispersion*, *type-token distribution* among collocates, and *connectivity*.

Distance (diversely called *collocation window* or *span*) determines the limits of the horizon within which collocates are looked for (Brezina et al. 2015). The collocation window, in fact, consists in “the number of word-forms, before and/or after the node (e.g. 4:4, 0:3), within which collocates are studied”, with *node* standing for the word whose collocational pattern is searched for (Stubbs 2001: 29). Once a researcher selects a node, they can choose to look for the immediate collocates of the term (e.g. one or two words to the left and/or to the right of the node) or to expand their collocation window so as to reach items that stand three, six, ten words on the left and/or on the right of the search word.

Stubbs (1995a: 8) writes that there is not a “real agreement in the literature about appropriate window size”. Some believe a 4:4 span to be the best solution to find consistent collocates (Jones and Sinclair 1974, Sinclair 1991) while others recommend expanding searches up to 10 word-forms to the left and to the right of the node, paying particular attention to the collocates falling within a 5 collocation window (Baker 2006).

For example, focusing on the expression *naked eye*, Sinclair (1996) observes that collocates found one word to the left of the node generally consist on function words like *the*, *your*, *our*, *a* or *to*. Among them, *the* is by far the most frequent, as well as the one founding the core meaning of *naked eye*, which is: something can be seen with the naked eye only when it is so close that a person does not to require any tool to recognise its shape. Sinclair (1996) moves then one step backwards and he includes *the* in the searched cluster; in *the naked eye*'s N-2 position, grammatical patterns emerge

with prepositions like *with* and *to*: the study of collocation expands on the observation of colligation. In the case of *naked eye*, the colligational power of *with* and *to* opens to a basic grammatical structure of the node's collocation networks, i.e. preposition-*the-naked-eye*. N-3 and N-4 spans shift the role of the collocates from syntactic to semantic (with semantic preference) and pragmatic (with semantic prosody). As far as semantic preference is concerned, *naked eye* co-occurs with words referring to visibility (as with various forms of *see* or *visible*) or with grammatical terms that build patterns to introduce visibility (prepositions like *to...*); in the case of semantic prosody, modal verbs such as *can* reduce the act of seeing to a difficulty and to something that needs to be boosted by technical instruments. With both collocation windows, *naked eye* is modified in its semantics and in its pragmatics, pointing at the concepts and attitudes that relate to the expression (Sinclair 1996).

Brezina (2018b: 273) summarises this controversy by noticing that, when analysing discourse,

A narrow span creates a more focused view on word co-occurrence but can lead to an omission of important meaning associations (think, for instance, of the phrase *long dreary time* in which only the second adjective will be captured by 1L, 0R span). A large span, on the other hand, can introduce noise into the collocation analysis, that is words that are not directly associated with the node such as words occurring across sentence and paragraph boundaries (... *metal cannulae could be inserted. To allow time for veins to recanalize ...*). It is important to remember that the span operates as a zoom helping us focus the analysis on the most relevant set of collocates as defined by the research question.

Frequency counts the co-occurrence of node and collocate within the chosen span (Brezina et al. 2015). According to Evert (2008: 1215), "any pair of words that cooccur at least twice in a corpus is a potential collocation according to this view" (Evert 2008: 1215). Nevertheless, only frequent co-occurrence is the "hallmark of an attraction between words".

As a consequence, several scholars suggest applying *frequency thresholds* when calculating collocation (among others, Baker 2016 and Evert 2004, 2008). Frequency thresholds isolate significant collocational patterns, while discarding weaker ones. The selection of these thresholds, that might consist in 3, 5, 10, 50, etc. words, is justified for mathematical reasons.

These mathematical reasons resume Zipf's law (Zipf 1935, 1949). According to Zipf's law, in most languages the distribution of the lexicon's frequency resembles a down-sloping asymptotic curve in which the first part of the curve is populated by high frequency terms and the final by low frequency terms or *hapax legomena* (Zipf 1935, 1949). For example, it "is not uncommon to find more than a million recurrent word pairs ( $f \geq 2$ ) in a corpus containing several hundred million running words, but only a small proportion of them will pass a frequency threshold of  $f \geq 10$  or higher, as a consequence of Zipf's law" (Evert 2008: 1215). As a result, frequency thresholds require to take into consideration the dimension of the corpus under inquiry (Evert 2004, 2008).

Setting frequency thresholds is further justified by the fact that collocations appearing only few times do not hold much power over lexical priming in the minds of speakers and listeners (or writers and readers; cf. § 1.2). On the contrary, those word-combinations emerging more frequently in a text or in a corpus are more likely to prime certain linguistic patterns (and so, specific concepts; Baker 2016).

In search for robust collocational patterns, Evert (2008: 1215) claims that sheer "recurrence is not a sufficient indicator for a strong attraction between words" and that an "additional measure of attraction strength is therefore needed in order to identify "true collocations" among the recurrent word pairs, or to distinguish between "strong" and "weak" collocations". This additional measure of attraction strength is statistical exclusivity or *association measure*.

An association measure quantifies the strength of the attraction that ties two words together and it is calculated with the frequency of the node, with the frequency of the collocate, with the frequency of their co-occurrence and with the frequency of the corpus's types (Evert 2008).

Association measures are characterised by diverse statistical natures and by formulas favouring different aspects of collocation (frequency, strength, direction, etc.). The choice of one of them depends on research goals, although testing multiple measures at the same time might bolster the strength of a study's results (Evert 2008).

When extracting collocational patterns in corpus-assisted analysis of discourse, it is recommended to apply statistical thresholds, "which are used as filters that enable us to see only the most relevant collocates" (Brezina 2018: 275).

Statistical thresholds consist in scores (2.0, 5.5, 13.0, ...) which separate strong and weak collocates. In general, collocational scores are figures derived from the computation of a certain association measure for a specific node-collocate combination and their value suggests how strong the collocation is. The higher a collocational score,

the more robust the collocation it represents (Hunston 2002). Strong collocates are characterised by a collocational score which is higher than the cut-off score, while weak ones are ranked below the threshold. Cut-off thresholds assist in ensuring that no co-occurrence is mentioned that does not reproduce one of the most salient word-combinations in a text or in a corpus (among others, Hunston 2002 and Durrant and Doherty 2010).

Directionality rules on the path taken by a collocational pattern matching two words ( $word_1$  and  $word_2$ ), i.e. on “whether  $word_1$  is more predictive of  $word_2$  or the other way round” (Gries 2013: 141; cf. Stubbs 1995a).

The relationship between node and collocate can diffuse symmetrically or asymmetrically: it is symmetrical in case the node triggers the collocate with the same readiness with which the collocate triggers the node; it is asymmetrical if only one of the two terms requires the other to complete its meaning. Brezina et al. (2015: 140) highlight that “the strength of the attraction between two words is rarely symmetrical” but most association measures are unidirectional, notwithstanding.

Dispersion is regarded as “the distribution of the node and the collocates in the corpus” (Brezina et al. 2015: 140; cf. Gries 2008). A node-collocate combination can be widespread in a corpus, representing a standard, recurrent linguistic tendency, or it could be confined to a single text or to a bunch of linguistic productions and signal in this way stylistic or thematic specificity (Brezina et al. 2015; Gries 2008). This notion is derived from the notion of dispersion of simple frequency (as in Scott’s Wordsmith) and it is also applied to keyness.

Type-token distribution (or *entropy* in Gries 2013: 158) among collocates “takes into account not only the strength of a given collocational relationship (say between *love* and *affair*), but also the level of competition for the slot(s) around the node word from other collocate types” (Brezina et al. 2015: 141). Type-token distribution calculates what other words or phrases could substitute the collocate in the same position and how strong the new collocational relationship could be (Brezina et al. 2015; Gries 2013).

Connectivity consists in the property of collocation of creating networks filled with semantic meaning thanks to the relationship between a node and a collocate first, and then between the node and the collocates of the first collocate (Brezina et al. 2015). In other terms, it reproduces the semantic potential of the networks created by node, first-order collocates, second-order collocates, etc. First-order collocates are the collocates of a word, while second-order collocates are the collocates of one of the first-order collocates (Brezina et al. 2015; cf. Baker 2016).

*Clusters and n-grams.* Clusters are defined as “any group of words in sequence” (Baker et al. 2006: 34) and they are also called *lexical bundles* (Biber et al. 1999, Biber et al. 2002, Biber et al. 2004, Cortes 2002, Cortes 2015). Lexical bundles are “sequences of word forms that commonly go together in natural discourse”, “regardless of their idiomaticity, and regardless of their structural status” (Biber et al. 1999: 990). Lexical bundles are thus sequences of words that co-occur frequently and without being interrupted by intruding lexical material (Biber et al. 1999). They are usually intended to include three or four words; two-word sequences are too short and common to be regarded as meaningful recurring patterns while longer sequences are often too rare to be included for research (Biber et al. 2002). Lexical bundles can also be meant as “extended collocations: bundles of words that show a statistical tendency to co-occur” (Biber et al. 1999: 989). They differ from collocations, however, “mostly in the word class of the components that make up lexical bundles” since they do not only trigger content words but they also track the co-occurrence of function words (Cortes 2015: 200).

N-grams are sequences “of n letters from a given string after removing any spaces” (Baker et al. 2006: 122).

*Concordances.* A concordance is “a collection of the occurrences of a word-form, each in its own textual environment” (Sinclair 1991: 32). This textual environment is nowadays usually displayed in *K.W.i.C.* format (or *Key Word in Context* format), which places the searched word at the centre and then lists lines according to the alphabetical order of the texts they belong to and, after that, according to the order of appearance of the lines within the texts. These concordance lines can be further sorted to the left or to the right of the search term. When sorting to the left or to the right of a word, *concordancers* (i.e. software for linguistic analysis that can extract concordances) catalogue the concordance lines according to the alphabetical order of the terms standing in a certain position to the left or to the right of the query term (Freddi 2014).

### **1.3.4. Corpus approaches to discourse**

The methodologies and tools offered by corpus linguistics assist critical and ecological discourse studies in the productive endeavour of corpus approaches to discourse. Corpus approaches to discourse profit both from methods and techniques which are typical of corpus-based approaches (e.g. frequencies and concordances) and they

improve them with methods and techniques that characterise corpus-driven studies (as with collocates; Partington et al. 2013).

The richness and robustness of the methodology offered by both corpus-driven and corpus-based approaches to the study of discourse contributes to corpus approaches to discourse. This combination of approaches makes “it easier to know where to start” when analysing language (Mautner 2009: 34). The use of corpora can thus offer a starting point for discourse research by revealing hidden patterns or by strengthening the confidence in handling renowned patterns.

The overarching methodology of corpus approaches to discourse enables to reduce research bias when carrying out a critical or ecological study of discourse (Baker 2006). In fact, the application of suitable methodologies, tools and corpora to discourse analytical research methods can help in discovering or confirming recurrent patterns that overtly or secretly reproduce an ideological standpoint (Baker 2006). To this extent, the combination of corpus linguistics and critical and ecological discourse studies is more than complementary, with the two fields intertwining since

the reason that drives researchers to work with corpora and discourse is epistemological in nature and their goal is achieving greater soundness and greater breadth in their research. (Marchi and Taylor 2018: 5)

The greater soundness and breadth in the research that emerge at the interplay of corpus linguistics and critical and ecological discourse studies can be better obtained by means of triangulation, namely “a research approach that takes two or more perspectives to investigate a research question” (Egbert and Baker 2020: 5). According to Egbert and Baker (2020: 7-8), in fact, “[d]iscourse analysis is one area that has seen an abundance of triangulation research with corpus linguistic methods”; at the same time, “[c]ritical discourse studies is triangulatory in of itself, regardless of whether corpus approaches are combined with other techniques within an individual study”.

Triangulation “can involve using multiple methods, analysts, or datasets” and it “has been used for decades by social scientists as a means of explaining behaviour by studying it from two or more perspectives” (Baker and Egbert 2016: 3).

To this extent, Marchi and Taylor (2009) identify four forms of triangulation: methodological triangulation, data triangulation, investigator triangulation, and theoretical triangulation. Methodological triangulation consists in “using more than one methodological approach to gathering and interpreting the data” (Marchi and Taylor 2009: 4). In the case of corpus approaches to discourse, “the two methodologies

of corpus linguistics and discourse studies are deliberately intertwined, and interaction is carefully exploited” (Marchi and Taylor 2009: 5). Egbert and Baker (2020: 5) state that methodological triangulation is “the most widely used and, arguably, the most useful and applicable type of triangulation for corpus research”. However, according to Marchi and Taylor (2009: 4), “integrating and complementing methodologies is not a solution to the vexed problem of objectivity and [...] there is more to *triangulation* than combined methods”. Data triangulation involves the employment of various data sources or of data collected from the same source but with a variety of sampling methods (Marchi and Taylor 2009). Investigator triangulation “refers to the use of more than one researcher to collect and/or interpret data” and it is rarely operationalised in corpus approaches to discourse analysis (Marchi and Taylor 2009: 5). Finally, theoretical triangulation “refers to the use of more than one theoretical position in interpreting data, and the explicit use of this type of triangulation is much less common than methodological or data variation” (Marchi and Taylor 2009: 5).

Triangulation can result into convergence, dissonance, and complementarity of the outcomes of the study (Marchi and Taylor 2009). In the case of convergence, “the results of the triangulation confirm one another” while when dissonance occurs, “the findings of one methodology, researcher, data-set or theoretical position are incompatible with the findings of another” (Marchi and Taylor 2009: 6-7). Convergence is the preferred outcome of triangulation because it seems to confirm the reliability of a research even though the falsification brought forward by dissonance is epistemologically interesting. In addition, complementarity regards findings as pieces of a broader picture, “which when put together may offer a more complete view of the construct which is being investigated” (Marchi and Taylor 2009: 7).

Corpus approaches to the analysis of discourse can rely nowadays on a well-established and sound theoretical and methodological framework that enables to carry out robust corpus-aided analyses of ideology in discourse. At the same time, some of their aspects are still under development and they allow some space for novel results or ideas (Marchi and Taylor 2018).

As a consequence, corpus studies of discourse, conducted from a critical perspective, have covered many discourse types so far. They have elaborated on media discourse, for instance. Baker et al. (2008) analyse the representation of refugees in the British press (see also Gabrielatos and Baker 2008). A similar research pattern is followed in Baker et al.’s (2013a, 2013b) study on the representation of Islam in British broadsheets and tabloids, in Mautner’s (2007) research on the concept of *elderly* in the Bank of English (a corpus of British English produced between early and mid-1990s)

and in Daryell and Urry's (2015) discussion of the issue of climate change, as dealt with by the Brazilian media.

Corpus researches have also addressed history. McEnery and Baker (2017: 2), for example, integrate corpus linguistics into the study of "women who traded sex for cash or some benefit in kind in the seventeenth century". They unfold their project from a historical and from a linguistic point of view: they report the historiographical treatment of the matter and they complement it with linguistic material. They clarify that corpus methodologies and techniques for the analysis of texts prove helpful in endorsing or refuting the hypotheses of historical research by guaranteeing availability of large amounts of linguistic data, more in comparison with the number of historical discourses that historians would probably use (McEnery and Baker 2017).

Also political discourse has been attracted into the wealth of corpus approaches to the analysis of discourse. Krizsán (2011), for instance, studies identity as it is linguistically portrayed in speeches uttered in English about the EU-enlargement that happened in 2004 by Hungarian, British and Finnish politicians; he focuses on collective identities and he studies the frequency and the contextual use of deictic pronouns designating the self. Mautner (2000) focuses on national identity in Euro-discourses published by four British newspapers (*The Guardian*, *The Daily Telegraph*, *The Daily Mirror*, *The Sun*) while Teubert (2000) researches Euro-sceptic discourse in the United Kingdom, trying to map the main themes revolving around the issue of Euro-scepticism.

Not only do corpus approaches to the critical study of discourse explore different discourse types but they also take different shapes, as Marchi and Taylor (2018: 5) affirm:

rather than one approach there is, in fact, a diverse range of kin approaches that go under different names, such as *corpus-based CDA* (e.g. Baker et al. 2008), *CADS* (e.g. Partington 2004), *discourse-oriented corpus studies* (Gabrielatos, private conversation), *corpora and discourse studies* (Baker & McEnery 2015) or under no particular label. Furthermore, much work that prefers the names *corpus stylistics* (e.g. Mahlberg & McIntyre 2011) or *corpus pragmatics* (e.g. Aijmer & Rühlemann 2015) and some *corpus-based sociolinguistics* (e.g. Friginal & Hardy 2014) will be very similar in scope and tools to projects using the 'discourse' range of labels.

In the present research, the approach of corpus-assisted discourse studies (Partington 2004, Partington et al. 2013) is preferred over the others and it is employed for the analysis of the discursive construction of sustainable development.

Corpus-assisted discourse studies are corpus approaches to the study of discourse that rely on a particular interpretation of triangulation. In the case of Partington et al.'s (2013) corpus-assisted discourse analysis, triangulation requires a good acquaintance with the corpus under inquiry and a sound knowledge of the historical, political and social context that produced the texts included in the collection. It is only such a historical, political and social engagement that can offer real nourishment to corpus-assisted analyses of discourse (Partington et al. 2013). Knowledge of the language-external context can be gathered through an interdisciplinary work that takes into consideration the theoretical and methodological assumptions of other disciplines (van Dusseldorp and Wigboldus 1994). According to Potts (2018: 1), however, interdisciplinarity needs to be well-balanced in order not to become a "Frankenstein monster".

Mautner (2016), however, sheds light on potential weaknesses of corpus approaches to discourse analysis and she pinpoints six thorny issues. The first is the "skills gap and lack of standardization", which involves both critical discourse analysts and corpus linguists; according to Mautner (2016), several critical discourse analysts are not confident with the methods and tools offered by corpus linguistics and, at the same time, methods and tools provided by corpus linguistics show an excessively high level of variability. The second flaw consists in "institutional barriers", and it depends on the fact that "[c]ritical discourse analysts and computer linguists do not necessarily work in the same departments, and, if they do, may not communicate well with each other" (Mautner 2016: 139). Mautner's (2016: 140) third observation deals with "resisting temptation in data collection", namely in avoiding both the analysis of a small sample of cherry-picked texts and the unguarded collection of huge amounts of data without careful planning (on this see also Baker and Levon 2015, Mautner 2007). In addition, Mautner (2016) warns against decontextualizing data and she recommends that corpus-assisted discourse analysis be carried out by balancing details and broad pictures. Finally, Mautner (2016; 142) focuses on the potentialities and limits of corpora for the investigation of language innovation and she raises some epistemological issues: she concludes by saying that in corpus-assisted discourse analysis a research proves an "eclecticism that is imaginative and productive" and not "aimless patchworking" thanks to

(1) a clear statement of the aims of your project, (2) a rigorous assessment of what each method can and cannot do, and (3) robust theoretical foundations capturing core assumptions about language and the social (Mautner 2016: 142).

On this point see also Fruttaldo and Cambria (2017).

In spite of the weaknesses highlighted by Mautner (2016), corpus approaches to discourse have proven extremely helpful for carrying out robust analyses of various kinds of discourse with various kinds of research goals. Among these, corpus approaches to discourse have been contributing to the corpus-assisted extraction of cultural keywords and to the corpus-assisted retrieval of meaning by collocation.

#### **1.3.4.1. Corpus approaches to the extraction of cultural keywords**

As far as cultural keywords are concerned, Jeffries and Walker (2017) suggest that, when adopting a corpus approach to discourse, cultural keywords are strongly intertwined with the notion of statistical keyword, which has been introduced in relation to corpus linguistics (on this see § 1.3.3).

Statistical keywords operate in a text or in a corpus towards the cohesive construction of the text's or corpus's aboutness and style and they can "direct the researcher to important concepts in a text (in relation to other texts) that may help to highlight the existence of types of (embedded) discourse or ideology" (Baker 2004: 347). In other words, they might "indicate the topics that are possibly of significant (or even ideological) interest to members of a discourse community" (Vessey 2016: 71). These topics might be regarded as examples of cultural keywords.

This holds true not only for words but also for multi-word expressions. Francis (1993: 155), for example, maintains that the study of key phraseological patterns might assist in identifying "the typical meanings that human communication encodes" and in recognising "untypical and hence foregrounded meanings whenever we come across them". Keyness, in fact, "applies equally to word forms, lemmas and word sequences" (Bondi 2010: 3; see also Scott 2010). It can be computed for single words but it can also be calculated for lemmas. In addition, it can be measured for word sequences. Statistical salience or lack of salience in a text or corpus should be extended to more complex word sequences and not only be associated with words, as the very term statistical keyword implies. Because of this, Wilson (2013: 3) proposes that the more neutral expression "key item" substitute the more common lexeme "keyword".

The fact that a key item appears more or less frequently in a corpus in comparison with another, especially when the study corpus is specialised, encourages to believe that that linguistic item is prominent in the construction of a stereotyped schema on reality that the author or the writer/s of that corpus have. As Scott and Tribble (2006: 87) put it, when analysing statistical keywords,

what we are accessing is stereotype. That is, the linkages created within and between texts are not claimed to be other than what writers decide to say, and what they choose to tell us is not at all the same as what co-occurs in real life.

In other words, statistical keywords do not depict the world but an “image of the world” as it is conceptualised by the author of the text or corpus (Scott 2002: 46).

The relationship between key items and reality, as well as between texts, text-types, and reality, however, is problematic (Stubbs 2010). Stubbs (2010: 40) affirms that “social institutions and text-types imply each other: they are different ways of thinking about the same thing”. Corpus approaches to discourse and the notion of statistical keyword can empirically trace the way in which linguistic items occur and co-occur to create speech events and, recursively, social agency, and social institutions (Stubbs 2010).

Furthermore, the notions of positive and negative statistical keyword advance towards the study of the discursive dichotomy of *presence* and *absence* (among others, Partington 2014, Duguid and Partington 2018, Schröter and Taylor 2018, Taylor 2012, Taylor 2013a), and, as a consequence of the ecolinguistic concepts of salience and erasure. According to Partington (2014: 129), the distinction between presence and absence is strongly associated with the concept of statistical keyword, since “perhaps the most obvious and most frequent way of identifying what is entirely or relatively absent – what is either missing or rare – in one set of texts is to compare that set with another set”.

Statistical keywords are adopted in much literature on corpus-assisted studies of discourse for the retrieval of cultural keywords and of other politically and culturally significant lexemes not explicitly regarded as cultural keywords (e.g. Baker 2010b, Bassi 2010, Jeffries and Walker 2017). Two exemplary studies of statistical keywords in corpus-assisted discourse studies are Gabrielatos and Baker (2008) and Baker et al. (2013). In their analysis of the representation of refugees in the British press, Gabrielatos and Baker (2008) search for the words that occur in a statistically more significant way in broadsheets compared to tabloids and then those that appear more

in tabloids compared to broadsheets. At the same time, when researching media attitudes in discourse, Baker et al. (2013a: 98) write that while in “the examination of statistical keywords newspaper by newspaper, [...] certain words were more likely to reveal a continuing focus in a particular paper over a long period of time”, when they contrast the same newspaper year after year they observe “the fact that the keywords of each year tended to be relatively transient, based as they are upon specific news events”.

Nevertheless, to O’Halloran (2010: 566) “corpus-comparative statistical keywords may or may not coincide with (corpus-based) cultural keywords”. Cultural keywords, in fact, are lexical words from a limited range of terms; statistical keywords, on the other hand, may be both lexical and grammatical items extracted from a large set of words. In addition, the former are intuitively collected by the analyst after their “cultural and political sympathies”; the latter are obtained through an objective process of frequency comparison (O’Halloran 2010: 567).

In this corpus-assisted framework, statistical keywords and cultural keywords belong “firmly in a text focus” (Scott and Tribble 2006: 7). In other words, they function as gears of *textuality*, as pivotal elements that assist in the construction of a *text*. A text is defined as “a COMMUNICATIVE OCCURRENCE” (de Beaugrande and Dressler 1981: 3) or as “coherent and interconnected pieces of language, as distinct from unorganized strings of sentences” (Halliday and Webster 2014: 183; see also, among others, Conte 1977, van Dijk 1977, Dressler 1972, Halliday and Hasan 1976, Mortara Garavelli 1979).

In Halliday and Webster (2014: 183), unorganised strings of sentences become a text whenever they display connections within them and among them, namely, when they are characterised by “inter-sentence texture” (*cohesion*) and “intra-sentence texture”.

Inter-sentence texture (or cohesion), in particular, “concerns the ways in which the components of the SURFACE TEXT, i.e. the actual words we hear or see, are *mutually connected within a sequence*” (de Beaugrande and Dressler 1981: 3). This connection produces lexical and grammatical patterns of presence and absence of intertwined elements which function as the reinforcing steel of a wall.

For Halliday and Hasan (1976), cohesion can be obtained by means of *reference*, *substitution*, *ellipsis*, *conjunction*, and *lexical cohesion*. Lexical cohesion, in particular, is realised when a word is reiterated throughout a text in the form of *repetition*, *synonym* or *near-synonym*, *superordinate*, or *general word*, or through collocation (namely the “occurrence of a different lexical item that is systematically related to the first one”; Halliday and Hasan 1976: 284). Halliday and Hasan (1976: 274-278) clarify this as

follows: repetition occurs whenever the same word is reiterated in the text (e.g. “*mushroom* refers back to *mushroom*”); synonym or near-synonym connect two words with the same meaning or with a very similar meaning (e.g. “*climb* refers back to *ascent*” for synonyms and “*brand* refers back to *sword*” for near-synonyms); superordinate is “a name for a more general class” (e.g. “*vehicle* is a superordinate of *car*”); general words, and in particular general nouns, function as an interplay between lexical items and grammatical items and they are “nouns having generalized reference within the major noun classes, those such as ‘human noun’, ‘place noun’, ‘fact noun’ and the like”.

Statistical keywords are viewed by Bondi (2010: 3) as “chains of repetition in text”. They might therefore be deemed to function as “a mechanism of global textual cohesion” whenever they are dispersed throughout a text (Stubbs 2010: 28).

As lexical cohesion involves content words, statistical keywords contribute to this with proper nouns first, then with lexical words (Scott 1999; also Baker 2006). Proper nouns highlight the specificity of a text in terms of content and contextual historical background of the study corpus. Lexical words (e.g. nouns, verbs, adjectives, adverbs) can be signals of the “aboutness” of a text, namely of what the text is about (after Phillips 1989).

Statistical keywords contribute to lexical cohesion when they gather in a single text. If they are dispersed across texts in a corpus, they cannot be said to be cohesive devices. They might be devices of intertextuality, instead. According to de Beaugrande and Dressler (1983: 10), in fact, intertextuality “concerns the factors which make the utilization of one text dependent upon knowledge of one or more previously encountered texts”. Statistical keywords could thus aid intertextuality in that they connect various texts through lexical repetition. Stubbs (1996: 91-92) claims that all “language is intertextual: it is shaped by prior texts, oriented to conventions and interpreted against the background of a very large corpus of linguistic experience”.

Finally statistical keywords can be indicative of the style of a text or writer. Style is defined by Leech and Short (1981: 10) as “the way in which language is used in a given context, by a given person, for a given purpose”. More precisely in relation to written text, it is meant to be “the linguistic characteristics of a particular text” (Leech and Short 1981: 12). These linguistic characteristics involve both lexical and grammatical patterns and they are thought to reflect the ideas of the author of a text or of the authors of a corpus of texts (Leech and Short 1981; see also, among others, Semino and Short 2014).

In keyword analysis, the stylistic peculiarities of the study corpus can be spotlighted not only by lexical words but also by grammatical words (Scott 1999). Notwithstanding, research on proper nouns and lexical words is frequently preferred to that of grammatical words. According to Baker (2006: 127), “as the style of a text may play some role in the discourses within it, it is recommended that such high frequency words are not discarded” without further exploration.

In this context, statistical keywords might shed light on cultural keywords and in so doing they can contribute to the textuality of a text by interspersing its components with politically and culturally paramount elements.

#### **1.3.4.2. Corpus approaches to the retrieval of meaning by collocation**

As in the case of cultural keywords, also meaning by collocation can be assessed within the framework of corpus-assisted discourse studies. Following from Stubbs, in contemporary corpus-assisted discourse analysis, collocation is generally recognised “when a word regularly appears near another word, and the relationship is statistically significant in some way” (Baker 2006: 95-96). Because of its nature, collocation helps “to summarize the most significant relationships between words in a corpus”, while “helping to spell out mainstream discourses”<sup>4</sup> (Baker 2006: 118). In other words, it helps to highlight linguistically expressed social constructions.

Corpus-assisted extraction of meaning by collocation involves first collocational patterns, organised then into colligational patterns, semantic preference and evaluative prosody, as corpus lexicology (Kosem 2015, Heid 2008, Moon 2010, Walter 2010 among others) has recently suggested within the broader field of corpus semantics (Stubbs 2015). Collocates can be listed according to their semantic area (semantic preference) or they can be grouped according to their connotational value opening thus to a broader evaluative harmony of the text or of the corpus (evaluative prosody). Semantic preference and evaluative prosody mirror ideology and evaluation in that they report shared positive, neutral or negative stances towards a concept (Mautner 2007). Additionally, they pinpoint divergence among the collocational patterns of words (Koller and Mautner 2004, Orpin 2005). This divergence is “likely to highlight differences in areas of activity, people or places with which each word is associated and may also highlight differences in the attitude encoded by the user of the item to

---

<sup>4</sup> In Baker (2006), discourse is meant as in Foucault’s terms as “practices which systematically form the objects of which they speak” (Foucault 1972: 49; in Baker 2006: 4). These discourses can be constructed, among other ways, “via language” (Baker 2006: 5).

what is being spoken of” (Orpin 2005: 49). In other terms, this difference can spotlight the various evaluative and ideological stances that the speaker or writer wishes to convey and that might become stabilized and naturalized. In fact,

From an ideological point of view, collocates are extremely interesting, as if two words are repetitiously associated with each other, then their relationship can become reified and unquestioned (Baker and McEnery 2015: 2)

Consequently, the recurrent company that words keep to each other might prime the receiver of a discourse to build the meaning of those lexical items according to their collocates. Moreover, it might prime the receiver to expect that a certain lexical item be characterised by a specific colligational pattern or by peculiar semantic or pragmatic associations. From a psychological point of view, this proves extremely powerful: the strengthening of linguistic patterns through recurrent collocation primes the mind of the receivers of the discourse to access reality through a very specific lens (Hoey 2004, 2005).

In the case of news discourse, this is even more so as “modern cultures harmonize the primings of a linguistic community [...] through the mass media” (Hoey 2004: 393). In the light of this, corpus-assisted analysis of discourse retains a paramount potentiality: “the computer corpus cannot tell us what primings are present for any language user, but it can indicate the kinds of data a language user might encounter in the course of being primed” (Hoey 2005: 14).

As it has already been stated in relation to drifts in priming, collocational patterns and the evaluative and ideological patterns that derive from them are not necessarily stable throughout time. The study of evaluation and ideology through collocation analysis can thus be conducted both synchronically and diachronically. While a synchronic study of collocational patterns portrays evaluative and ideological stances as they are conceived when the corpus’s texts are produced, a diachronic study of collocation might track the development of the same stances throughout time. Marchi (2010), for instance, describes the evolution of the collocational patterns of words regarding morality in the UK press. McEnery et al. (2019) focus on the change in the collocational behaviour of the terms *whore* and *harlot* in historical discourse and, to do so, they adopt a new methodology for the diachronic study of large corpora, namely Usage Fluctuation Analysis. Usage Fluctuation Analysis plots the variation of collocation networks in a graph and it allows to graphically visualize the change of a word’s collocates (McEnery et al. 2019).

However, collocational patterns (together with patterns of colligation, semantic preference, and evaluative prosody) need not be generalised across a language but they need to be kept still for the language or discourse variety represented by a corpus. In fact,

Approaches that focus on different discourses need to acknowledge that the concept of discourses as discrete and separate entities is problematic. Discourses are constantly changing, interacting, merging, reproducing and splitting off from each other. Therefore a corpus-based analysis of any discourse must be aware that it can only provide static snap-shots that give the appearance of stability but are bound to the context of the data set (Baker 2005: 17)

A corpus-assisted extraction of collocates can serve several practical purposes in addition to being a pivotal means of spotlighting ideology and evaluation in text or discourse. For example, it can aid in tracing the representation of groups or individuals, such as Baker et al. (2013b) who explore the representation of Muslims in the UK press. It might also assist in digging into the collocational construction of certain lexical items (e.g. *climate change* and *global warming*) in order to unveil the potential meaning that readership might associate to the words or multi-word expressions (as in Fiammenghi and Pinnavaia 2019).

In this research, the corpus-assisted study of cultural keywords and meaning by collocation is bolstered with an ecological interpretation of the results.

#### **1.3.4.3. Corpus approaches to ecological analysis of discourse**

Corpus approaches to the ecological study of discourse have developed since the early 2000s (Alexander 2002, 2009) and they are still evolving (Poole 2019). Their potentiality is explained by Alexander (2018: 202), who states that ecological discourse analysis

can benefit from employing such methods as computer-generated concordances, where texts are electronically available, to specifically investigate environmental discourse. Together with quantitative counts the perusal of concordance items from particular texts may provide us with more explicit data about a writer's categorical scheme. We thus set out to see how specific linguistic features are associated with or serve to uphold larger discourse processes, such as evaluation, argumentative strategies and discourse tactics. This enables the analyst

to support empirically what readers otherwise just infer concerning the ideological or principled positions which speakers or writers adopt.

Corpus approaches to ecological discourse analysis inherit most of their theoretical and methodological background from corpus approaches to critical discourse studies. As with corpus approaches to critical discourse studies, also in this case corpus linguistics offers a valid and elaborated methodological assistance to the quantitative and qualitative analysis of texts while ecological discourse analysis provides with a robust theoretical framework and with countable categories of ecological investigation that can be assessed with the aid of corpora (Alexander 2018).

Notwithstanding, even though Alexander's (2018) approval of corpus approaches to ecological discourse analysis resembles the defence of corpus approaches to critical discourse studies in content and strength, the actual employment of these approaches is scant in comparison with other strands of corpus-assisted discourse analysis. Much ecolinguistic research, in fact, is carried out without the use of corpora (e.g. Caimotto and Molino 2011, Stibbe 2012). At the same time, corpus studies on environmentally engaged discourse are sometimes carried out without any reference or allegiance to the ecolinguistic perspective (e.g. Bevitori and Johnson 2017, Potts et al. 2015, Sealey 2018).

What is more, some studies that define themselves as corpus approaches to the ecological analysis of discourse are only superficially corpus-aided, as in the case of the very Alexander (2018). They deploy basic methods of corpus analysis (e.g. frequency) and they do not engage in more sophisticated quantitative research. Nevertheless, there are also positive examples of this synergetic methodology: Poole (2016a, 2016b), for instance, studies the discussion on the Rosemont Copper Mines in corporate discourse with the aid of corpus-assisted keyword analysis and he praised the extensive use of corpora for ecological analysis of discourse. At the same time, Castello and Gesuato (2019) study Pope Francis's encyclical letter *Laudato si'* in order to portray how the religious document unfolds and how it handles environmental issues from an ecological perspective.

Not only is the current corpus-assisted study of cultural keywords and meaning by collocation conducted within an ecological framework, but it does also revolve around a cross-linguistic approach to the study of discourse.

#### **1.3.4.4. Corpus approaches to cross-linguistic studies of discourse**

Corpus approaches to cross-linguistic studies of discourse consist in “the application of the theory and methods of both corpus linguistics and discourse analysis to multilingual data” (Freake 2012: 3) and they aim at identifying lingua-cultural similarities, differences and peculiarities thanks to the aid of corpora (Taylor 2014).

The identification of lingua-cultural similarities, differences, and peculiarities posits some challenges. Taylor and Del Fante (2020) recognise these challenges at the stages of corpus design, planning of corpus analysis, and interpretation. As far as corpus design is concerned, Taylor and Del Fante (2020: 34) observe that corpus approaches to cross-linguistic studies usually involve comparable corpora that the authors call “multiple monolingual sub-corpora”; they are, in fact, monolingual corpora planned and constructed with the same design criteria so that the results obtained by analysing them with the same methods can be easily compared (e.g. Del Fante 2018; in Taylor and Del Fante 2020). These cross-linguistic studies can also be performed on parallel corpora or on multilingual corpora “which contain multiple languages within the same documents/document sets” (e.g. Freake et al. 2011; in Taylor and Del Fante 2020: 35). Taylor and Del Fante (2020) notice that designing such corpora requires to reflect on the existence and availability of comparable text types in the languages under inquiry and on the identity of the search terms to be entered when retrieving this data. In relation to the planning of corpus analysis, the authors distinguish between studies that act on the lexical level, those that operate above the lexical level, and those that investigate discourse analytic categories. The studies that focus on the lexical level adopt a corpus-driven approach and they let the corpora guide the analyst in search for similarities, differences, and peculiarities; these studies resort to corpus linguistic concepts like frequency, statistical keywords, and collocation, and they require an attentive identification of translational equivalents in different languages. The studies that expand on an above-the-lexical level disengage with the concreteness of word frequency and statistical counts; they deal with more abstract linguistic aspects like semantic categorisation, and they force the researchers to compare semantic and pragmatic associations. Research on discourse analytical categories assesses the way discourse and rhetorical frames are employed cross-linguistically and they compel to find and systematise frames before analysing them across languages. Problems with interpretation can depend on “attributing any variation identified in cross-linguistic discourse studies to the variables of language or culture when other factors could have intervened”; it is thus fundamental that “the

researcher be sufficiently familiar with the context of production to enable them to restrain interpretation and expand the analysis where required” (Taylor and Del Fante 2020: 37). This is a matter of triangulation.

At the same time, Vessey (2013) identifies four main challenges for corpus approaches to the cross-linguistic study of discourse. The first is related to the kind of corpora that should be used. The author observes that comparable monolingual corpora bear with them a whole set of social and demographic values related to the status of the investigated languages. Thus, Vessey (2013: 21) believes that “to adequately compare single-language corpora that index diverse populations, as much information as possible must be gathered regarding corpus composition and population demographics”. In addition, “particular attention should be paid to minority discourses, which may be further marginalised by cross-linguistic comparison” (Vessey 2013: 21). Then, “essentialisation and reification of categories” refers to the fact that sometimes essentialised and reified differences across corpora “are attributed to the different medium of communication – rather than, for example, internal diversity or context-specific differences” (Vessey 2013: 10). Another challenge in cross-linguistic, corpus-assisted studies of discourse appears when looking for the semantic prosody of translation equivalents (for a discussion of semantic prosody see § 1.2). Being semantic prosody an inherently cultural matter, “when comparing discourses across bilingual (or multilingual) corpora, familiarity with each corpus medium is essential and the semantic prosodies of translation equivalents cannot be presumed” (Vessey 2013: 14). Finally, Vessey (2013) notices the difficulties in the extraction of statistical keywords and the problems with the identification of suitable comparator corpora.

The tradition of these studies is less well-established than that of the generalist corpus approaches to discourse analysis (Vessey 2013, Taylor and Del Fante 2020). Cross-linguistic corpus-assisted studies have come to the fore only recently, starting with works by Freake (2012), Freake et al. (2011), Taylor (2014), and Vessey (2013). So far, some studies have been produced within SiBol’s CADS school (e.g. Partington et al. 2013, Taylor 2014, 2016) or by other sparse researchers (Baker and Vessey 2018, Bayley and Williams 2012, Vessey 2016). Taylor (2016) compares patterns of politeness and impoliteness in two internet forums written in English and Italian using a corpus-assisted methodology. Baker and Vessey (2018) compare English and French extremist texts to investigate common and divergent linguistic and thematic strategies in the construction of extremist discourse. Bayley and Williams (2012) explore the representation of the European identity in the media published in several languages

(e.g. English, French, Italian, and Polish). Vessey (2016) investigates the construction of language ideology in the Canadian press by comparing English and French news discourse.

Taylor (2014) writes that corpus approaches to the cross-linguistic study of discourse align to one of three tendencies. According to the first tendency, “the researcher focuses very explicitly on the language differences (and similarities)” (Taylor 2014: 373). The second tendency observed by Taylor (2014: 373) witnesses a comparative interest that is “largely cultural” so that “the research is cross-linguistic out of necessity”; in this kind of studies, “the fact that the corpora are in different languages is almost irrelevant”. The third tendency in corpus approaches to cross-linguistic studies is characterised by “no explicit comparative drive” (Taylor 2014: 373). Taylor and Del Fante (2020) add a fourth category for the classification of cross-linguistic corpus-assisted studies of discourse. They write that some approaches deal with the comparison of cultural keywords or discourse keywords so that “by adopting a lexicological approach, the focus of the analysis is concerned with the investigation of the discourse usage and functions of a word or a set of words and the subsequent comparison across cultures and languages” (Taylor and Del Fante 2020: 33).

The first approach is typical of traditional corpus-aided studies in contrastive linguistics, that are generally shaped on the needs of translation studies (among others, Altenberg and Granger 2002a, Ebeling and Ebeling 2013, Johansson 2007, Johansson and Oksefjell 1998, Mikhailov and Cooper 2016). Altenberg and Granger (2002a) collect a series of theoretical and applied chapters on corpus-based studies of lexis in contrast. Ebeling and Ebeling (2013) produce a theoretical and empirical work on the contrastive study of linguistic patterns across English and Norwegian translational equivalents. Mikhailov and Cooper (2016) outline the main points that involve corpus linguistics in translation and contrastive studies.

While adopting an approach to the study of language that is similar to the one employed by corpus-assisted discourse studies, most of this contrastive research differs in goals and kind of social and political engagement with the matter in comparison with cross-linguistic analyses of discourse. In fact, while the former focus on similarities and differences in the forms and functions of the linguistic choices adopted in a set of languages for the sheer purpose of contrast and comparison, the latter commit with the historical, social and political issues at stake in discourse and they try and combine an overview of linguistic matters with a broader commentary on the contextual reasons that justify pinpointed linguistic choices (Taylor and Del Fante 2020, Vessey 2013).

Another issue that needs to be considered is the availability of previous corpus-assisted, discourse-analytical research on the languages under inquiry, both taken singularly and contrastively. Some languages are overwhelmingly represented while others are neglected; some languages are frequently compared whereas others are rarely employed for cross-linguistic studies (Vessey 2013).

Starting with a monolingual approach, among the languages considered for the present research (i.e. English, Hungarian and Italian), English is the focus of most corpus analyses of discourse; also Italian is well represented while Hungarian seems to be quite neglected, at least internationally (but see, for instance, Barát 2017).

The study of statistical keywords for discourse analytical purposes in English is well-documented in edited works like Bondi and Scott (2010) and it can be found, for instance, in Jeffries and Walker (2017), who adopt a corpus-assisted approach to search for cultural keywords in the discourse of New Labour, or in Zottola (2019: 461), who “focuses on the representation of transgender people in the British press as social actors”. Model Italian studies of statistical keywords are Spina (2014), who examines the statistical keywords emerging in political discourse spread through Twitter, or Castello and Gesuato (2019), who research on the aboutness of the Pope’s encyclical *Laudato si* with the extraction of key terms.

The importance of collocational patterns for discourse analytical purposes is highlighted for English in works like Partington (1998). Collocational patterns in English are researched on by authors like Balfour (2019), who outlines the representation of people affected by schizophrenia with the use of collocations, or by Fusari (2016), who adapts the corpus-assisted extraction of collocational patterns to the investigation of the European debt crisis within a systemic-functional perspective. The theory of collocation in Italian and for Italian is accounted for by authors like Spina (2001, 2010) and Rossini Favretti (1998, 2002), and it is exemplified by works like Formato (2019), who discusses the use of gendered language in Italian political and news discourse.

Continuing with a cross-linguistic approach, corpus-assisted analyses of discourse seem to favour exclusively the comparison of English and Italian, excluding Hungarian. Cross-linguistic studies of statistical keywords comparing English and Italian can be found, for instance, in Bassi (2010), who writes on the statistical keywords popping up in Italian and American news discourse on the Kyoto protocol. Cross-linguistic studies of collocational patterns can be read in Tognini-Bonelli (2002), but also in Aragrande (2018), who compares the representation of migrants in British and Italian news broadcasting channels, Lo Cascio (1997), who attempts to build a

dictionary of collocations for lexicographical purposes that includes also English and Italian, Rossini Favretti et al. (2007), who contrast legal discourse through collocates.



## 2. Data

The corpus-assisted analysis of cultural keywords and meaning by collocation is put into practice in the present study in order to outline the discursive construction of sustainable development in discourse. The second chapter describes the data used for the research. This includes the 2030 Agenda Corpus, a multilingual parallel corpus, and the Sustainable development Corpus, a multilingual comparable corpus. For both corpora, I detail first the design and annotation issues and then I sketch the characteristics of the collections.

### 2.1. The 2030 Agenda Corpus

The 2030 Agenda Corpus is a multilingual parallel corpus consisting of three documents: the English version of the United Nations' *Transforming our world: The 2030 Agenda for Sustainable Development* (United Nations 2015a); the Hungarian version of the same document (*Világunk átalakítása: Fenntartható Fejlődési Keretrendszer 2030*; United Nations 2015b); the Italian agenda (*Trasformare il nostro mondo: l'Agenda 2030 per lo Sviluppo Sostenibile*; United Nations 2015c).

The corpus is a multilingual parallel corpus because it comprises the same text translated into three languages (English, Hungarian, and Italian). The direction of the translation is not easy to determine, as it frequently happens in relation to documents released by the European Union. Teubert (2002: 213), for example, writes that the "EU documents as a parallel corpus are unique in the sense that it is not possible to distinguish source language from target language". Teubert's (2002) statement about the EU documents might be easily applied also to the UN documents that shape the 2030 Agenda Corpus. The 2030 Agenda Corpus, in fact, incorporates resolutions similar to the EU documents: the UN resolutions are produced in an official, international context comparable to the one that originated the aforementioned EU reports.

#### 2.1.1. Corpus collection and annotation

For the construction of the corpus, the three versions of the Agenda are downloaded in PDF format and then they are converted into plain text files with the AntFileConverter software (Anthony 2017a). AntFileConverter is a tool developed by

Laurence Anthony “to convert PDF and Word (DOCX) files into plain text for use in corpus tools” (Anthony 2017a). After the conversion the text files are carefully checked to make sure that they correspond to the original document.

The conversion of the documents from .pdf to .txt format saves headers and additional footnotes that depend on the editorial choices of the documents’ translators and publishers and that do not match throughout the corpus. Therefore, the files are cleared of those parts that do not correspond across languages in order to allow for better alignment of the texts. The English and Italian versions of the agenda are similar in format and they are both polished by eliminating the header. This includes page number, the label A/RES/70/1 (corresponding to Resolution 70/1), and the title *Transforming our world: The 2030 Agenda for Sustainable Development* for the English document and *Trasformare il nostro mondo: l’Agenda 2030 per lo Sviluppo Sostenibile* for the Italian document. In the Hungarian version of the agenda, page numbers and the table of contents are erased. Footnotes marked by a number or by an asterisk are moved at the end of the document so as not to split sentences. The number of the footnotes is erased for the same reason.

The three documents of the 2030 Agenda Corpus are aligned at sentence level with the AntPConc software (Anthony 2017b). AntPConc software is a “parallel corpus analysis toolkit for concordancing and text analysis” and it was developed by Laurence Anthony to assist in the alignment and analysis of parallel corpora (Anthony 2017b). All sentences of the English document are matched with the corresponding sentences of the Hungarian and Italian documents. This is done to ease the retrieval of correspondences at the level of words and sentences.

Apart from sentence alignment, the corpus is utilised in raw form, without any annotation in order to suit the requirements of one of the software employed for the analysis (i.e. #LancsBox; Brezina et al. 2020). #LancsBox is a software package developed at Lancaster University for the analysis of language data and corpora. It has been selected for the current study because of its potentialities for the extraction of collocations. In its current version, #LancsBox supports annotation only for English and Italian, while it does not allow to study annotated files for Hungarian.

### **2.1.2. Corpus features**

The English document *Transforming our world: the 2030 Agenda for Sustainable Development* consists of 15,858 tokens and 2,435 types, as it can be seen in Table 1. The Hungarian document *Világunk átalakítása: Fenntartható Fejlődési Keretrendszer 2030*

counts 14,915 tokens and 4,076 types. The Italian document *Trasformare il nostro mondo: l'Agenda 2030 per lo Sviluppo Sostenibile* includes 17,945 running tokens and 3,264 types.

	2030 Agenda Corpus (English)	2030 Agenda Corpus (Hungarian)	2030 Agenda Corpus (Italian)
Word types (n°)	2,435	4,076	3,264
Tokens (n°)	15,858	14,915	17,945
MATTR	0.6804	0.7373	0.7257

**Table 1.** Number of word types, number of tokens, and MATTR of the English, Hungarian, and Italian versions of the 2030 Agenda.

Considering the differences in the number of types and tokens across languages, the lexical richness of the document is calculated with *lexical diversity* statistics (Jarvis 2013). As Brezina (2018a) observes, lexical diversity can be expressed through three major statistical counts: type-token ratio (TTR), *standardized type-token ratio* (STTR), and *moving average type-token ratio* (MATTR).

Type-token ratio is simply the ratio between the number of types and the number of tokens included in a corpus. It is a valuable statistical measure for esteeming a corpus's richness in vocabulary and it is frequently expressed through a percentage. The closer the percentage to 100%, the richer the lexical diversity of the corpus (McEnery and Wilson 2001). TTR, however, depends on corpus size and it can make cross-corpus comparisons difficult (Brezina 2018a).

The standardized type-token ratio, on the other hand, allows such comparisons by considering the TTR of various chunks of the corpus. Scott (2015), for instance, explains the mode of functioning of the STTR by referring to the WordSmith Tools software, where the measure was first launched. Scott (2015) writes that STTR

is computed every n words [...]. By default, n = 1,000. In other words the ratio is calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of your text or corpus. A running average is computed, which means that you get an average type/token ratio based on consecutive 1,000-word chunks of text. (Scott 2015: 303)

Another measure that improves the computation of lexical diversity is the moving average type-token ratio. Convington and McFall (2010: 96) write that while STTR “is computed on successive non-overlapping segments of the text [...] MATTR uses a smoothly moving window”. While applying MATTR, the corpus is divided into fixed-

size subsequent windows, gently drifting from the first to the last token of the corpus; TTR is calculated for every window and MATTR is the result of the computation of the average of those subsequent TTRs. MATTR seems thus “better for tracking changes within texts” and it is “not affected by accidental interactions between segment boundaries and text unit boundaries” (Covington and McFall 2010: 96). This makes it more suitable than the previous measures to be employed with corpora comprising more than a text.

MATTR is computed for the three versions of the 2030 Agenda using the #LancsBox software, with windows of 100 words. As it can be seen in Table 1, the MATTR of the English document is 0.6804. The Hungarian version of the Agenda displays a MATTR of 0.7373. MATTR for the Italian Agenda is 0.7257. In other terms, this count progressively increases from the English Agenda to the Italian and then until the Hungarian.

The growth of MATTR from English to Hungarian can be tentatively interpreted in two ways. The first relates to the essence of translation and the second to purely linguistic reasons. First, according to Baker (1995), the type-token ratio of a translated text should be lower than that of its original version. This is explained by Laviosa (2002) as a form of lexical impoverishment, in that translated texts are generally less rich in lexis compared to their original counterparts. Following Baker’s (1995) and Laviosa’s (2002) observations, the highest value of MATTR for the Hungarian and Italian versions of the document would hint at their being the source texts for translation. The English version of the 2030 Agenda displays a lower value for MATTR, which in Baker (1995) and Laviosa (2002) is a clue for considering it as a translation equivalent of the UN’s resolution. Nevertheless, it seems unlikely that the Hungarian and Italian agendas might be the original versions of the document. In fact, the preferred linguistic medium of official, international meetings such as the one that ended with the publication of the Agenda is usually English and the official languages of the United Nations are Arabic, Chinese, English, French, Russian, and Spanish<sup>5</sup>. Therefore, I argue that among the English, Hungarian, and Italian documents, the only original is the English text, while the Hungarian and the Italian texts are translations.

Second, the growth of the value of MATTR from English to Hungarian might be adduced as a reflex of morphological differences across the three languages (on this see § 3.5). In the Hungarian agenda, MATTR is higher than in the English and Italian

---

<sup>5</sup> A list of the official languages of the United Nations can be found at <https://www.un.org/en/our-work/official-languages> (last accessed on 29<sup>th</sup> March 2021).

documents: at first sight, this might suggest that the former is rendered with a richer vocabulary; however, such a difference in the ratio between types and tokens could be due to the fact that the documents are analysed in plain text format, which leads the software to count all the inflected forms of a lemma as different types. To this extent, in fact, English and Hungarian are morphologically rather different: while English words are only lightly inflected, Hungarian morphology is more elaborated. For instance, while the English lemma *COUNTRY* is found only in two forms (*country* and *countries*), one of its Hungarian translational equivalents, i.e. *ORSZÁG*, is declined as *ország* ‘country’, *országaink* ‘our countries’, *országban* ‘in (the) country’, *országok* ‘countries’, etc. In this case, #LancsBox counts only two types for English and more than four for Hungarian, making the MATTRs in the two languages meaningfully diverse. The MATTR for the Italian text is closer to the Hungarian than to the English figure. This similarity might lie in that morphological difference that has already been mentioned or in a wider variety of translational equivalents in the rendering of some lexemes (for more on this see § 3.5).

## **2.2. The Sustainable development Corpus**

The Sustainable development Corpus (SusCorp in short) is a specialised, multilingual, and comparable corpus of British, Hungarian, and Italian news discourse on sustainable development, published after the release of the 2030 Agenda. The corpus is built on purpose for the current study.

The corpus is built on purpose for the current study and it incorporates texts written in three languages (English, Hungarian, and Italian). The articles of SusCorp are collected from the most read national quality papers published in the United Kingdom, in Hungary, and in Italy. The articles included in the corpus are published between 1<sup>st</sup> January 2016 and 31<sup>st</sup> December 2018 in the print or in the web version of national daily broadsheets and they all relate to sustainable development by mentioning it.

The Sustainable development Corpus counts 1,500 articles and 982,444 running tokens. The corpus is divided into three sections, namely *SusCorp (English)*, *SusCorp (Hungarian)*, and *SusCorp (Italian)*. The three sections are comparable in quality and size: they all include articles sampled with the same design criteria (see § 2.2.1 in this chapter on this), and they count the same number of texts and a similar number of tokens (see § 2.2.3 in this chapter on this).

Plain text articles make up the core of the corpus. These articles, however, are also saved in annotated form. Texts are annotated with light descriptive metadata to easily scan through some key discursive characteristics of the texts (e.g. author and date of issue) and with a word-level linguistic annotation (tokenisation, POS-tagging, and lemmatisation) that could help during the analysis of the texts.

### 2.2.1. Corpus design

The Sustainable development Corpus is designed to represent news discourse published in the most read British, Hungarian, and Italian national daily broadsheets. Traditional newspapers are chosen as sources for the corpus because of their playing a paramount role in the construction, reconstruction, or confirmation of ideologies (see § 1.3 on ideology). The relationship existing between newspapers and their readers in the creation and reproduction of ideology, in fact, is bidirectional and dynamic: ideologies expressed in newspapers might be gathered and reproduced by their customary readers; or, on the way round, the ideologies of the readers might induce newspaper writers to adhere to them so as to sell more (cf. among others, Fowler 1991 and Gabrielatos and Baker 2008).

This bidirectional and dynamic relation is reflected in the linguistic choices adopted by newspapers. Therefore, an analysis of the linguistic patterns associated with a debated issue can interestingly aid to capture the view that readers might have gathered on the matter (Fowler 1991; see § 1.3 on this). In relation to the present study, recurrent linguistic patterns emerging in British, Hungarian, and Italian news discourse on sustainable development might spotlight the ideologically charged construction that the press and their readership assign to sustainability in the United Kingdom, in Hungary, and in Italy. In particular, the statistical and cultural keywords scattered in this kind of discourse highlight the most salient topics associated with sustainable development by newspapers after the publication of the 2030 Agenda; the collocational patterns that surround *sustainable development* and other lexical items related to sustainability boost the semantics of the issue in the press (see § 4 for a thorough explanation of these points).

The newspapers considered for the Sustainable development Corpus are selected according to circulation figures. Among the most read newspapers, quality papers are preferred, i.e. newspapers publishing articles about serious issues and printed on large sheets in opposition to newspapers focusing on light matters with a glossy layout and small dimensions (i.e. tabloids; Bednarek and Caple 2012).

The decision to exclude tabloids is one of opportunity and comparability. In terms of opportunity, a pilot survey conducted prior to the collection of SusCorp shows that the issue of sustainable development is only scantily dealt with by tabloids. In 2016, for instance, the most distributed British tabloid (*Metro*) mentions *sustainable development* only 5 times while the most popular British broadsheet (*The Times*) quotes the same expression in 59 texts.

In addition, as far as comparability is concerned, the reason for choosing quality newspapers only depends on the fact that the distinction between quality papers and tabloids is almost meaningless among the Italian newspapers, while it is clear-cut in the British and in the Hungarian daily press panorama. In Italy, in fact, the only two daily papers that might be deemed tabloids are the freely distributed *Leggo* and *Metro (Italia)*.

*Leggo* and *Metro (Italia)* reach a more limited readership in comparison with the most popular Italian daily papers. A survey carried out by Audipress<sup>6</sup>, for instance, shows that between July and September 2018, *Leggo* was read daily by 556,000 people on average whereas *Corriere della Sera*, the most printed daily broadsheet, was bought by 2,034,000 readers. On the contrary, circulation figures show that in the UK selling rates are surely higher for tabloids in comparison with broadsheets and that the most popular broadsheet in terms of circulation is read less than the ninth most popular British tabloid<sup>7</sup>.

Because of this cultural difference, if both broadsheets and tabloids were welcomed in the corpus with the criterion of circulation still holding true, it would be difficult to discern how to balance the presence of quality and tabloid papers in the three collections. In the British and Hungarian cases, the selection of the most read daily papers would be skewed towards tabloids, hardly including broadsheets, while in the Italian case, no tabloids would be comprised.

Consequently, the selection of the daily newspapers within the list of the three countries' quality papers is determined by their circulation rates and the Hungarian press panorama is established as baseline because of its featuring only four national

---

<sup>6</sup> Audipress is a privately owned company which collects and distributes data on the circulation figures of daily, weekly, and monthly papers and magazines in Italy. Audipress's survey of the Italian reading habits between 2017 and 2018 can be found at the URL <http://audipress.it/pubblicati-dati-dellindagine-audipress-2018-ii/> (last accessed on 29<sup>th</sup> March 2021).

<sup>7</sup> Press Gazette is a privately owned media trade magazine dealing with journalism and the press. Press Gazette's survey of the British reading habits at the end of 2018 can be found at the URL <https://www.pressgazette.co.uk/national-newspaper-abcs-telegraph-y-o-y-circulation-decline-slows-as-bulk-sales-distortion-ends/> (last accessed on 30<sup>th</sup> March 2021).

broadsheets overall (i.e. *Magyar Hírlap*, *Magyar Nemzet*, *Népszava*, and *Világgazdaság*). Articles from all these Hungarian quality papers are incorporated in the corpus.

As far as the press in the United Kingdom is concerned, according to a survey of the Audit Bureau of Circulation<sup>8</sup> that assessed the circulation figures of the British national newspapers in December 2018, the five most read daily broadsheets of the UK are *The Times*, *The Daily Telegraph*, *i* (a newspaper that developed from *The Independent*), *Financial Times*, *The Guardian* (in decreasing order of circulation; see also Hill 2016 for the circulation of newspapers in the United Kingdom). Among these, the four newspapers selected are *The Times*, *The Daily Telegraph*, *Financial Times*, and *The Guardian*. *Financial Times* is included in the corpus so that the English section of the SusCorp can be comparable to the Hungarian: the Hungarian broadsheet *Világgazdaság*, in fact, is a business newspaper as *Financial Times*. *The Guardian* is preferred to *i* after a pilot search for articles mentioning sustainable development in the newspapers: the topic is reported in the former much more than in the latter.

The design of the Italian subcorpus follows the circulation figures of the Italian national newspapers in December 2018 according to the Italian company ADS<sup>9</sup>. Only the three most popular national newspapers are included in the Italian section of the Sustainable development Corpus, namely *Corriere della Sera*, *La Repubblica*, and *La Stampa*. The business newspaper *Il Sole 24 Ore* is added to these to save comparability with the British and Hungarian sections of SusCorp.

All articles extracted from the British, Hungarian, and Italian newspapers appeared after the publication of the 2030 Agenda. The document was published on 25<sup>th</sup> September 2015 after a meeting organised by the United Nations. Thus, it is deemed appropriate to collect texts that were issued from the very beginning of 2016 to the end of 2018 in order to track how the most read British, Hungarian, and Italian newspapers dealt with the issue of sustainable development also in relation to the United Nation's Agenda.

The articles retrieved for the Sustainable development Corpus explicitly refer to sustainable development.

---

<sup>8</sup> The aforementioned survey of the British newspapers' circulation figures at the end of 2018 was conducted by the Audit Bureau of Circulation and it was published by the media trade magazine Press Gazette. It can be found at the URL <https://www.pressgazette.co.uk/national-newspaper-abcs-telegraph-y-o-y-circulation-decline-slows-as-bulk-sales-distortion-ends/> (last accessed on 30<sup>th</sup> March 2021).

<sup>9</sup> The aforementioned survey of the Italian newspapers' circulation figures at the end of 2018 was conducted by the Italian company ADS and it could be found at the URL [http://www.adsnotizie.it/dati\\_DMS.asp](http://www.adsnotizie.it/dati_DMS.asp) (last accessed on 6<sup>th</sup> February 2019).

## 2.2.2. Corpus collection and annotation

To guarantee that all articles included in the corpus explicitly mention sustainable development, they are found with the query terms *sustainable development* for English, *fenntartható \*fejlődés\** for Hungarian, and *sviluppo sostenibile* for Italian. The Hungarian query term *fenntartható \*fejlődés\** is characterised by two wildcards that help capture variants: the first wildcard allows to obtain occurrences of the lemma *FENNTARHATÓ FEJLŐDÉS* when it is modified by an interpolated adjective (e.g. *fenntartható gazdasági fejlődés* ‘economic sustainable development’, in the nominative); the second wildcard enables the identification of all inflectional variants of the lemma (e.g. *fenntartható fejlődést* ‘sustainable development’, in the accusative).

These seeds are chosen with an iterative procedure that follows the study of the 2030 Agenda, whose detailed account can be read in § 4, and after the construction of pilot corpora of British, Hungarian, and Italian news discourse. The founding hypotheses of the seeds’ collection are the following: is *sustainable development* referred to only explicitly or through cognates like *sustainability*? Is it also alluded to with other terms (e.g. synonyms, near-synonyms, or pronouns)? To test these hypotheses, the most frequent and statistically significant lexical items of the 2030 Agenda are considered, partly following Gabrielatos’s (2007) method. The aim of this is to guarantee both recall and precision by incorporating in the corpus the highest number of articles dealing with sustainable development without lingering on those writings that feature the pinpointed search terms without regarding sustainable development at all.

The sets of seeds gathered for the collection of the English, Hungarian, and Italian pilot corpora are equivalent across the three languages. The lexical items that appear to be most linked to the concept of sustainable development in the English Agenda are *2030 Agenda*, *resilient*, *sustainable development*, and *sustainable*. These lexical items are selected as seeds for the pilot corpus by going through the most frequent and most significant words of the 2030 Agenda (a thorough presentation of the analysis of the 2030 Agenda can be read in § 4). Also *resilience* and *sustainability* are added to the list of seeds after a manual examination of the document and with an introspective integration of the results. For Hungarian and Italian, the corresponding translational equivalents are adopted in the first stage of corpus planning and collection. These seeds are used as query items to retrieve and collect newspaper articles from the selected British, Hungarian, and Italian broadsheets.

An exploratory analysis of the resulting pilot subcorpora shows that some of the seeds employed for the collection of newspaper articles refer to topics that are totally unrelated to sustainable development. *resilient*, for example, is frequently used to describe the attitude of sportsmen, as in

Chelsea were rugged and resilient across the capital at Crystal Palace on Saturday and courtesy of their first headed goal of term, produced by Diego Costa just before the interval, secured a club record 11<sup>th</sup> league win in succession (*The Guardian*, 18<sup>th</sup> December 2016).

At the same time, *sustainability* frequently hints at the feasibility of economic practices, as in

If Rome's growth bet fails to pay off, the concerns of financial markets about the sustainability of its debt may be violently reawakened (*Financial Times*, 9<sup>th</sup> November 2018)

Articles featuring these seeds often do not include mention to sustainable development.

As a consequence, the seeds for the collection of SusCorp are cut down to the sole *sustainable development* for English and the corresponding *fenntartható \*fejlődés\** for Hungarian and *sviluppo sostenibile* for Italian as the goal of the present research is the investigation of the semantics of sustainable development and of the statistical and cultural keywords that help construct it discursively. This cutting down of seeds ensures that all texts included in the corpus mention sustainable development at least once.

The English section of the Sustainable development Corpus is constructed using the LexisNexis Academic<sup>10</sup> database. LexisNexis Academic is the academic component of LexisNexis, a database that stores vast collections of legal, business and news discourse, gathered especially in the United States but produced also in Europe and in other areas of the world.

The LexisNexis database comes in several versions, apt for companies, lawyer firms, universities, etc. Among these versions, the Academic database is considered to be particularly useful for research carried out within academia and it includes material

---

<sup>10</sup> The interface for the exploration of the LexisNexis Academic database can be found at the URL <https://www.lexisnexis.com/hottopics/lnacademic/> (last accessed on 30<sup>th</sup> March 2021).

classified as *News, Legal, International Legal, Companies, and People, Places and Things*. It provides a wide range of texts issued by media sources such as newspapers or newswires, but also writings composed in legal contexts (like trials), or documents released by companies or about companies.

The News section of LexisNexis Academic is further divided into *All News, Foreign Language News, and Broadcast Transcripts*. The *All News* section includes news discourse delivered worldwide, especially in English. It can be accessed through a platform that has to be interrogated with a list of seeds. The platform permits to filter the results of a query according to the date of issue of the searched articles, according to source (as one is looking for a specific newspaper), according to source type (e.g. newspapers, magazines, blogs, etc.), or to a specific article type and geographic location. The *Foreign Language News* incorporates news published in several languages (i.e. Arabic, Danish, Dutch, Finnish, French, German, Italian, Malay, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, Turkish, and other). Its platform filters the material only for date and language. The *Broadcast Transcripts*, finally, is a collection of transcripts of the *All English Language News Group*, that features channels like *ABC, CNN*, etc. The database *Broadcast Transcripts* can be interrogated by filtering the transcripts only for date and source.

The English section of SusCorp is compiled by querying the *All News* store. *All News* is explored in search for all articles published between 1<sup>st</sup> January 2016 and 31<sup>st</sup> December 2018 on a selected list of broadsheets (*Financial Times, The Daily Telegraph, The Guardian, The Times*) and featuring the seed *sustainable development*.

The archives with the chosen newspapers are explored separately and the searches are split for periods of six months in order not to exceed the number of texts that can be retrieved at once. This, in fact, would cause LexisNexis Academic to automatically select the texts according to their relevance with the consequent spoiling of the representativeness of the corpus. Then, the texts retrieved with such a query are downloaded as text files and they are randomly selected to reach a total number of 500 articles. Any duplicate articles coming out of the stratified random selection are removed.

The search with LexisNexis Academic, in fact, produces a list of articles that include single, duplicate, and quasi-duplicate documents. Duplicate articles are recognised whenever two articles are published with the same headline and the same body or with the same body but different headlines. These articles are either published the same day or in different occasions. Moreover, they frequently appear in different versions of the newspaper, being it the national in opposition to the local or the print

opposed to the web version of the paper. Quasi-duplicates are signalled in case the bodies of two articles differ only because of minor changes. These changes might be due to the length of the texts and to their relationship: for example, an article might be 600 words long and the other 830, having the former at its core. Quasi-duplicate articles are maintained. In fact, as Dayrell (2019) observes for the Brazilian press, newspapers tend to publish similar texts repeatedly, by simply adding minor changes.

The initial collection of texts is therefore cleared by excluding all duplicate articles. In the exclusion of the articles, if the duplicates are published the same day, the first in the list is the one to be selected. If they are published in different dates, the first to appear is chosen among the range. At the same time, in case the duplicate articles appear in different editions of the same newspaper (e.g. the national, the international, and the local one), the national edition is preferred over the others. Moreover, when the same article is published both in the print and in the web version of the newspaper, the print is chosen because of its being already cleaned off metadata reproducing the hyperlinks that can be seen in the webpage from which the article was extracted. Hyperlinks are always cleared from the texts.

Duplicates are excluded to avoid that the repeated presence of the same article could skew the results of statistical analyses of the corpus. It might be argued, nevertheless, that the presence of duplicates reflects the editorial panorama more faithfully than their absence. This is certainly true if one is interested in studying the impact that the repeated publication of news dealing with a certain topic has on the readers of a newspaper. On the other hand, if one is interested in examining the linguistic strategies that surround that topic irrespective of their impact, a clean corpus might fit the purpose better. It would, in fact, offer a clearer picture of the variety of linguistic choices that span throughout the years in a newspaper.

The same process is followed also for the collection of a part of the Italian subcorpus (i.e. *Corriere della Sera* and *La Stampa*). This subsection of SusCorp is constructed through the LexisNexis platform, from the *Foreign Language News* section. In this case, since it is not possible to query the stores of the newspapers separately but for language, the database is searched for all articles published in Italian between 1<sup>st</sup> January 2016 and 31<sup>st</sup> December 2018 that include in their heading or in their body the search term *sviluppo sostenibile* 'sustainable development'. Then, only those coming from *Corriere della Sera* and *La Stampa* are singled out and saved in separate .txt files. These files are polished of duplicate articles.

Duplicate articles are excluded also from the list of the texts collected manually for the remaining part of the Italian subcorpus (i.e. *Il Sole 24 Ore* and *La Repubblica*) and

for the whole Hungarian subcorpus (i.e. *Magyar Hírlap*, *Magyar Nemzet*, *Népszava*, and *Világgazdaság*).

These two components of the Sustainable development Corpus are constructed by directly exploring the websites of the newspapers. The newspapers' websites are queried in search for all texts of the paper's archive that include the seeds *fenntartható \*fejlődés\** 'sustainable development' for Hungarian and *sviluppo sostenibile* 'sustainable development' for Italian and that were published between 1<sup>st</sup> January 2016 and 31<sup>st</sup> December 2018. Texts are then downloaded in .txt format and stored in the corresponding folder. The same storage procedure is followed also for the articles saved through LexisNexis. 500 articles per language are sampled from the whole list of texts through stratified sampling. The comparability of the number of articles saved for every section of SusCorp contributes to the comparability of the size of every subcorpus.

The articles are stored according to three parameters: language, broadsheet, and date of issue. They are labelled by mentioning first the language they are written in (*En* for English, *Hun* for Hungarian, and *It* for Italian), second, the newspaper they are published on (as in Table 2), and, finally, the date they were released with the order of year, month, and day (YYYYMMDD, e.g. 20160812 for an article appeared on 12<sup>th</sup> August 2016).

SusCorp (English)		SusCorp (Hungarian)		SusCorp (Italian)	
Name of the broadsheet	Label	Name of the broadsheet	Label	Name of the broadsheet	Label
<i>Financial Times</i>	FT	<i>Magyar Hírlap</i>	MH	<i>Corriere della Sera</i>	CS
<i>The Daily Telegraph</i>	TDT	<i>Magyar Nemzet</i>	MN	<i>La Repubblica</i>	LR
<i>The Guardian</i>	TGU	<i>Népszava</i>	NSz	<i>La Stampa</i>	LS
<i>The Times</i>	TT	<i>Világgazdaság</i>	VG	<i>Il Sole 24 Ore</i>	SO

**Table 2.** Labels for the texts in the English, Hungarian, and Italian subcorpora.

In case more than one article was published on the same day, its publication rank is signalled as an increasing cardinal number going up to two figures. The parameters are divided by a full stop, as it might be seen in the following exemplary storage label:

En.TG.20170526.04.txt

The label stands for “article appeared in the English broadsheet *The Guardian* and published on 26<sup>th</sup> May 2017 as fourth article of the day among those collected”.

The Sustainable development Corpus is stored both in raw and in annotated format. In its annotated version, it is marked with descriptive metadata.

Metadata are defined by Burnard (2005: 40) as “data about data”, or, in other terms, as descriptive and interpretative information about the origin and content of the corpus. According to Burnard (2005), metadata are extremely important when it comes to the study of the corpus. In fact, they provide contextual notions that the analyst could hardly retrieve without indication and that they could deploy for the interpretation of the data.

Descriptive metadata, that Burnard (2005) distinguishes from editorial and analytic metadata, encode information on the social context that produced the texts included in the corpus. They might list features like the date of issue of the document or its author. Both in the case of big and small corpora, descriptive metadata offer contextual details that might enrich the comparison between texts.

Descriptive metadata of the Sustainable development Corpus contain the following traits:

- *language* of the text;
- *name of the newspaper* the text has been retrieved from;
- *section* where the article is found;
- *date* of issue of the article;
- *author* of the article;
- *heading* of the article;
- *body* of the article.

The language of the texts (English, Hungarian, and Italian) is included in the annotation because of SusCorp’s multilingual and comparable nature: specification of every text’s language is deemed essential for cross-linguistic investigations of the collection.

The name of the newspaper is mentioned so that analyses of the linguistic idiosyncrasies or similarities that unfold in the various papers could be accessible; as Bednarek and Caple (2012) notice, in fact, stylistic and thematic patterns might change significantly among news sources.

The section where the article is found is noted to observe trends in the indexing of articles mentioning sustainable development.

The article's date of issue is annotated for two main reasons. First, it is included for potential diachronic evaluations of the texts: the articles could be studied across time to trace the evolution and change that might have occurred to the linguistic patterns that emerge when dealing with sustainable development in the news. Second, it could prove helpful to associate any peaks in the use of certain lexical items to extraordinary events and dates as Gabrielatos (2007) does in relation to his corpus of news discourse about refugees and asylum seekers.

At the same time, keeping track of the author of the articles might aid to exclude that very frequent linguistic patterns are highlighted by statistical counts only because of a single person's idiolect and not as a general trend in news discourse about sustainable development.

The articles are then divided into heading and body. This might shed light on the distribution of linguistic patterns in two very different parts of the articles: while the heading presents the topic in a way that tries to be catchy in order to make the reader stick to the text, the body develops at length the theme introduced in the heading. A research on the linguistic strategies adopted in these very different parts of the articles might be very fruitful (Bednarek 2006).

After an overview of the annotation standards used within the community of corpus linguists and computational linguists and listed, for instance, by Lehmborg and Wörner (2009) and by Ide et al. (2017), I have decided to employ the scheme put forward by Hardie (2014b) as *Modest XML* (Hardie 2014b). XML, or *eXtensible Markup Language*, is a mark-up system that was created as an extension of the previously created *Standard Generalised Markup Language* (SGML) and that encodes mark-up within angle brackets. It witnesses a wide variability of attributes and tags provided they conform XML's grammar (Bray et al. 2008). In the present research, the texts are annotated as in the following example:

```
<text id="I.LS.001">
  <header language="it" newspaper="La Stampa" section="" author="Paolo Mastroiilli"
    date="2017/09/23">
    <head>
      I big del mondo riuniti da Bloomberg fra crescita e sviluppo sostenibile
    </head>
    <body>
      «Voi siete qui perché, in un modo o nell'altro, credete che la
      moltiplicazione sia una strategia superiore alla divisione. Perché credete
      che l'addizione sia migliore della sottrazione, in economia,
      nell'inclusione sociale e nella politica.
      [...]
    </body>
  </header>
```

As it can be seen from the example, articles are labelled as text in general and the aforementioned metadata are all inserted in the header with the tags `language`, `newspaper`, `author`, `date`, `head`, and `body`. The attributes of language are “en” for English, “hun” for Hungarian, “it” for Italian. The name of the newspaper is mentioned in full as for the section and the author of the article. The date is included by typing the year first, followed by month and date of issue of the text.

Other annotation standards such as the *Text Encoding Initiative*, or *TEI*, have also been considered because of their being widespread within the community of computational linguists and corpus linguists and because of their offering enormous opportunities for the annotation of text (Sperberg-McQueen and Burnard 2019). The complexity and time-expense of annotating with TEI, though, has been judged excessive for an addition of descriptive metadata only intended at preserving contextual information that might be precious during the interpretation of the results.

The Sustainable development Corpus is also annotated linguistically on word level. SusCorp is first tokenised, then tagged for parts-of-speech, and finally lemmatised. Only letters and numbers are considered when defining token classes, while punctuation, symbols, and mark-up are not counted. POS-tagging is automatically carried out on the Sketch Engine platform. POS-tagging for English is conducted with the Modified English TreeTagger part-of-speech tagset applied to the TreeTagger developed by Helmut Schmit at the University of Stuttgart. POS-tagging for Hungarian is rendered available by the emMorph-based part-of-speech tagset. POS-tagging for Italian is carried out with the TreeTagger tool and Marco Baroni’s tagset. POS-tagging paves the way for the lemmatisation of the corpus. Also lemmatisation is achieved within the Sketch Engine environment.

### **2.2.3. Corpus features**

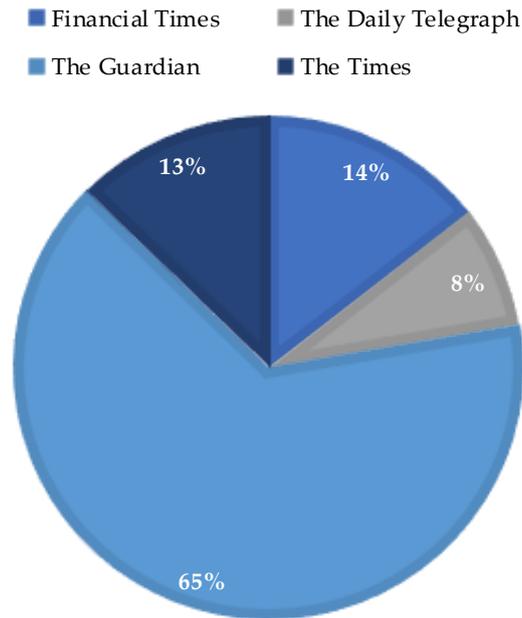
As it can be seen in Table 3, the English section of SusCorp counts 26,416 word types and 422,545 running tokens; its MATTR is 0.7483338. The Hungarian section of SusCorp consists of 43,515 word types and 257,789 tokens, with a MATTR of 0.7914573. The Italian section of SusCorp is made of 28,600 word types and 302,101 tokens, and it features a MATTR of 0.7922072.

	SusCorp (English)	SusCorp (Hungarian)	SusCorp (Italian)
<b>Word types (n°)</b>	26,416	43,515	28,600
<b>Tokens (n°)</b>	422,545	257,798	302,101
<b>MATTR</b>	0.7483338	0.7914573	0.7922072
<b>Average sentence length (n° tokens)</b>	23.345028	20.176723	26.34755
<b>Average text length (n° tokens)</b>	845.090	515.596	604.202

**Table 3.** Number of word types, number of tokens, MATTR, average sentence length, and average text length of the English, Hungarian, and Italian sections of SusCorp.

As in the case of the 2030 Agenda corpus, also in SusCorp the MATTR of the three subcorpora differs. The lowest value is found for the English section, followed by the Hungarian and the Italian sections. This variable MATTR is paired and explained by the different number of word types and tokens running in the subcorpora. Taking MATTR's lowest and highest values, for instance, it can be observed that the English section of SusCorp features a lower number of word types compared to the Italian, while the number of tokens is higher in English. Italian's higher number of word types depends also on the higher inflectional variability of the language. English's high number of tokens, on the other hand, depends on the length of the articles included in the corpus: English texts roughly include 845.090 words on average, while the average text length of the Italian subcorpus is 604.202 words. Nevertheless, the average sentence length is higher in Italian compared to English and Hungarian: in SusCorp, Italian sentences consists of roughly 26 words on average, English sentences count 23 words on average, whereas Hungarian sentences include an average number of 20 words. The English subcorpus is characterized by the longest texts on average, the Italian subcorpus is marked by longest sentences on average, while the Hungarian subcorpus witnesses the shortest text and sentence length.

The distribution of SusCorp's tokens among the English, Hungarian, and Italian newspapers can be seen in Graphs 1, 2, and 3. Graph 1 shows that 65% of the tokens in the English subcorpus is to be found in *The Guardian* section, which counts 301 articles and 275,164 words; 14% of the tokens belongs to the *Financial Times* section, which includes 79 articles and 61,144 words; 13% of the tokens comes from *The Times* section, which is made of 74 articles and 53,227 words; only 8% of the tokens is included in *The Daily Telegraph* section, which witnesses 46 articles and 33,010 words.

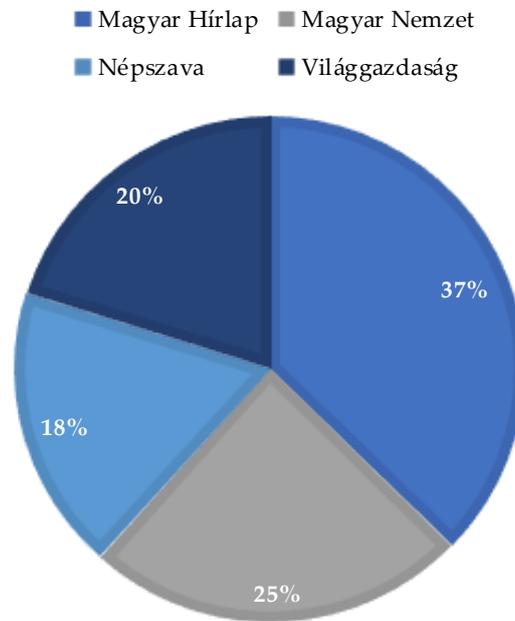


**Graph 1.** Distribution of tokens across newspapers in SusCorp (English).

*The Guardian* is by far the most represented broadsheet in the collection both in terms of number of articles and in terms of number of tokens per article. In fact, its section in SusCorp (English) comprises roughly four times as many articles as those included in the *Financial Times* section, namely the second most represented broadsheet. This depends on the different coverage of the issue of sustainable development in the four quality papers: *The Guardian* writes about sustainable practices the most, and it is followed by *Financial Times*, *The Times*, and *The Daily Telegraph*. The political orientation of the newspapers and their economic, social, and environmental interests explain this order: *The Guardian* is a centre-left oriented broadsheet particularly keen on touching upon environmental matters, *Financial Times* is a liberal, financial paper devoted to economic issues, while *The Times* and *The Daily Telegraph* are conservative papers that tend to neglect environmental and economic concerns if they are compared to *The Guardian* and *Financial Times*. In addition, as far as article length is concerned, while *The Guardian* counts 914.166 words per article on average, the average article length for the other three newspapers ranges between 717.609 and 773.975 words. Articles in *The Guardian* are significantly longer than in the other papers, especially because they frequently consist of reports or transcripts of meetings and conferences.

Graph 2 reports the distribution of SusCorp's tokens across the Hungarian newspapers: 37% of the tokens comes from the *Magyar Hírlap* section, which consists of 207 texts and 96,173 words; 25% of the tokens cluster in the *Magyar Nemzet* section,

which is made of 92 articles and 62,666 words; 20% of the tokens belongs to the *Világgazdaság* section, which includes 123 articles and 51,935 words; 18% of the tokens can be found in the *Népszava* section, which counts 78 texts and 46,994 words.

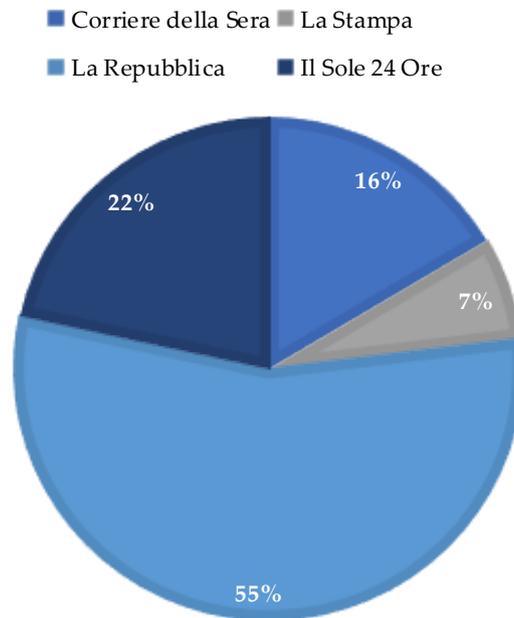


**Graph 2.** Distribution of tokens across newspapers in SusCorp (Hungarian).

*Magyar Hírlap* is the most represented broadsheet of the collection in terms of number of articles and tokens, although words are more evenly distributed across SusCorp’s (Hungarian) newspapers compared to the British panorama. Furthermore, also the average length of the articles is more evenly distributed in the Hungarian press compared to the British. The longest articles can be usually read in *Magyar Nemzet* and *Népszava*: their average length is of 681.152 and 602.487 words respectively. The shortest articles are often published in *Magyar Hírlap* and *Világgazdaság*, with an average length of 464.603 and 422.236 words respectively. The average number of tokens per article in the Hungarian quality papers is usually lower than in the English newspapers. In addition, similarities in terms of political orientation are not always matched in similarities in words’ distribution: the broadsheet with the longest articles is the national conservative *Magyar Nemzet*, and the left-oriented *Népszava* comes only second; it is followed by the conservative *Magyar Hírlap* and by the financial paper *Világgazdaság*, which features the shortest articles.

As it can be seen in Graph 3, most tokens of the Italian section of the Sustainable development Corpus come from the newspaper *La Repubblica*, which includes 55% of

the words of SusCorp (Italian) by counting 277 articles and 173,380 running tokens. *La Repubblica* section is followed by the *Il Sole 24 Ore* section, which contains 22% of SusCorp’s (Italian) tokens as it encompasses 87 articles and 67,732 running tokens. The sections *Corriere della Sera* and *La Stampa* follow, with 16% and 7% of the subcorpus’s words respectively; the *Corriere della Sera* section consists of 92 articles and 51,718 tokens, whereas the *La Stampa* section is made of 44 articles and 20,610 tokens.



**Graph 3.** Distribution of tokens across newspapers in SusCorp (Italian).

This distribution of tokens across the Italian broadsheets resembles to some extent the distribution of tokens across the British broadsheets encompassed in SusCorp. The most represented quality paper of the Italian collection is the left-oriented *La Repubblica*, which appears to refer most frequently to environmental issues in general and to sustainable development in particular. Because of its importance in the corpus and because of its political orientation, the Italian *La Repubblica* can be compared to the British *The Guardian*, which proves to be the most represented British quality paper in SusCorp (English). Unlike *The Guardian*, however, *La Repubblica* does not feature the longest articles among the group of Italian broadsheets. In fact, the articles in *La Repubblica* newspaper are 625.921 words long on average and their length is exceeded by the length of *Il Sole 24 Ore*’s articles, which count 778.529 words on average. The second most represented Italian broadsheet is the financial newspaper *Il Sole 24 Ore*, followed by the centre-oriented *Corriere della Sera* and *La Stampa*. *Corriere della Sera* and

*La Stampa* are less represented than *La Repubblica* and *Il Sole 24 Ore* both in terms of number of tokens and as far as the average article length is concerned: the articles of *Corriere della Sera* are approximately 562.152 words long and the articles of *La Stampa* are 468.409 words long on average. Unlike the British and the Hungarian cases, the newspapers in SusCorp (Italian) are almost exclusively centre or centre-left oriented, apart from the financial paper *Il Sole 24 Ore*.



### 3. Methodology

The third chapter describes the methodology employed in the present work for the study of the discursive construction of sustainable development in the 2030 Agenda Corpus and in the Sustainable development Corpus, which have just been introduced (§ 2). After resuming the research questions listed in the introduction, I focus on the methods and tools adopted when exploring the 2030 Agenda Corpus and the Sustainable development Corpus in search for cultural keywords (for more on cultural keywords see § 1.1). I also summarize the rationale used to identify additional lexemes related to economy, society and environment. Then, I account for the methodology and the tools selected to study the meaning by collocation of the lexical items *SUSTAINABLE DEVELOPMENT*, *SUSTAINABLE*, and *SUSTAINABILITY* and their Hungarian and Italian translational equivalents (on meaning by collocation see § 1.2 and § 1.3).

#### 3.1. The research questions

As it has been stated in the introduction, the present study aims at exploring the discursive construction of sustainable development in the British, Hungarian, and Italian versions of *The 2030 Agenda for Sustainable Development* (2030 Agenda Corpus) and in British, Hungarian, and Italian news discourse on sustainable development (SusCorp). The discursive construction of sustainable development is investigated by answering the following research questions. First, in relation to the 2030 Agenda:

1. What are the cultural keywords emerging in the English, Hungarian and Italian versions of the 2030 Agenda? And how are the economic, social and environmental dimensions of sustainable development lexically represented in the document?
2. What meaning do *SUSTAINABLE DEVELOPMENT*, *SUSTAINABLE*, and *SUSTAINABILITY* and their Hungarian and Italian translational equivalents acquire through collocation in the UN's resolution?

Second, regarding news discourse

3. What are the cultural keywords emerging in the English, Hungarian, and Italian news discourse on sustainable development?
4. What meaning does the English *SUSTAINABLE* and its Hungarian and Italian corresponding lexical items acquire through collocation in news discourse?

These research questions are answered under the light of the linguistic theories on cultural keywords (Williams 1983) and on meaning by collocation (Sinclair 1991). Cultural keywords provide a strong connection between discourse and the world by lexically reproducing concepts that are meaningful for the society where the discourse under inquiry was generated. In addition, not only do cultural keywords provide links between discourse and the world, but they also contribute to the unfolding of connections within discourse by means of cohesion: they disseminate discourse with repeated, culturally paramount lexemes.

The semantics of these cultural keywords, as well as of words like the English *SUSTAINABLE*, *SUSTAINABILITY*, *SUSTAINABLE DEVELOPMENT* and their Hungarian and Italian counterparts, can be assessed by means of collocation. Collocation showcases the lexemes that habitually co-occur with a pinpointed word; colligation organises these habitually co-occurring words to outline the grammatical behaviour of the lexical items under inquiry; through semantic preference, the collocates cluster into specific semantic areas that reproduce the events, people, things that contribute to the semantics of the investigated lexemes; thanks to evaluative prosody, this semantics contributes to clarify the ideological assumptions that underlie the discursive construction of a notion by imbuing the lexeme with a connotation that can be positively or negatively judged according to the ideological framework adopted for the analysis.

In the present work, cultural keywords and meaning by collocation are analysed and interpreted with a corpus-assisted approach (Taylor and Marchi 2018; see § 1.3.4) and from the perspective of ecological discourse analysis (Alexander and Stibbe 2014; see § 1.3.2).

Ecological discourse analysis offers a theoretical and interpretative background to study discourse within an ecological framework. The ecological framework consists in a personal, structured opinion about ecological matters; this structured opinion gathers its tenets from well-known ecological theories. Within the ecological framework adopted in this study and presented in this chapter (§ 3.6), cultural keywords provide clues to salient or erased elements. In other terms, they foreground elements that are deemed worthy of consideration, and they background elements that are deemed unworthy of consideration. Foregrounding and backgrounding are interpreted under the light of the ecological framework of the study. Also the evaluative prosody highlighted when searching for meanings by collocation is estimated under the light of the ecological framework of the study in order to assess whether the ideologies hidden in the semantics of the lexemes can be positively or

negatively evaluated in relation to the ecosophy of the author. The semantics of the lexemes are positively judged when they might be regarded beneficial for the well-being of the environment, while they are negatively judged when they might be regarded detrimental for the well-being of the environment.

The corpus-assisted approach to the study of cultural keywords and of meaning by collocation strengthens the qualitative observations made within the theoretical framework of ecological discourse analysis with quantitative, empirical data. Cultural keywords can be signalled among the most frequent or among the statistically most significant lexemes of a corpus, namely through frequency lists and keyword lists, and meaning by collocation can be extracted through collocation networks. Both keywords and collocational patterns can be further explored with the aid of concordances (for more on corpus approaches to discourse see § 1.3.4)

This corpus approach to the study of cultural keywords and meaning by collocation within the framework of ecological discourse analysis is carried out from a cross-linguistic perspective. Cultural keywords and meaning by collocation are explored across three languages (i.e. English, Hungarian and Italian) to observe whether sustainable development is discursively constructed in a comparable way or if patterns of frequency and statistical salience change cross-linguistically.

### **3.2. The identification of cultural keywords**

The first step in this corpus-assisted, ecological study of the discursive construction of sustainable development across languages consists in the retrieval of the cultural keywords that characterize the corpora under inquiry. Cultural keywords are found among the most frequent and among the statistically significant lexemes of the corpora. The most frequent lexemes are esteemed paramount for the collection of cultural keywords because they mirror the topics that are most frequently dealt with in the corpora. The statistically significant lexemes are deemed essential for the identification of cultural keywords because they reflect the topics that concern the corpora in a statistically significant way in comparison with other corpora; as Stubbs (2010: 23) claims, statistical keywords are “the tips of the iceberg: pointers to complex lexical objects which represent the shared beliefs and values of a culture”.

### 3.2.1. The most frequent lexemes

The most frequent words of the corpora are found in frequency lists (or wordlists), namely lists of all the word types running in the corpora. The items of a wordlist are generally ranked according to frequency of occurrence. This ranking empowers researchers to understand the grammatical and lexical peculiarities of the corpus by showing first the corpus's most frequent function and content words; this opens a breach towards the understanding of the corpus's style and aboutness (Phillips 1989; see also Hoover 2009). In the case of the 2030 Agenda Corpus, the wordlists frame the grammatical and lexical features of the United Nations' resolution on sustainable development. In the case of the Sustainable development Corpus, the frequency lists help identify the most frequent grammatical and lexical items of news discourse on sustainable development published in some of the most widely read British, Hungarian, and Italian quality papers between 2016 and 2018.

The most frequent word types of the corpora are computed with AntConc's Word List tool (Anthony 2020). Only high-ranking word types are considered for further analysis. High-ranking word types are distinguished via the notion of *frequency bands*: high frequency bands comprise the words that are used the most; high frequency bands slip over middle frequency bands in the precise ranking position where two word types have the same frequency; low frequency bands incorporate rare word-forms and *hapax legomena*, i.e. those items occurring only once in a corpus (Freddi 2014).

Among high-frequency word types, function words are separated from content words and only the latter are examined in detail. Content words, in fact, are isolated as representative of the themes that were addressed the most in the English, Hungarian and Italian sections of the 2030 Agenda Corpus and of SusCorp.

### 3.2.2. The statistically significant lexemes

The statistically significant word types of the corpora are identified through the keyword lists. Keywords are collected by comparing the frequency lists of the corpora with the frequency lists of other corpora chosen as reference corpora.

The keyword study combines an *exploratory* and a *focused* approach (Gries 2010, Partington 2009). The exploratory approach is one in which no hypothesis is formulated *a priori* on the quality of the words that appear statistically more or less

significantly than expected in the study corpus. The focused approach is one in which particular attention is paid to linguistic items that prove particularly interesting for the analyst (Gries 2010, Partington 2009). I believe that this exploratory-focused dichotomy reflects the opposition between corpus-driven and corpus-based approaches, and, to some extent, it can be regarded as a mirror of one of the basic distinctions between statistical keywords and cultural keywords. While both reach society and culture, the former are collected with a bottom-up, exploratory, or corpus-driven approach, while the latter result from a top-down, focused or corpus-based approach. Both concepts are thus employed in the current study: statistical keywords assist in the exploration of the lexico-grammatical profile of the corpora, while cultural keywords are gathered introspectively to conduct an in-depth research on the topic of sustainable development.

For the keywords' computation in the 2030 Agenda Corpus, the English, Hungarian and Italian versions of the document are compared to the English, Hungarian and Italian sections of a corpus of proceedings of the European Parliament (*EUROPARL7*, Koehn 2005). *EUROPARL7* is a corpus that collects the minutes of the European Parliament and that is deemed as "a highly valuable resource for various fields of contrastive linguistics and translation studies" (Cartoni et al. 2013: 39). Carapinha and Plag (2019), for example, employ *EUROPARL7* to study the function and the German translational equivalents of the Portuguese discourse marker *na verdade*. Ghivirigă (2020) explores *EUROPARL7* to focus on the options for the translation of English gerunds into Romanian.

The comparison between the frequency lists of the 2030 Agenda Corpus and the frequency lists of the *EUROPARL7* corpus allows to highlight the specificities in aboutness and style of the United Nations' 2030 Agenda compared to another specimen of international, official political discourse. The *EUROPARL7* corpus has been deemed suitable for this purpose for several reasons. First, the genre of *EUROPARL7* is comparable to the 2030 Agenda's: *EUROPARL7* is made of edited transcripts of debates held at the European Parliament; the 2030 Agenda is a resolution issued as the written account of a meeting held by the United Nations. They can be both regarded as instances of political discourse. This renders it more likely for differences in the two corpora's word frequencies to be the result of different topics being dealt with rather than of different stylistic choices dictated by the genre. In addition, there seem to be no problems also as far as medium is concerned: as Zufferey and Cartoni (2012: 238) observe, in fact, the "language used in the [*EUROPARL7*] corpus is intermediate between speech and written language, as deputies' statements

are spoken during the session and these statements are later transcribed and edited". At the same time, the 2030 Agenda was issued as a written product, sharing consequently lexical and grammatical features with the partially written nature of EUROPARL7. Moreover, EUROPARL7 and the English, Hungarian and Italian versions of the 2030 Agenda can be regarded as examples of parallel corpora: they are all official publications of proceedings with their translational equivalents. This suggests that the idiolects used for writing the EUROPARL7 proceedings and the 2030 Agenda might be similar in the three languages. Furthermore, the two corpora are almost contemporary: the 7<sup>th</sup> version of EUROPARL that was used for this study collects transcripts of parliamentary meetings held until 2011, only four years before the 2030 Agenda was published.

Some features of the English, Hungarian and Italian sections of the EUROPARL7 corpus can be seen in Table 4. The features were gathered using AntConc's Word List tool.

	<b>EUROPARL7 (English)</b>	<b>EUROPARL7 (Hungarian)</b>	<b>EUROPARL7 (Italian)</b>
Texts (n°)	9,672	8,763	9,486
Word types (n°)	88,282	291,987	159,392
Tokens (n°)	57,885,375	13,374,057	55,656,092

**Table 4.** Number of texts, number of word types, and number of tokens of the English, Hungarian, and Italian sections of the EUROPARL7 corpus.

As it can be noted in Table 4, the English subsection of the EUROPARL7 corpus counts 9,672 texts that feature 88,282 word types and 57,885,375 tokens. The Hungarian subcorpus includes 8,763 texts with 291,987 word types and 13,347,057 running tokens. The Italian subsection of EUROPARL7 consists of 9,486 texts, counting 159,392 word types and 55,656,092 tokens.

For the 2030 Agenda Corpus, keywords are extracted by comparing the English, Hungarian and Italian versions of the 2030 Agenda (study corpus) with the corresponding components of the EUROPARL7 corpus (reference corpus). Keywords are computed with AntConc's Keyword List tool (Anthony 2020).

As Gabrielatos (2018) and Pojanapunya and Watson Todd (2018) recommend, in the calculation of keywords, a statistical significance test is paired to an effect-size test to strengthen the weight of the keywords within the 2030 Agenda Corpus. Thus, keywords are calculated with the Log Likelihood (4 term) statistical significance test

( $p < 0.0001$ ) together with the Difference Coefficient (Relative) effect-size test. As Rayson et al. (2004) suggest, the statistical threshold is set at 15.13 for the Log Likelihood statistical test; in fact, such a threshold is necessary for the results of keyword analysis to be at the same time robust and accurate. As far as the minimum frequency of occurrence of the keywords in the corpus is concerned, the frequency threshold is fixed to 5. A frequency threshold of less than 5, in fact, might result in the extraction of weak keywords while a threshold of more than 5 might leave out meaningful word types. The three versions of the 2030 Agenda, in fact, count only between 14,493 and 17,510 words and a 5-times-occurring term makes already for the 0.029-0.034% of the document.

Keywords are extracted for the 2030 Agenda Corpus and for the SusCorp with two different software or platforms (i.e. AntConc and Sketch Engine respectively) and with slightly different statistical measures (i.e. a combination of Log Likelihood and Difference Coefficient for the 2030 Agenda Corpus and the default simple maths for the SusCorp). This depends on feasibility reasons. In fact, while the EUROPARL7 corpus can be easily handled with offline software like AntConc, no comparable reference corpus of valuable size could be found for the SusCorp for a remote calculation of keywords. Furthermore, a calculation of the 2030 Agenda's keywords with the use of the EUROPARL7 corpora available on the Sketch Engine platform shows that no significant difference can be observed when the keywords are extracted using the Log Likelihood statistical significance test together with the Difference Coefficient (Relative) effect-size test in AntConc and using the simple maths statistical test in Sketch Engine.

Thus, the keywords of the Sustainable development Corpus are computed through the Sketch Engine platform because of the reference corpora available there. They are extracted by comparing the frequency list of the English, Hungarian and Italian sections of the SusCorp (study corpus) with the frequency lists of the English, Hungarian and Italian sections of the *TenTen corpora* first and of the *Timestamped JSI webcorpus 2014-2020* second (reference corpora). The TenTen corpora are a family of corpora "created from the Web" and "prepared according to the same criteria" for every language they include, so that they "can be regarded as comparable corpora"; they are built "using technology specialized in collecting only linguistically valuable web content" (on the TenTen corpora see Jakubíček et al. 2013). Consequently, the TenTen corpora represent the language of the Web in a precise period. The comparison between the frequency lists of the SusCorp and the frequency lists of the TenTen corpora allows to emphasise the specificity in aboutness and style of online-published

news discourse on sustainable development compared to a general use of language in the Web. The TenTen corpora have been regarded as suitable reference corpora for the SusCorp because of their encompassing a wide amount of textual material drawn from the Web and deriving from a large variety of sources. The TenTen corpora, in fact, consist of all linguistically valuable content published on the Web in a certain year and thus they cover such a broad range of genres, registers, and text types that the corpora can be valued as representative of Web language. In addition, the valuable size of the TenTen corpora strengthens the robustness of the statistical comparisons done when extracting statistical keywords. The TenTen corpus employed as a reference corpus for the English section of the SusCorp is the *English TenTen2012*, the one adopted as a reference corpus for the Hungarian section of the SusCorp is the *Hungarian TenTen2012*, and the corpus utilized as a reference corpus for the Italian section of the SusCorp is the *Italian TenTen2016*. As it can be seen in Table 5, the English TenTen 2012 corpus (also *enTenTen 2012*) includes 22,223,066 texts, 27,894,538 word types, and 12,968,375,937 tokens; the Hungarian TenTeb 2012 corpus (also *huTenTen 2012*) counts 6,447,178 documents, 28,127,413 word types, and 3,161,920,362 tokens; the Italian TenTen 2016 corpus (also *itTenTen 2016*) is made of 12,967,535 texts, 17,119,528 word types, 5,864,495,700 tokens.

	<b>enTenTen 2012</b>	<b>hunTenTen 2012</b>	<b>itTenTen 2016</b>
Texts (n°)	22,223,066	6,447,178	12,967,535
Word types (n°)	27,894,538	28,127,413	17,119,528
Tokens (n°)	12,968,375,937	3,161,920,362	5,864,495,700

**Table 5.** Number of texts, number of word types, and number of tokens of the English, Hungarian and Italian sections of the TenTen corpus.

The second reference corpus for the SusCorp, namely the Timestamped JSI webcorpus (also called *JSI newsfeed*) “is a clean, continuous, real-time aggregated stream of semantically enriched news articles from RSS-enabled sites across the world” (Bušta et al. 2017: 1). The corpus is collected with a news aggregator that scans the web daily in search for news feeds; the news aggregator collects news feeds, it clears them, and it renders them accessible for research (Trampuš and Novak 2012). This news aggregator has been collecting articles since 2014 and it has produced collections that include data from 2014 to 2016, from 2014 to 2020, and from 2020 on. For the present research, the Timestamped JSI webcorpus 2014-2020 is preferred over the others. Moreover, the Timestamped JSI webcorpus is available in several languages, including English, Hungarian and Italian.

Being the Timestamped JSI webcorpus a collection of news appeared in the Web, the comparison between the frequency lists of the SusCorp and the frequency lists of the Timestamped JSI webcorpus underlines the aboutness and style of online-published news discourse on sustainable development compared with general Web news discourse. In fact, the Timestamped JSI webcorpus 2014-2020 has been selected as the second reference corpus because of its being comparable to SusCorp on many extents. First, the genre of the SusCorp and of the Timestamped JSI is the same, since they both include newspaper articles discovered in online archives and repositories. Second, the sources are also comparable, although the Timestamped JSI includes texts scraped from a wider variety of websites compared to the SusCorp. The corpora contain documents published roughly in the same period: the SusCorp's articles appeared between 2016 and 2018 and the articles of the Timestamped JSI were aggregated between 2014 and 2020. In addition, the Timestamped JSI webcorpus has been considered as the most suitable reference corpus due to its generic content and valuable size. Although specialised in genre, in fact, the Timestamped JSI can be regarded as a general reference corpus because of its opening to documents on a wide range of matters. This enables to compare the SusCorp to a corpus that is not skewed towards any particular topic. This might result in a better highlighting of the words that are typical of the news discourse on sustainable development. In terms of size, the Timestamped JSI webcorpus includes a very meaningful number of word types and tokens, as it can be seen in Table 6. For instance, the English Timestamped JSI webcorpus 2014-2020 consists of 148,531,127 documents counting 53,106,755,084 word types and 61,833,890,155 tokens. The Hungarian corpus consists of 3,059,143 documents, 714,951,341 word types, and 864,172,414 tokens. The Italian corpus is made of 22,573,257 texts with 6,509,458,717 word types and 7,691,374,904 running tokens.

	<b>Timestamped JSI webcorpus 2014-2020 (English)</b>	<b>Timestamped JSI webcorpus 2014-2020 (Hungarian)</b>	<b>Timestamped JSI webcorpus 2014-2020 (Italian)</b>
Texts (n°)	148,531,127	3,059,143	22,573,257
Word types (n°)	60,275,417	7,808,453	7,040,194
Tokens (n°)	61,833,890,155	864,172,414	7,691,374,904

**Table 6.** Number of texts, number of word types, and number of tokens of the English, Hungarian and Italian sections of the Timestamped JSI webcorpus 2014-2020.

The keywords of the Sustainable development Corpus are identified on the Sketch Engine platform with the simple maths measure, described by Kilgarriff (2009). The statistical threshold is set to 10.0. Statistical keywords are searched for in the form of lemmas with the TenTen corpora and in the form of word types with the Timestamped JSI 2014-2020 corpora. The Timestamped JSI 2014-2020 corpora, in fact, do not allow to retrieve lemmas for all languages. Unlike the keywords extracted from the 2030 Agenda Corpus, the SusCorp's keywords are calculated with a frequency threshold of 20, as the number of tokens of the English, Hungarian and Italian sections of the SusCorp is between 17 and 28 times as big as the number of tokens of the English, Hungarian and Italian sections of the 2030 Agenda Corpus. A frequency threshold lower than 20 might spotlight rare words.

When extracting statistical keywords, only the statistically most frequent types of the corpora (namely positive keywords) are taken into consideration. Positive keywords, in fact, allow to list salient contents. Among positive keywords, proper nouns are focused on because of their referential function: they introduce the protagonists of statements and actions towards sustainable development. Other content words are considered as carriers of the aboutness of the texts, of the most significant topics that the texts deal with. Also function words are included in the study as indicative of the stylistic choices that characterise the corpora and that help the discursive construction of sustainable development.

### **3.2.3. Economic, social and environmental lexemes**

In addition, the lexemes reproducing the economic, social and environmental dimensions of sustainable development are pinpointed in the 2030 Agenda Corpus. The lexemes belonging to these dimensions assist in tracing the focus of the 2030 Agenda and in better understanding the political, social and cultural engagement of the United Nations' resolution in relation to sustainable development.

The economic, social and environmental lexemes are found among the most frequent and statistically significant lexemes of the 2030 Agenda Corpus, but they are also identified through the whole list of words included in the document. The small size of the corpus, in fact, allows to scan through its complete wordlist. Lexemes are grouped according to their reproducing the economic, social or environmental dimensions of sustainable development. In the case of the social and the environmental dimensions of sustainable development, four pivotal groups are

recognized for living creatures and non-living elements: human beings, animals, plants, and natural elements.

Only content words are considered for this classification. Moreover, in the case of human beings, animals, plants and natural elements, concrete nouns are preferred over abstract ones: so, for example, while *child* can be inserted in the group of humans, *childhood* would not be accepted because of its referring to an abstract entity.

### 3.3. The reading of concordance lines

The study of the most frequent and statistically significant lexemes of the 2030 Agenda Corpus, as well as the analysis of the corpus's economic, social and environmental lexemes is complemented by inquiring further on the broader semantic patterns they belong to. As Baker (2006) recommends, concordances and collocations (on this see § 3.4 in this chapter) are explored to identify the semantic patterns of the lexemes.

A concordance “is a collection of the occurrences of a word-form, each in its own textual environment” (Sinclair 1991: 32). It allows to zoom in on the co-text surrounding all the instances of a lexical item and it enables to peruse the relationship between a word-form and its usual co-text.

Concordance lines are generally displayed in *Key Word in Context* format (or *K.W.i.C.* format), where the “word-form under examination appears in the centre of each line, with extra space on either side of it” (Sinclair 1991: 33). Concordance lines are usually listed according to the order of appearance of the lines within the corpus. They can also be listed so that the words appearing either on the left or on the right of the node are sorted according to their alphabetical order (Freddi 2014).

When sorting to the left or to the right of a word, the recurrent lexical environment of a node is highlighted, and patterns become more easily observable. It is generally good practice to examine the concordances sorted in several ways, because different pieces of information about the search term usually emerge. Concordances sorted to the left, for instance, emphasize recurrent patterns that precede the node, while concordance lines sorted to the right spotlight recurrent patterns that follow the node (see, among others, Sinclair 1991 and Tognini-Bonelli 2001).

In this study, concordances are extracted with AntConc's Concordance tool and sorted to the left and to the right in relation to the type of analysis that needs to be performed time after time.

When analysing concordance lines including verbs, verbs are classified according to Halliday and Matthiessen's (2004) transitivity theory. The study of

transitivity patterns for certain lexemes aims at identifying the kind of processes that social and natural actors are involved in (Halliday and Matthiessen 2004; originally in Halliday 1994). In Halliday and Matthiessen's (2004) transitivity system, process types are classified into *material processes*, *mental processes*, *relational processes*, *behavioural processes*, *verbal processes*, and *existential processes*. Material processes report the outer experience of the world through "actions and events: things happen, and people or other actors do things, or make them happen" (Halliday and Matthiessen 2004: 170). They involve an *actor*, who is "the source of the energy bringing about the change" and who directs this energy to a *goal* (Halliday and Matthiessen 2004: 179). Mental processes reflect the inner experience of the world and they are "partly a kind of replay of the outer, recording it, reacting to it, reflecting on it, and partly a separate awareness of our states of being" (Halliday and Matthiessen 2004: 170). They are conveyed thanks to the interaction between a *senser* (i.e. a human-like participant "that 'senses' – feels, thinks, wants or perceives") and a *phenomenon* (i.e. "that which is felt, thought, wanted or perceived" (Halliday and Matthiessen 2004: 201-203). Relational processes are processes "of identifying and classifying" (Halliday and Matthiessen 2004: 170). They can be *attributive* and assign an *attribute*, namely a class, to a *carrier*, namely the entity to which the attribute is ascribed; they can also be *identifying* and associate an *identified* and an *indentifier*. Behavioural processes "represent the outer manifestations of inner workings, the acting out of processes of consciousness [...] and physiological states" (Halliday and Matthiessen 2004: 171). Behavioural processes concern a *beholder*, namely a conscious being "who is 'behaving'" and a *behaviour* (Halliday and Matthiessen 2004: 250). Verbal processes express "symbolic relationships constructed in human consciousness and enacted in the form of language, like saying and meaning", whose main participant is a *sayer* (Halliday and Matthiessen 2004: 171). Existential processes testify that "phenomena of all kinds are simply recognized to 'be' – to exist, or to happen" and those entities and events who are recognized to be are called *existents* (Halliday and Matthiessen 2004: 171).

### **3.4. The analysis of meaning by collocation**

After the identification of cultural keywords and of economic, social and environmental lexemes, their semantics is investigated either by reading their concordance lines, as it has just been stated, or by means of collocation. Meaning by collocation is sought also for the English lexical items *SUSTAINABLE*, *SUSTAINABILITY*, *SUSTAINABLE DEVELOPMENT* and their Hungarian and Italian translational equivalents,

namely *FENNTARTHATÓ*, *FENNTARTHATÓSÁG*, *FENNTARTHATÓ FEJLŐDÉS* for Hungarian, and *SOSTENIBILE*, *SOSTENIBILITÀ*, *SVILUPPO SOSTENIBILE* for Italian, in order to strengthen the study of the discursive construction of sustainable development.

Meaning by collocation is gathered with the retrieval of collocational patterns, as it is recommended in corpus lexicology (Moon 2010). Collocational patterns, in fact, signal words that co-occur with a lexical item significantly frequently. The habitual company that words keep to each other can be systematised in terms of content and connotation with the concepts of colligation, semantic preference, and evaluative prosody. Colligation highlights recurrent grammatical patterns; semantic preference guides a researcher in classifying collocates according to their semantic field; evaluative prosody eases the scholar in discovering evaluative contents concealed behind collocates and the discursive co-text they can be found in. Put together, they can assist in determining a word's semantics with corpus-aided techniques (Stubbs 2015).

Meaning by collocation is calculated first for the 2030 Agenda Corpus and then for the Sustainable development Corpus. As far as the Sustainable development Corpus is concerned, even though the texts sampled in the SusCorp span throughout a range of three years (from 2016 to 2018) and they could consequently be analysed both with a synchronic approach and with a diachronic approach (on the two approaches see Marchi 2018), in the current study they are considered as a unitary whole because the number of texts sampled for each year is not sufficient to carry out a diachronic analysis of the corpus. The same holds true also for a potential analysis of meanings by collocation across the SusCorp's newspapers: the size of the sampling for all newspapers is not enough to guarantee robust results for the extraction of collocational patterns. Thus, collocates are extracted from the whole corpus.

### **3.4.1. Word sketches**

Collocational patterns are retrieved first through one of Sketch Engine's functions, namely the *word sketch* function (Kilgarriff et al. 2014). A word sketch is "a one-page summary of a word's grammatical and collocational behaviour". Word sketches display the grammatical relations that tie a node to its collocates. For example, if the node is a verb, a word sketch will show the most peculiar subjects, objects, modifiers, and prepositional phrases of the verb, etc.; if it is a noun, it will show information on the modifiers of the noun, the nouns and verbs that are modified by the noun, the verbs that feature the noun as object or as subject, etc.

The word sketch function operates by collecting the collocates of a word and by analysing them through a word sketch grammar. A word sketch grammar is a grammar written in the CQL programming language that combines POS-tags and regular expressions to assign every token of a corpus to a specific grammatical relation as far as the selected node is concerned. The function adopts the logDice statistical measure to test the statistical significance of collocates. It operates better on big corpora or, at least, with a hundred occurrences of a word. Corpora used to extract word sketches are annotated morphologically by the Sketch Engine with POS-tagging and lemmatization, while syntactic parsing is not needed. Thus, the strength and weakness of word sketches depends on the precision of the morphological tagset used for the language under inquiry.

When searching for meaning by collocation, word sketches perform very well. In fact, according to Kilgarriff et al. (2014: 10), word sketches are useful in many ways, but especially in lexicography: a word sketch “can be seen as a draft dictionary entry” since the Sketch Engine system works “its way through the corpus to find all the recurring patterns for the word” and organizes them so they are “ready for the lexicographer to edit, elucidate, and publish”.

Word sketches have been used not only for lexicographical purposes but also in corpus-assisted discourse analysis. They are defined by Baker et al. (2013a) as “a relatively recent concept in corpus linguistics that identifies and groups together the salient lexical patterns of particular words within different grammatical structures” (Baker et al. 2013a: 35). For example, word sketches were used by Baker et al. (2013a) to explore how the words *Muslim* and *Islam* are characterized in the British press. Also Balfour (2019) investigates the representation of people affected by schizophrenia in the British press through word sketches and he finds out that “while schizophrenic people were not referred to explicitly as violent and dangerous unusually frequently, several discursive strategies cumulatively operating across multiple texts help imbue the lexeme with a semantic prosody of “dangerousness”” (Balfour 2019: 58-59).

For the 2030 Agenda Corpus, word sketches are extracted by typing the following search terms on the Sketch Engine platform: sustainable, sustainability, sustainable development for the 2030 Agenda (English); fenntartható, fenntarthatóság, fenntartható fejlődés for the 2030 Agenda (Hungarian); sostenibile, sostenibilità, sviluppo sostenibile for the 2030 Agenda (Italian). In addition, the part of speech of these lexical items is explicitly marked: the English lexical items sustainability, sustainable development, the Hungarian fenntarthatóság, fenntartható fejlődés, the Italian sostenibilità, sviluppo sostenibile are classified as nouns; the English lexeme

sustainable, the Hungarian *fenntartható*, the Italian *sostenibile* are classified as adjectives. This marking helps in the recognition of grammatical patterns among the lexemes' collocations. The minimum typicality score used for the default statistical measure (i.e. LogDice) is set at 6.0, whereas the minimum frequency of co-occurrence of node and collocate to be included in the word sketch is set at 5.

For the Sustainable development Corpus, word sketches are computed by typing the following lexemes on the Sketch Engine platform: *sustainable* for the SusCorp (English); *fenntartható* for the SusCorp (Hungarian); *sostenibile* for the SusCorp (Italian). All lexemes were marked with the "adjective" tag. As in the case of the 2030 Agenda Corpus, the minimum typicality score used for LogDice is set at 6.0; on the contrary, the minimum frequency of co-occurrence of node and collocate is set at 20.

Word sketches are labelled following a schema suggested by Brezina et al. (2015). The first piece of information to be signalled is the statistical measure chosen for the extraction of the collocates. The statistical measure is followed by the statistical threshold selected, which is inserted into brackets. For example, if a word sketch is calculated with LogDice as a statistical measure and with 6.0 as statistical threshold, the first part of the label will be LogDice(6.0). Then, the label indicates the frequency threshold as the minimum co-occurrence of node and collocate and the minimum occurrence of the collocate. For example, the label NC5-C5 means that node and collocate need to co-occur at least 5 times to be considered for the extraction of the collocational pattern and that the collocate should have at least a frequency of 5 within the corpus. Thus a collocational pattern signalled with LogDice(6.0), NC5-C5 is searched for with the LogDice statistical measure and by setting the statistical threshold to 6.0; the collocates are found making sure that the node and the collocate appear together under these conditions at least five times and that the collocate features at least five times in the corpus.

The collocational patterns depicted through word sketches are better explained through the reading of the collocates' concordance lines. In this case, concordance lines are extracted on the Sketch Engine platform by clicking on every collocate of the lexical items under inquiry.

### **3.4.2. Collocation networks**

Meaning by collocation is further researched with the extraction of collocation networks. Collocation networks are retrieved from the raw versions of the corpora.

Collocational patterns are built by searching for lemmas, while the collocates are not lemmatized. The choice of building collocation networks from lemmas but of considering all forms that collocates can come with adheres to Baker's (2006) idea that collocational patterns tend to change according to the different forms of a lemma. For instance, the list of collocates of a noun might vary in relation to its being used in singular or plural form or according to the case it is bound to. Thus, the choice of constructing collocation networks by searching for lemmas depends on the intention to outline a broad semantics for the lexical items under inquiry. At the same time, collocates are extracted in the form of word types in order to track the lexicogrammatical variety of the clauses including the nodes.

Collocation networks are gathered using #LancsBox's GraphColl tool. GraphColl is a tool "for investigating collocation networks" which enables researchers to visualise collocational patterns in the form of graphs (Brezina et al. 2015: 141). The graphs are constructed around a central dot (the node) by means of other small circles (the collocates). The node is connected to the collocates by lines. The length of the lines depends on the strength of the collocation: the shorter the line, the stronger the relationship between node and collocate. The position of the dots in the graph, namely whether they lie on the left or on the right of the node, is the result of their collocating more on the left or on the right of the search term. Their colour helps distinguish between function and content words (Brezina et al. 2015).

GraphColl displays first-order collocates (i.e. the collocates of a node) but it also allows to visualize second-order collocational patterns in a single graph. Second-order collocates are the collocates of a collocate. GraphColl admits graphs where the main collocation network is crowned by a secondary net, diffused from one of the node's collocates. Graphs can be carefully planned by users of the tool. In fact, researchers can set the following parameters: span, association measure, statistical and frequency thresholds. In addition, they can decide whether collocates should be found among types, lemmas, or parts of speech (Brezina et al. 2015). GraphColl is praised by Baker (2016: 139) because it "brings a new dimension to corpus-based analysis of collocation, plotting networks between multiple words simultaneously, rather than simply showing relationships between two words at a time".

In the current study, collocation networks are computed using the GraphColl tool to boost the results obtained through the word sketches. While the Sketch Engine's word sketches operate on annotated corpora, collocation networks are extracted from raw corpora in #LancsBox's GraphColl. #LancsBox, in fact, automatically operates basic linguistic analyses of the uploaded text (e.g. tokenization), but it adds further

morphological annotation like POS-tagging and lemmatisation only for some languages. English and Italian are included in the languages for which morphological annotation is available, whereas Hungarian is not. As a consequence, since it would be possible to automatically retrieve collocation networks for lemmas only for English and Italian, GraphColl is set so that all sections of the 2030 Agenda Corpus and of SusCorp are queried for word forms, but the query string is actually planned so that collocation networks are extracted for lemmas.

To do so, for the study of the 2030 Agenda Corpus, the 2030 Agenda (English) is queried with the strings */sustainable/, /sustainability/, /sustainable development/*, since the English lemmas *SUSTAINABLE, SUSTAINABILITY, SUSTAINABLE DEVELOPMENT* are invariable. For the same reason, in this study they are simply written about as *sustainable, sustainability, sustainable development* from now on. The 2030 Agenda (Hungarian) is queried with the strings */fenntartható\*/, /fenntarthatóság\*/, /fenntartható fejlődés\*/*, to search for the collocation networks of the Hungarian *FENNTARTHATÓ* ‘sustainable’, *FENNTARTHATÓSÁG* ‘sustainability’, *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’. The strings end with the wildcard *\** standing for “any character”, so that the software is instructed to consider both the nominative forms of the lemmas (i.e. *fenntartható, fenntarthatóság, fenntartható fejlődés*) and all forms where the lexical items are modified by bound morphemes. However, since the search for the string */fenntartható\*/* when extracting the collocation networks of the adjective *FENNTARTHATÓ* ‘sustainable’ might result also in lexemes irrelevant for the study of its meaning by collocation, including for instance occurrences of the lemma *FENNTARTHATÓSÁG* ‘sustainability’, all occurrences of these irrelevant lexemes are excluded manually. The 2030 Agenda (Italian) is queried with the strings */sostenibile|sostenibili/, /sostenibilità/, /sviluppo sostenibile/*, that allow to access the collocation networks of the Italian *SOSTENIBILE* ‘sustainable’, *SOSTENIBILITÀ* ‘sustainability’, *SVILUPPO SOSTENIBILE* ‘sustainable development’. The collocations of *SOSTENIBILITÀ* ‘sustainability’ and *SVILUPPO SOSTENIBILE* ‘sustainable development’ are calculated with a simple query because the two lexical items are usually invariable in Italian (for more on this point see § 4.4.3); the collocations of the *SOSTENIBILE* ‘sustainable’ are found by querying #LancsBox with the regular expression */sostenibile|sostenibili/* to find all the collocates of the singular and plural forms of the lemma. Collocation networks are calculated with the Z statistical measure (Barnbrook 1996). This statistical measure is preferred over the traditional MI and Log Likelihood for two main reasons: first, because it tries to balance the inclusion of rare linguistic events with more common ones (Barnbrook 1996); second, because it seems

to be in tune with studies on the psychological validity of collocation (Hughes 2018). In terms of psychological validity, Hughes (2018) suggests that association measures like z-score, MI3 and Dice coefficient outperform measures frequently praised in the literature on collocation extraction such as Mutual Information and Log Likelihood. In fact, the cognitive load observed in case of violation of a collocational pattern seems to be higher for association measures like z-score, MI3 and Dice coefficient and this shows that these measures are psychologically more robust (Hughes 2018; see also Hughes and Hardie 2020). The statistical threshold is set at 10.0 and the frequency threshold at 5. The frequency threshold is fixed on the 5 value because of the size of the corpus under inquiry: while a lower threshold might signal casual co-occurrences of words, a higher threshold could ignore interesting patterns. The initial window span employed for the analysis is one word to the left and one word to the right of the node, and it is then gradually widened to reach three words to the left and three words to the right of the search term, and finally five words to the left and five words to the right of the node. As Sinclair (2004) hints at, such collocation windows should encompass both morphosyntactic and discursive features of the corpora under inquiry. A larger window span is adopted in the comparative part of the analysis. It is not employed extensively throughout the search because it captures co-occurrences of words that belong to different sentences and it is thus retained not useful for the purposes of the current research.

For the analysis of the Sustainable development Corpus, the SusCorp (English) is queried with the string `/sustainable/` to retrieve the collocation networks of the English *sustainable*, the SusCorp (Hungarian) is queried with the string `/fenntartható*/` to collect the collocation networks of the Hungarian *FENNTARTHATÓ* 'sustainable', the SusCorp (Italian) is queried with the string `/sostenibile|sostenibili/` to extract the collocation networks of the Italian *SOSTENIBILE* 'sustainable'. Collocation networks are computed with the Z association measure, a statistical threshold of 10.0, a frequency threshold of 20, and spans of  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -words.

Both for the 2030 Agenda Corpus and for the Sustainable development Corpus, collocation networks are labelled in roughly the same way as the word sketches. The first element to be mentioned is the statistical measure used for the computation of the collocation networks, namely Z. The statistical measure is followed by the statistical threshold manually selected for the search. For example, the label `Z(10.0)` would stand for the use of the Z measure with a statistical threshold of 10.0. Then, the label includes the span that borders the collocational pattern both on the left and on the right of the node. For instance, if the collocate is looked for within three words to the left and three

words to the right of a search term, then the label will be 3L-3R with 3L standing for “three words to the left” and 3R for “three words to the right”. Finally, the label features the minimum frequency of co-occurrence for the node and the collocate and the minimum frequency of occurrence of the collocate. NC20-C20 means that the node and the collocate appear together 20 times and that the collocate is at least 20 times frequent in the corpus. The hypothetical label Z(10.0), 1L-1R, NC5-C5 would thus describe a word sketch built with the Z statistical measure set at a minimum score of 10.0 and used to extract collocates that appear at least 20 times in the corpus and at least 20 times together with the node within a collocation window of five words to the left and five words to the right of the node.

Concordances are used to detail the linguistic patterns traced by collocation networks and to enhance the collection of collocational patterns with their organisation into colligational patterns, semantic preference and evaluative prosody. The sorting of concordances to the left or to the right of a word, in fact, enables to outline the colligational tendencies of a lexeme, as it stresses the most frequent grammatical patterns that the node is included in. These colligational patterns are manually gathered by perusing the concordance lines and by assigning collocating words their part of speech; the frequency of these colligations is exploited to draw conclusions on the most significant grammatical patterns involving the lexical items under inquiry. The reading of concordances together with a glance at collocation networks allows also to group collocates according to the semantic area they belong to, obtaining thus a lexical item’s semantic preference. The results of the analysis of colligation and semantic preference offer the basis for outlining the evaluative prosody of the lexical item under inquiry. Through evaluative prosody, the semantics of the lexical item is judged positively or negatively in relation to the tenets of the ecological framework adopted for the analysis, thus introducing the ecological interpretation of meaning by collocation.

### **3.5. The cross-linguistic interpretation of the results**

In the present research, the study of cultural keywords and of meaning by collocation is carried out from a cross-linguistic perspective. This implies that the observations gathered throughout the analysis of cultural keywords and collocational patterns are interpreted under the light of the linguistic features of the three languages under inquiry, namely English, Hungarian and Italian.

### 3.5.1. English

English is a West Germanic language, and it belongs to the Indo-European family (König 1994).

As far as the morphological typology of English is concerned, according to Comrie (1981: 39), English is an isolating language: it is a language “which has no morphology, i.e. at least ideally, a language where there is a one-to-one correspondence between words and morphemes”. This means that inflection is maximally reduced, and lexical and grammatical functions are encoded by isolated words combined with each other according to specific syntactic constraints (Comrie 1981).

In English clauses, constituents usually have a rather fixed position. The basic word order of English simple clauses is SVO, with the SV preferred order for subject and verb and the VO preferred order for verb and object, eventually VOX if the valency of the verb requires also an oblique. In case of negation the word order is SNegVO (see, for instance, Quirk et al. 1985). English word order is stable, and it is rarely affected by pragmatic features like focus. On focus in English, Comrie (1981) affirms that “in general, there is no grammaticalized indication of focus, although focus is usually shown intonationally in the spoken language by being assigned sentence stress” (Comrie 1981: 57).

English regularly has a nominative-accusative alignment. This tendency notwithstanding, some verbs behave in an ergative way: they allow for both a transitive and an intransitive use and the kind of object chosen for the transitive perfectly equals the kind of subject selected for the intransitive (König 1994). However, the use of S is compulsory in English, being it a non-pro-drop language (Quirk et al. 1985).

The English word order influences also other relative orders within clauses: English prefers prepositions over postpositions (Prep), adjectives tend to precede nouns (AdjN), nouns tend to precede relative clauses (NRel), while there is no preference for the relative position of noun and genitive, as a genitive can both precede and follow a noun (Comrie 1981).

### 3.5.2. Hungarian

Hungarian is a Uralic language; it belongs to the Ugric branch of the Finno-Ugric languages, which are Western members of the Uralic stock (Abondolo 1998, Manzelli 1993).

From the point of view of morphological typology, Hungarian is an agglutinative language. According to Comrie (1981: 40), in agglutinative languages, “a word may consist of more than one morpheme, but the boundaries between morphemes in the word are always clear-cut; moreover, a given morpheme has at least a reasonably invariant shape, so that the identification of morphemes in terms of their phonetic shape is also straightforward”. Agglutinative languages like Hungarian tend to possess a wide range of lexical and grammatical morphemes that frequently combine within a single word (Comrie 1981).

As a consequence of this agglutinative tendency, in Hungarian, grammatical features are frequently signalled by bound morphemes. Bound morphemes tend to follow word stems: Hungarian has a strongly suffixing tendency. In addition, grammatical features can also be expressed through postpositions or with a combination of bound morphemes and postpositions (Kenesei et al. 1998).

The position of words in clauses is rather free in Hungarian. Hungarian, in fact, does not experience a dominant order in the distribution of syntactic functions in unmarked sentences: both the SOV and the SVO word orders are quite common. However, the preferred order for verb and subject is SV and the preferred order for verb and object is VO. (Kenesei et al. 1998). The relatively free position of syntactic functions depends on pragmatic reasons. In unmarked sentences, the topic of the sentence precedes the comment independently of their syntactic identity. In sentences marked by the presence of a focused word, focused words need to precede the finite verb (Kiefer 1967).

From a typological perspective, Hungarian has a nominative-accusative alignment. Being Hungarian a pro-drop language, when personal pronouns are used as subjects of a sentence, they can be omitted, as number and person of the subject are already encoded in the verb ending (Abondolo 1998).

In relation to Hungarian word order, other tendencies are experienced within sentences: the language prefers the use of postpositions instead of prepositions (Post), genitives and adjectives tend to precede nouns (GenN and AdjN), relatives can both precede and follow the noun they modify (RelN or NRel; Kenesei et al. 1998).

### **3.5.3. Italian**

Italian is a Romance language of the Indo-European family, belonging to the Italic branch (Vincent 1988).

As far as morphological typology is concerned, Italian is a fusional language, namely it has “no such clear-cut boundary between morphemes, the characteristic of a fusional language being that the expression of different categories within the same word is fused together to give a single, unsegmentable morph” (Comrie 1981: 41). In fusional languages like Italian, the morphemes that constitute words can express more than one lexical or grammatical function at a time (Comrie 1981).

Italian basic word order is SVO, although S and V can switch their position for pragmatic reasons. Since Italian is a pro-drop language, the subject can be omitted without impairing the meaning of the sentence (Vincent 1988).

The SVO word order triggers the following syntactic tendencies within sentences: Italian prefers the use of prepositions instead of postpositions (Prep), genitives and adjectives tend to follow nouns (N<sub>Gen</sub> and N<sub>Adj</sub>) even though the order of adjective and noun can be inverted for semantic or pragmatic reasons (AdjN), relative sentences tend to follow the noun they modify (N<sub>Rel</sub>), but they can precede it to serve semantic or pragmatic necessities (RelN; Dardano and Trifone 1997)

## **3.6. The ecological interpretation of the results**

At the end of the study, the results of the identification and analysis of cultural keywords and of meanings by collocation are interpreted under the light of the ecological framework adopted for the research. Cultural keywords are explained in terms of salience, namely the foregrounding of a linguistic pattern and of the concept it embodies; meanings by collocation are described in terms of ideology and evaluation by highlighting the ideological standpoint they recreate and the positive or negative connotation assigned to lexical items.

The ecological interpretation of the results in the current study relies on an ecological framework drawn from the reflections of sustainable development and of social ecology.

### 3.6.1. Sustainable development

The importance of sustainable development for the well-being of our planet and of the living creatures and non-living elements that our planet hosts has been debated for decades in our Western civilization. The first notes on the matter date back to the 1970s'. In 1972, the Stockholm Conference on the Human Environment forewarned the United Nations of the relationship between sustainable development and climate: development sustainability was believed to be an essential requirement for climate change to pace down (United Nations 1973). Back then, sustainable development was inherently tied to ecological matters. This bond continued up to the early 1980s'. In 1980, for instance, the document *World Conservation Strategy* of the International Union for the Conservation of Nature and Natural Resources described sustainable development as a form of development particularly keen on ecological matters (IUNC 1980). What is more, the document circulated the lexeme *sustainable development* for the very first time. In IUNC's *World Conservation Strategy*, the meaning of the unprecedented lexeme *sustainable development* conserved a bias towards embodying sustainability only in relation to ecology. The semantics of the lexical item broadened to encompass social and economic issues in 1987 with the report *Our Common Future*, released by the World Commission on Environment and Development (WCED) and known as the Brundtland Report (WCED 1987). In the Brundtland Report, sustainable development was written about as an economic, social and environmental matter that ought to be searched for if governments wished to promote a respectful and feasible kind of development for their countries (for a thorough account of the evolution of the notion of sustainable development see Baker 2006 and Meadowcroft 2005).

From the 1980s' on, the notion of sustainable development put forward by the Brundtland Report was accepted and adopted by most debates and reflections on sustainability and the growing economic, social and environmental concerns that humanity had to face stimulated much discussion and committing to sustainability actions. The United Nations, for instance, have engaged with the issue by devoting a series of meetings to sustainable development. The latest crucial forum of the United Nations on sustainability was held in New York between 25<sup>th</sup> and 27<sup>th</sup> September 2015 and it culminated in the publication of the document *Transforming our World: The 2030 Agenda for Sustainable Development* (United Nations 2015a).

In the 2030 Agenda, the UN discussed sustainable development in relation to five key matters, namely *people, planet, prosperity, peace, and partnership*. Drawing from these, they insisted that the route towards sustainable development should be

measured in terms of seventeen Sustainable Development Goals (SDGs), which are portrayed in Figure 1.



**Figure 1.** The seventeen Sustainable Development Goals of the United Nations' 2030 Agenda for Sustainable Development.

The Seventeen Sustainable Goals focus on social matters (Goals 3, 4, 5, 6, 7, 10, 11, 16, 17), but they also linger on economic issues (Goals 1, 2, 8, 9, 12) and touch on environmental concerns (Goals 13, 14, 15). The description provided by the United Nations (2015a: 14) can be read in Table 7.

<b>Goals</b>	<b>Description</b>
Goal 1	End poverty in all its forms everywhere
Goal 2	End hunger, achieve food security and improved nutrition and promote sustainable agriculture
Goal 3	Ensure healthy lives and promote well-being for all at all ages
Goal 4	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
Goal 5	Achieve gender equality and empower all women and girls
Goal 6	Ensure availability and sustainable management of water and sanitation for all
Goal 7	Ensure access to affordable, reliable, sustainable and modern energy for all
Goal 8	Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all
Goal 9	Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation
Goal 10	Reduce inequality within and among countries
Goal 11	Make cities and human settlements inclusive, safe, resilient and sustainable

Goal 12	Ensure sustainable consumption and production patterns
Goal 13	Take urgent action to combat climate change and its impacts
Goal 14	Conserve and sustainably use the oceans, seas and marine resources for sustainable development
Goal 15	Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
Goal 16	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
Goal 17	Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development.

**Table 7.** The seventeen Sustainable Development Goals of the United Nations' *2030 Agenda for Sustainable Development*.

The United Nations envisage these seventeen goals as a thorough description of the difficulties and needs born by world development, but they also wish they might be adopted as “a strategy of global governance” (Kanie et al. 2017: 5).

### 3.6.2. Social ecology

The strategies of global governance hinted at by Kanie et al. (2017) in relation to sustainable development can be uncovered also in the theories of social ecology, which keeps together the study of the characteristics and development of human communities with the observation of the environment that human communities influence and are influenced by.

Social ecology developed since the 1920s-1930s as a discipline that applied the tenets of 19<sup>th</sup> century bioecological theories such as Darwin's theory to the investigation of human communities. The first research group of social ecology, namely the Chicago School of Human Ecology, studied human communities with the conceptual and methodological tools employed in bioecology and they maintained that material, environmental conditions unidirectionally impacted on social conditions and phenomena. This unidirectional impact was changed in the 1970s into a more organic influence involving bidirectionally material, environmental conditions and social conditions and phenomena: according to this evolved version of social ecology, material, environmental conditions influence social conditions and phenomena, but at the same time they are influenced as well. Within this more encompassing theory of social ecology, it is nowadays believed that

social ecology generally refers to the study of communities from a broad, interdisciplinary perspective that encompasses bioecological and macro-economic concerns, but gives greater attention to the social, psychological, institutional, and cultural contexts of people-environment relationships that did earlier human ecology research (Stokols et al. 2013: 3).

According to Stokols et al. (2013: 3), the core principles of social ecology reside in its emphasising the multidimensional organisation of human environments, in its incorporating “multiple levels of analysis and diverse methodologies for assessing the resilience and healthfulness of settings and the well-being of individuals and groups”, and in its encompassing theoretical concepts and assumptions from economic and cultural theories.

Economic and cultural theories contribute to paying attention to various forms of capital for the description of the multidimensional relations that shape human communities. Among these forms of capital, social ecology focuses on social capital as it is defined by Bourdieu (1986) and on human capital, as it is formulated by Coleman (1988). Bourdieu (1986: 21) describes social capital as “the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition”. Coleman (1988: S100) states that human capital “is created by changes in persons that bring about skills and capabilities that make them able to act in new ways”. Social and human capital belong to a broader set of human resources that cooperate with material resources for the resilience and well-being of human communities (Stokols et al. 2013).

Within social ecological theories, the economic and social dynamics crossing human communities are intertwined with environmental dynamics. This is affirmed most famously by Murray Bookchin in volumes published since the 1970s (among these see, for instance, Bookchin 1989, 1995, 2005). Bookchin devotes a keen attention to the relations occurring between human communities and environmental problems. In relation to this, Best (1998: 337) writes that according to Bookchin

all current environmental problems are ultimately social problems, rooted in an irrational and antiecological society whose crises cannot be solved through piecemeal, single-issue reform measures. Dislocations in the human-nature relation stem from dislocations within the human world itself; environmental problems emerge from a long history of hierarchical social relations that culminate in a class-ridden, profit-driven, accumulation-oriented capitalist society.

The concern of social ecology for environmental issues as well as for economic and social matters ties the discipline to the theories of sustainable development that have been previously described.

### **3.6.3. The ecological framework of this study**

The ecological framework adopted in this study gathers elements from the theories of sustainable development and from social ecology theory and it can be summarised as follows:

*Inclusion and care:* In the global world we live in, human communities should be inclusive so that both indigenous people and newcomers might feel at home. Humans should respect each other and care for each other's wellbeing with empathy and firmness.

*Respect:* Humans should respect and protect the other beings residing in nature, namely animals, plants and the elements of the natural world, by regarding them as beings and elements with their own peculiarities and needs. In addition, environment in general should be preserved from extreme human impacts like pollution.

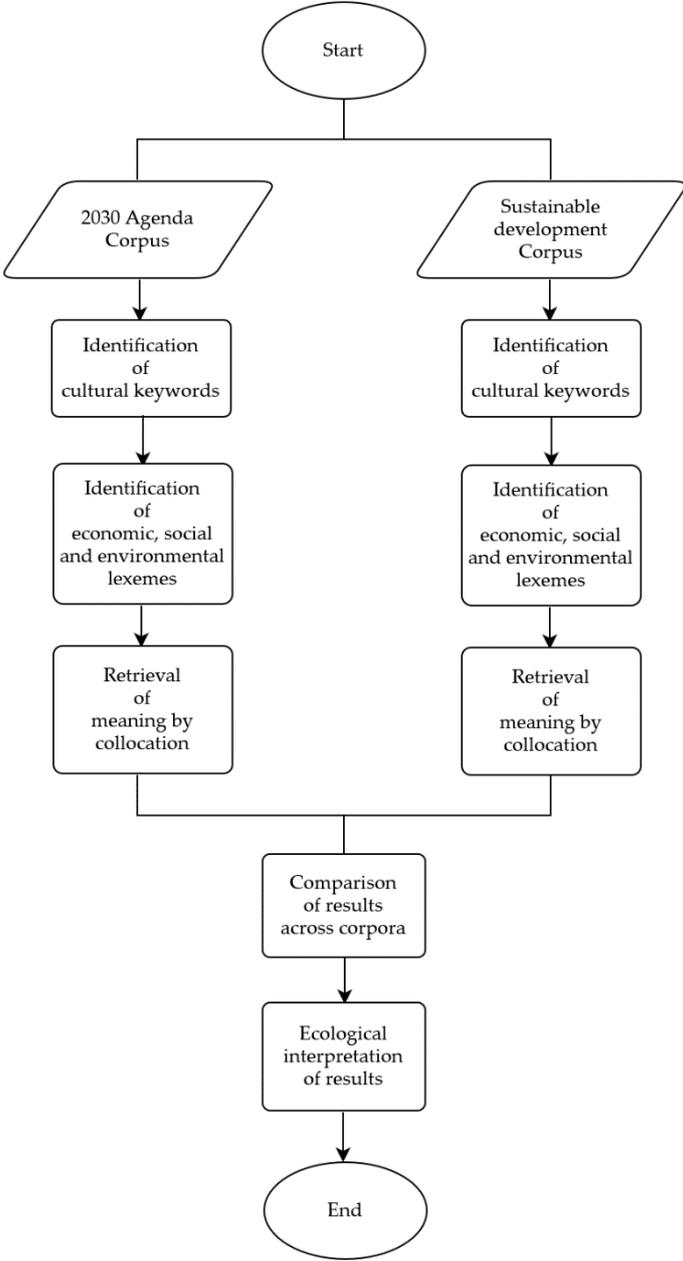
*No exploitation:* Nature provides human communities with valuable elements to live by, but its capacity to reproduce them is not unlimited. Consequently, animals, plants and natural elements should not be spoiled or treated as resources to exploit.

*Sustainable growth:* The economic and social growth of human communities should follow the real needs of communities, and it should be matched with a general decrease in consumption.

This ecological framework is employed to interpret the results of the identification and study of cultural keywords and meanings by collocation. The salience of cultural keywords and the ideology and evaluation that evaluative prosody is imbued with are assessed under the light of the ecological framework. As far as cultural keywords are concerned, the ecological framework allows to establish whether their being salient is beneficial or detrimental for the semiotic relationship that they supply to the human-nature relation. As far as meanings by collocation are concerned, the ecological framework aids in understanding whether the ideology and evaluation that load the evaluative prosody of the lexemes under inquiry is beneficial or detrimental for the saving of a balance between humans and nature.

### 3.7. The structure of the study

The ecological interpretation of the results of the analysis of cultural keywords and meanings by collocation is the final step of the study, whose complete structure can be seen in the flow chart represented in Figure 2.



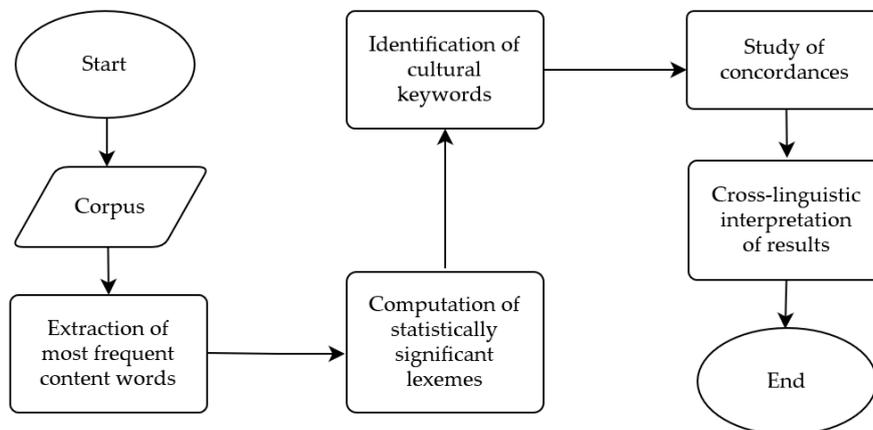
**Figure 2.** The structure of the present research in a flowchart.

As it can be seen in Figure 2, the study is carried out first on the 2030 Agenda Corpus and then on the Sustainable development Corpus. For both corpora, the analysis

begins with the identification of cultural keywords; the identification of cultural keywords is followed by the spotting of economic, social and environmental lexemes and then by the retrieval of the meaning by collocation of pinpointed lexical items. The outcome of these analytical steps is enriched with a comparison of the results found in the two corpora and with an ecological interpretation of these results.

For the Sustainable development Corpus, economic, social and environmental lexemes are identified by gathering words that refer to economic, social and environmental issues from the lists of the most frequent content words and of the statistically significant lexemes of the corpora. For the 2030 Agenda Corpus, the lexemes found among the most frequent and statistically significant lexemes are added with all other economic, social and environmental words found in the whole frequency list.

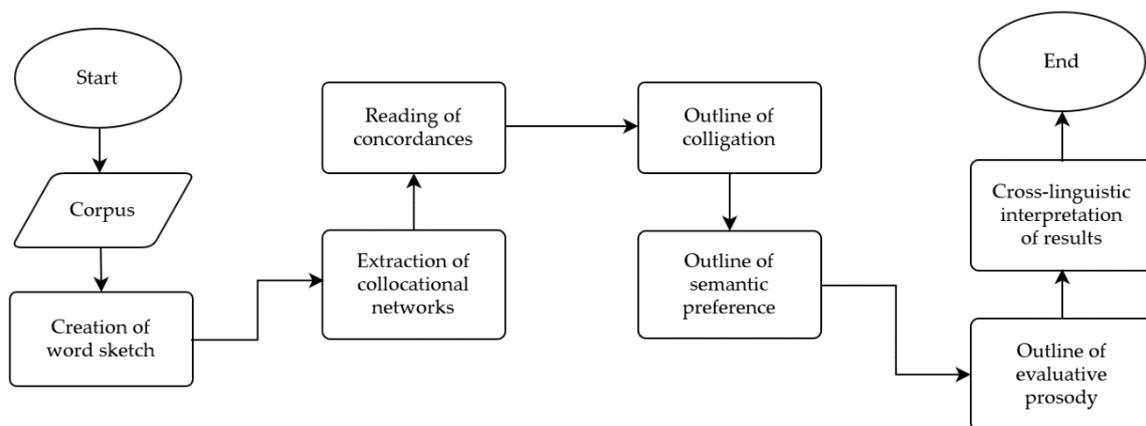
For both corpora, the identification of cultural keywords is organised as in Figure 3.



**Figure 3.** The identification of cultural keywords in a flowchart.

As Figure 3 details, the study of cultural keywords involves both corpora with the extraction of the most frequent content words of the collection, followed by the computation of statistically significant lexemes, and by the identification of cultural keywords among the most frequent content words and statistically significant lexemes of the corpus. Then, cultural keywords are explored with the study of their concordances and with a cross-linguistic interpretation of the results.

The retrieval of meaning by collocation is achieved both for the 2030 Agenda Corpus and for the SusCorp with the process represented in Figure 4.



**Figure 4.** The retrieval of meanings by collocation in a flowchart.

Figure 4 shows that the retrieval of meanings by collocation for both corpora generally follows the following process: word sketches are created as a basis for the analysis and their results are enhanced through the extraction of collocation networks; collocation networks are explored by reading the concordance lines of the collocates and by drawing from them colligational patterns, semantic preference and evaluative prosody; results are then interpreted cross-linguistically.

## 4. Analysis of the 2030 Agenda Corpus

In the fourth chapter I present the results of the corpus-assisted study of the 2030 Agenda Corpus in search for the discursive construction of sustainable development. First, I collect and explore the cultural keywords represented in the English, Hungarian and Italian versions of the 2030 Agenda. I isolate them among the most frequent and statistically significant lexemes of the three documents. I also pinpoint the lexemes that belong to the economic, social and environmental dimensions of sustainable development. Then, focusing on the social and environmental dimensions of sustainable development, I single out the human, animal, plant and natural actors that feature in the resolution. Finally, I sketch the meaning by collocation that the English lexical items *SUSTAINABLE*, *SUSTAINABILITY*, *SUSTAINABLE DEVELOPMENT* and their corresponding Hungarian and Italian translational equivalents display in the 2030 Agenda Corpus.

### 4.1. Cultural keywords of the 2030 Agenda Corpus

The cultural keywords identified in the 2030 Agenda Corpus mirror the aboutness of the United Nations' *2030 Agenda for Sustainable Development* as they point at the most meaningful lexemes mentioned by the UN's resolution. By embodying historically, politically, socially and culturally salient matters, these paramount lexemes reflect and help shape the historical, political, social and cultural background that witnessed the Agenda's release. The identification of cultural keywords is carried out under the light of the main tenets of the literature on sustainable development (Baker 2006, Kanie et al. 2017, Meadowcroft 2005).

#### 4.1.1. The most frequent lexemes

Cultural keywords are collected first through the analysis of the most frequent lexemes of the English, Hungarian and Italian versions of the document. As Hoover (2009) states, in fact, high-ranking content words can reflect the aboutness of a corpus and signal the most peculiar issues addressed by the collection.

The most frequent lexemes of the Agendas are extracted with AntConc's Word List function. Among the most frequent lexemes of the documents, the twenty most frequent content words are focused on in search for cultural keywords and they are

analysed further through the perusal of their concordance lines. The concordance lines are scanned to highlight recurrent patterns of co-occurrence, which are identified by means of frequency. Raw frequency is reported in brackets with an *f*; for instance, if the word type *journey* appears four times, then it is reported as “*journey (f=4)*”.

Since the Hungarian and Italian Agendas are probably translations of the English document, the list of the most frequent content words of the English Agenda is adopted as baseline and as a *tertium comparationis* for the analysis of the Hungarian and Italian wordlists. Similarities and differences among the three lists of lexemes are highlighted as they could be justified by linguistic reasons, but they might also be signposts of cultural and political tendencies reflected by the translations. The translations of the Hungarian and Italian lexemes correspond to the equivalent word forms used in the English version of the Agenda. Moreover, Hungarian and Italian examples are glossed with the Leipzig glossing rules when grammatical issues are debated.

#### 4.1.1.1. English

The identification of the cultural keywords emerging from the most frequent lexemes of the 2030 Agenda Corpus starts with the analysis of the English version of the resolution.

The six most frequent words of the English Agenda are function words (*and*, *f*=1,237; *the*, *f*=837; *of*, *f*=553; *to*, *f*=428; *in*, *f*=305; *for*, *f*=220). The prominence of function words in frequency lists does not come as a surprise since the underlying structure of Standard Average European is shaped around this category of words (Heine and Kuteva 2006). In the English version of the 2030 Agenda these function words are followed in rank by four content words (i.e. *development*, *countries*, *sustainable*, *all*), whose high frequency and whose appearing among the ten most frequent word types of the document set them as cornerstones of the aboutness of the text.

The twenty most frequent content words of the English section of the 2030 Agenda Corpus can be seen in Table 8. In Table 8, content words are listed with their raw frequency (*f*).

Rank	Content word	<i>f</i>			
			4	<i>developing</i>	95
1	<i>development</i>	202	5	<i>global</i>	78
2	<i>countries</i>	184	6	<i>2030</i>	73
3	<i>sustainable</i>	177	7	<i>agenda</i>	71
			8	<i>national</i>	71

9	<i>goals</i>	64	15	<i>support</i>	53
10	<i>international</i>	61	16	<i>access</i>	52
11	<i>developed</i>	60	17	<i>ensure</i>	50
12	<i>economic</i>	60	18	<i>implementation</i>	49
13	<i>nations</i>	58	19	<i>levels</i>	45
14	<i>united</i>	56	20	<i>promote</i>	45

**Table 8.** The twenty most frequent content words of the 2030 Agenda Corpus (English).

Among the most frequent content words of the English version of the 2030 Agenda, *development*, *sustainable*, *global* and *goals* deserve the right to be defined cultural keywords because of their distinguishing themselves within the “vocabulary we share with others” when conversing about politically salient issues in our Western society (Williams 1983: 13). The four lexemes symbolize some of the paramount topics of the politics and culture of the mid-2010s, when the 2030 Agenda was released by the United Nations: the noun *development* recalls the constant, eager ambition for progress that characterises Western societies; the adjective *sustainable* evokes the paramount quality that progress should have to be respectful of all components of the environment, from the human to the natural; the adjective *global* stresses that nowadays most decisions and actions tend to be conceived with an international scope; the noun *goals* alludes at the Sustainable Development Goals and it spotlights them outright.

The first ranking cultural keyword is *development*, which holds a broad collocational behaviour. *development* is modified by the adjective *sustainable* in about 60% of its occurrences (108 out of 202; for more on the semantics of *sustainable development* see § 4.4.1). The other noun and adjective phrases that modify *development* can be seen in Table 9.

<b>Noun and adjective phrases</b>	<b><i>f</i></b>		
<i>sustainable</i>	108	<i>post-2015</i>	2
<i>Millennium</i>	8	<i>social</i>	2
<i>official</i>	5	<i>urban</i>	2
<i>Doha</i>	4	<i>early childhood</i>	1
<i>United Nations</i>	4	<i>economic</i>	1
<i>economic and social</i>	4	<i>enhanced</i>	1
<i>national</i>	3	<i>global</i>	1
<i>technology</i>	3	<i>infrastructure</i>	1
<i>Africa's</i>	2	<i>national and regional</i>	1
<i>industrial</i>	2	<i>social and economic</i>	1
<i>international</i>	2	<i>pro-poor and gender sensitive</i>	1
		<i>rural and urban</i>	1

**Table 9.** Noun and adjective phrases modifying *development* in the 2030 Agenda Corpus (English).

Thanks to these noun and adjective phrases, *development* is described as a condition that touches on various dimensions, among which the economic (*economic, industrial*) and the social (*social, early childhood, and pro-poor and gender sensitive*), which can also be treated together (*economic and social, social and economic*). It involves various fields of human activity (*technology, urban, infrastructure, and sustainable agriculture, pastoralist and fisheries*) and it depends on political and geographical conditions (*national, Africa's, international, global, and national and regional*). Development is lexicogrammatically associated with political events (as in *Millennium, official, Doha, United Nations, post-2015*). Among these, the most interesting for the purposes of this study is *Millennium*.

*Millennium* belongs to the noun phrase *Millennium Development Goals* ( $f=8$ ), which is very important for the existence of the very Agenda. The Millennium Development Goals (or MDGs) are the eight goals set by the United Nations Millennium Declaration. The United Nations Millennium Declaration is a resolution that condensed years of debates on the most urgent social problems of the world and that were finally gathered during the United Nations' Millennium Summit in 2000. The Millennium Declaration mainly aimed at tackling the issue of extreme poverty and hunger in underdeveloped countries, encompassing also other social concerns like universal primary education, gender equality and the empowerment of women, child mortality, maternal health, diseases like HIV/AIDS and malaria. It devoted some space also to environmental sustainability and a global partnership for development (on the MDGs see McGillivray 2008). These topics were included in eight goals and twenty-one related targets which should have been met by 2015 according to the UN and whose results have been regarded as debatable (for instance, on MDGs and poverty see Cimadamore et al. 2016, on MDGs and human rights see Langford et al. 2013). In 2015, the Millennium Development Goals were substituted by the *2030 Agenda for Sustainable Development*.

*development* features as a modifier or the head of noun phrases in a variety of expressions. These noun phrases are displayed in Table 10.

<b>Noun phrases</b>	<b><i>f</i></b>		
<i>Goals</i>	41	<i>assistance</i>	5
<i>Agenda</i>	5	<i>system</i>	5
		<i>strategies</i>	4

Goal	3	planning	1
challenges	2	priorities	1
cooperation	2	processes	1
policies	2	progress	1
efforts	1	Report	1
Finance	1	Round	1
impacts	1	spending	1
Organization	1	strategies and policies	1
-oriented policies	1	workers	1
outcomes	1		

**Table 10.** Noun phrases modified by *development* in the 2030 Agenda Corpus (English).

Because of the noun phrases it modifies, *development* is characterised as a condition that needs to be reached through an active and well-designed engagement (*assistance, strategies, system, Policies, efforts, Organization, -oriented policies, planning, priorities, processes, spending, strategies and policies*), which needs to be constantly assessed (*impacts, outcomes, progress, Report*). Individuals or communities (*workers*) should strive to collaborate (*cooperation*) towards development, even when this requires overcoming difficulties (*challenges*). Development can also have to do explicitly with official resolutions (*Goals, Agenda, goal, Round*) and it seems to favour an economic dimension (*Finance*).

The lexeme *development* also belongs to noun phrases formed with the conjunction *and*. In these noun phrases, *development* can either precede or follow the conjunction. When it precedes the conjunction, it is followed by the lexical items *capabilities* (f=2), *capacities* (f=1), *human well-being* (f=1), *lifestyles* (f=1), *management* (f=1), *operationalization* (f=1), *other relevant ongoing processes* (f=1), *peace* (f=1), *plant and livestock gene banks* (f=1), *sustainable lifestyles* (f=1). When *development* follows the conjunction, it is preceded by the lexical items *Environment* (f=3), *Population* (f=3), *research* (f=2), *national policies* (f=1), *private research* (f=1), *Trade* (f=1). The lexical items coupled with *development* stress that change should be aimed at with a proactive attitude (*research, capabilities, capacities, management, national policies, operationalization, other relevant ongoing processes*) and that not only should it tackle economic and social issues (*Population, human well-being, lifestyles, peace, sustainable lifestyles, Trade*), but that it should also contemplate environmental matters (*Environment, plant and livestock gene banks*).

The second cultural keyword identified among the most frequent content words of the English Agenda is the adjective *sustainable*. *sustainable* emerges among the highest-ranking content words as a symbol of the Agenda. *sustainable* mainly

appears together with *development* in the expression *sustainable development* (108 occurrences out of 177, roughly 61% overall) but it is also associated with expressions like *economic growth* (f=6), *consumption and production* (f=6), *agriculture* (f=4), etc. Broadly speaking, the adjective modifies a noun phrase which refers to activities and processes that are carried out in the light of sustainability (for more on the semantics of the adjective *sustainable* in English see § 4.2.1).

The third cultural keyword spotted among the most frequent content words of the English Agenda is the adjective *global*. *global*, together with the other very frequent lexemes *national* and *international*, testifies that the enterprise of the 2030 Agenda is envisioned both at a national and at an international level. These words are either used independently (e.g. in *Global Partnership*, f=11; *international cooperation*, f=9; *national parliaments*, f=2) or they are used in combination through the conjunction *and*. In fact, they feature together in the following patterns: *national and global* (f=3), *national and international* (f=3). *global* and *international* are also found together in *global international economic and financial institutions* (f=6). Although it does not appear among the most frequent words of the document, also *regional* can be found among these patterns (e.g. in *regional and global*, f=6; *national, regional and global*, f=3; *national, regional and international*, f=2; *regional and national*, f=2). Also *domestic* (in *domestic and international*, f=2), *local* (in *national and local*, f=1) and *subnational* (in *national and subnational*, f=1) provide an overview of the contexts in which sustainable actions can be taken.

The fourth cultural keyword singled out in the English Agenda is the noun *goals*. The *goals* are most frequently the *Sustainable Development Goals* (f=33), but they can also be the *Millennium Development Goals* (f=8). These goals are sometimes mentioned only with the noun *Goals* and they are frequently associated with *targets* with the conjunction *and*.

The remaining most frequent content words of the 2030 Agenda contribute the aboutness of the document, although they cannot be considered cultural keywords. The noun *agenda* usually refers to the very UN's resolution. It is related to the concept of sustainable development in expressions like *this Agenda* (f=15), *the Agenda* (f=13), *the new Agenda* (f=12), and *2030 Agenda for Sustainable Development* (f=4). Moreover, the noun *agenda* is also incorporated in the *Addis Ababa Action Agenda* expression. The *Addis Ababa Action Agenda* is a document published in 2015 at the conclusion of the United Nations' Third International Conference on Financing for Development held in Addis Ababa between 13<sup>th</sup> and 16<sup>th</sup> July 2015. The Addis Ababa Action Agenda is not explicitly related to the topic of sustainability, but it supports substantially the implementation of the *2030 Agenda for Sustainable Development*.

*united* and *nations* go in pairs and their high frequency depends on the Agenda's mentioning the *United Nations* quite often. For instance, United Nations are recalled in relation to conferences; in fact, the pattern *United Nations (World) Conference on [NOUN PHRASE]* is repeated eight times with variations. The UN are also recognized as the authors of conventions and charters. *United Nations* co-occurs six times with *charter* in the pattern *Charter of the United Nations* and it collocates with *convention* in seven instances, among which *United Nations Framework Convention on Climate Change* ( $f=4$ ) and *United Nations Convention on the Law of the Sea* ( $f=2$ ). UN are also spoken about in terms of the *system* they have implemented (11 occurrences like *United Nations system* or *United Nations development system*) or of the *framework* they operate with (5 occurrences, among which 4 of *United Nations Framework Convention on Climate Change*).

The United Nations assume that sustainable development unfolds in three dimensions, namely an economic, a social and an environmental dimension. The economic dimension of sustainable development is the only one that can be traced among the most frequent content words of the 2030 Agenda Corpus. Among the most frequent content words of the document, in fact, the only word type that addresses one of the three issues related to sustainability is *economic*. The adjective most closely modifies the noun *growth* ( $f=13$ ), but it also collocates with *council* (in *Economic and Social Council*;  $f=12$ ) and with *development* ( $f=6$ ), as Table 11 shows. The economic dimension of sustainable development is most significantly associated to a kind of growth that needs to be governed by institutions like the United Nations Economic and Social Council. The United Nations Economic and Social Council is one of the organs of the United Nations. As it claims in its website, the United Nations Economic and Social Council operates to boost the three dimensions of sustainable development by means of “fostering debate and innovative thinking, forging consensus on ways forward, and coordinating efforts to achieve internationally agreed goals”; in addition, it is “responsible for the follow-up to major UN conferences and summits”.

<b>Noun phrases</b>	<b><i>f</i></b>	<b><i>Affairs</i></b>	<b></b>
<i>growth</i>	13	<i>barriers</i>	1
<i>Council</i>	12	<i>benefits</i>	1
<i>development</i>	6	<i>empowerment</i>	1
<i>governance</i>	3	<i>environment</i>	1
<i>resources</i>	3	<i>fields</i>	1
<i>dimensions</i>	2	<i>foundations</i>	1
<i>integration and interconnectivity</i>	2	<i>inclusion</i>	1
<i>activity</i>	1	<i>institutions</i>	1

<i>life</i>	1	<i>productivity</i>	1
<i>links</i>	1	<i>progress</i>	1
<i>losses</i>	1	<i>shocks and disasters</i>	1
<i>measures</i>	1	<i>status</i>	1
<i>objectives</i>	1		

**Table 11.** Noun phrases modified by *economic* in the 2030 Agenda Corpus (English).

The full list of the noun phrases modified by the adjective *economic* implies that economy is a field (*dimensions, fields*) that demands political engagement (*governance, Affairs, foundations, institutions, measures, status*) and the usage of resources (*activity, resources*) in order to boost profits (*benefits, empowerment, objectives, productivity, progress*) and to minimize damage (*barriers, losses, shocks and disasters*). This can be achieved through cooperation (*integration and interconnectivity, inclusion, links*) and it can reach also social and environmental dimensions (*environment, life*).

The adjective *economic* participates also in adjective phrases that include the adjectives *social* ( $f=29$ ), *inclusive* ( $f=10$ ), *sustainable* ( $f=10$ ), *sustained* ( $f=8$ ), and *environmental* ( $f=8$ ). The adjectives *social* and *environmental* tell the story of the two other main dimensions of sustainable development: one involving society and people, the other caring about the environment.

According to the most frequent lexemes of the English Agenda, the economic, social and environmental development that the 2030 Agenda promotes should be tackled with a supportive and reassuring attitude. The word types *support* and *ensure*, in fact, are very frequent in the document. *support* is used both as a verb and as a noun and it is mainly aimed at backing actions for *developing countries*. *ensure*, on the other hand, is employed as a promise to guarantee the implementation of measures that could help reach the Sustainable Development Goals and that could offer *access* to opportunities for growth and wellbeing. Also *access*, in fact, stars among the most frequent words of the document. It is always followed by a prepositional phrase introduced by *to* and completed with a wide range of noun phrases that reproduce the elements that should be accessed according to the Agenda. To this extent, the two most prominent noun collocates of *access* are *information* ( $f=5$ ) and *education* ( $f=5$ ). What should be bolstered according to the document, in fact, is the education of people having less opportunities to study compared to the ones dwelling in developed countries (on this see, among others, McCowan 2019).

Both developed and developing countries are prompted to act in favour of sustainable development. They are represented among the most frequent content

words of the English Agenda by the lexemes *developing*, *developed* and *countries*. *developed* modifies *countries* (or *country*) in roughly 93% of the cases. In most cases (44 occurrences out of 52), *developed countries* does not refer to countries that are characterized by a high level of development but to *least developed countries*. *least developed countries* are also called *developing countries* ( $f=69$ ) or *developing states* ( $f=23$ ). When country development is explicitly referred to in the Agenda, it seems that underdevelopment is more frequently dealt with compared with development. This is in tune with the purpose of the document, that wishes to contribute to the improvement of life conditions in developing countries. The noun *countries*, in fact, is more frequently modified by adjectives and nouns that refer to underdevelopment rather than to ones that refer to development. Apart from *developing* and *least developed*, other modifiers of *countries* are *middle-income* ( $f=7$ ), *most vulnerable* ( $f=3$ ), *post-conflict* ( $f=2$ ), and *highly indebted poor* ( $f=1$ ). Also *African* ( $f=10$ ) seems to identify least developed countries: it is always combined with phrases like *least developed countries*, *landlocked developing countries* or *small island developing States*. These occurrences make up for approximately 74% of the overall instances of *countries* (on the engagement for sustainable development in Africa see, among others, Froehlich 2019).

#### 4.1.1.2. Hungarian

The spotting of cultural keywords among the most frequent content words of the 2030 Agenda Corpus continues with the analysis of its Hungarian section so as to check whether similar or different content words could exemplify the most meaningful political, social and cultural concerns of the Hungarian society.

The three most frequent lexemes of the Hungarian Agenda are the function words *a* 'the' ( $f=1,433$ ), *és* 'and' ( $f=1,018$ ) and *az* 'the; that' ( $f=471$ ). In Hungarian the first content word of the frequency list is *fenntartható* 'sustainable', which is followed by two function words and then by two content words interspersed in the list of the ten most frequent word types of the Agenda. The second and the third content words of the frequency list are *fejldő* 'developing' and *országok* 'countries', as Table 12 shows.

Rank	Content word	f			
			6	2030 '2030'	71
1	<i>fenntartható</i> 'sustainable'	181	7	<i>nemzetközi</i> 'international'	63
2	<i>fejldő</i> 'developing'	95	8	<i>fejlett</i> 'developed'	57
3	<i>országok</i> 'countries'	94	9	<i>gazdasági</i> 'economic'	57
4	<i>globális</i> 'global'	76	10	<i>fejldés</i> 'development'	55
5	<i>nemzeti</i> 'national'	74	11	<i>biztosítása</i> 'ensure; provide'	49

12	<i>agenda</i> ‘agenda’	47	17	<i>jelentős</i> ‘substantially;	41
13	<i>való</i> ‘-’	47		significantly’	
14	<i>fejlesztési</i> ‘development’	44	18	<i>cél</i> ‘goal’	40
15	<i>történet</i> ‘-’	44	19	<i>célok</i> ‘goals’	40
16	<i>szintű</i> ‘level’	43	20	<i>különösen</i> ‘in particular’	39

**Table 12.** The twenty most frequent content words of the 2030 Agenda Corpus (Hungarian).

The cultural keywords identified for the English Agenda apply also to the Hungarian document in the form of their translational equivalents. As in the English case, also in Hungarian the lexemes *fenntartható* ‘sustainable’, *globális* ‘goal’, *fejlesztés/fejlesztési* ‘development’ and *célok* ‘goals’ reflect and shape the cultural and political concerns of the Hungarian society at the beginning of the 21<sup>st</sup> Century.

Moreover, the most frequent content words of the Hungarian version of the 2030 Agenda shape a similar aboutness of the text compared to the English version of the document. Both versions, in fact, share a focus on sustainability with the lexemes *sustainable-fenntartható* and *development-fejlesztés/fejlesztési*. They frame the notion of sustainable development into the United Nation’s resolution with the lexemes *2030, agenda-agenda* and *goals-célok*. Both the English and the Hungarian texts point out that the achievements of sustainable practices depend on the level of development of the countries involved in the process (*countries-országok, developed-fejlett, levels-szintű, developing-fejlődő*) and that this process ought to be tackled both at in a national and in an international setting (*global-globális, international-nemzetközi, national-nemzeti*). The achievement of sustainable development is regarded by the English and the Hungarian Agendas as requiring *support-biztosítása* and as devoting paramount importance to the *economic-gazdasági* dimension.

The remaining content words of the Hungarian frequency list of Table 12 list the lexemes *cél* ‘goal’, *való* ‘-’ and *történet* ‘-’, *jelentős* ‘substantially; significantly’ and *különösen* ‘in particular’. *cél* ‘goal’ is used as an alternative to *célok* ‘goals’ when the document lists the SDGs. The English way of listing the goals involves the use of a numeral (e.g. *Goal 2*); also the Hungarian goals are listed by employing the numeral, which is signalled by a full stop following the number (e.g. *6. cél* ‘Goal 6’).

*történet* and *való* are active, present participles that modify nouns by functioning as the head of adjective phrases. *történet* is derived from the verb stem *történ* ‘happen’ with the addition of the suffix *-ő*. *történet* literally means ‘happening’; it is generally used to construct adjective phrases that modify nominalized verbs, whose semantics needs to be completed with arguments, as in

A víz-hez és szanitáció-hoz történő hozzáférés-Ø  
 DET water-ALL and sanitation-ALL happen-PTCP.PRS availability-NOM  
 'availability [...] to water and sanitation'

*való* is one of the two present participle forms of the verb 'to be', which is characterised by two stems (i.e. *le-* and *val-*). *való* consists of the verb stem *val-* 'to be' and of the suffix *-ó*. *-ó* and *-ő* are the two forms that the suffix of the present participle in Hungarian has in relation to the vocalism of the word that it is added to, following the principle of vowel harmony. The use of *való* is similar to that of *történő*.

The adjective *jelentős* 'substantially; significantly' mainly modifies the nouns *CSÖKKENTÉSE* 'reduce; reduction' (in 11 instances) and *NÖVELÉSE* 'enhance; increase' (in 8 instances), which consist of a deverbal nominal stem (*csökkentés* and *növelés* respectively) and of a possessive suffix *-e*. The significant reduction or increase can be of a certain degree (*jelentős mértékű* 'substantially', literally 'of a significant degree', *f=5*) and they mainly involve questions of human, environmental, and economic protection. The *CSÖKKENTÉSE* 'reduce; reduction' that is wished for is mainly one of human losses, of corruption and economic degradation, or of environmental degradation. The *NÖVELÉSE* 'enhance; increase' that is wished for is mainly one of economic growth and improvement in human conditions, as the noun phrases that modify *NÖVELÉSE* hint at. Significant reduction and rise are temporally confined in a time that precedes 2030 (*2030-ig* 'by 2030').

The adverb *különösen* 'in particular' is a stylistic peculiarity of the Hungarian version of the agenda. *különösen* 'in particular' is usually followed by noun phrases referring to countries, clearly distinguished according to their being developed and developing and according to their geographical location (as in *a fejlődő országok* 'developing countries', *f=3*, *a legkevésbé fejlett országok* 'least developed countries', *f=3*, and *az afrikai országok* 'African countries', *f=2*).

The bulk of the most frequent content words of the Hungarian Agenda depicts for the document an aboutness that faithfully resembles the aboutness of the English Agenda. The peculiarity of the Hungarian document in terms of content word frequency is of a linguistic and of a stylistic kind: lexemes like *cél* 'goal', *való* '-l' and *történő* '-l' emerge among the most frequent content words of the Agenda because of the linguistic characteristics of Hungarian; lexemes like *jelentős* 'substantially; significantly' and *különösen* 'in particular' shed light on diverse stylistic choices that can be adopted when translating the Agenda into Hungarian.

### 4.1.1.3. Italian

A comparative study between the most frequent content words of the Italian version of the 2030 Agenda and the most frequent content words of the English and Hungarian versions of the resolution is carried out to emphasise similarities and differences in the series of cultural keywords highlighted in the document.

The six most frequent lexemes of the Italian agenda happen to be the function words *e* 'and' ( $f=1,114$ ), *di* 'of' ( $f=751$ ), *la* 'the' (feminine singular,  $f=397$ ), *il* 'the' (masculine singular,  $f=395$ ), *per* 'for' ( $f=358$ ) and *in* 'in' ( $f=322$ ). The most frequent content word of the Italian frequency list is *sviluppo* 'development'; it is ranked seventh in the frequency list of the 2030 Agenda (Italian) subcorpus and it is followed by function words up to the fourteenth lexeme of the list, which is *paesi* 'countries'. Thus, *paesi* 'countries' is the second most frequent content word of the 2030 Agenda (Italian), as it can be noted from Table 13.

Rank	Content word	<i>f</i>			
1	<i>sviluppo</i> 'development'	290	12	<i>unite</i> 'united'	56
2	<i>paesi</i> 'countries'	167	13	<i>accesso</i> 'access; facilitation'	52
3	<i>sostenibile</i> 'sustainable'	151	14	<i>internazionale</i> 'international'	52
4	<i>via</i> (part of <i>in via di sviluppo</i> 'developing')	83	15	<i>sviluppati</i> 'developed'	51
5	<i>obiettivi</i> 'goals'	76	16	<i>capacità</i> 'capacity; capabilities'	47
6	<i>2030</i> '2030'	75	17	<i>risorse</i> 'resources'	47
7	<i>livello</i> 'level'	70	18	<i>fine</i> (in <i>al fine di</i> 'to' and <i>porre fine</i> 'end')	45
8	<i>agenda</i> 'agenda'	65	19	<i>particolare</i> 'particular'	44
9	<i>globale</i> 'global'	63	20	<i>promuovere</i> 'promote; foster'	40
10	<i>nazioni</i> 'nations'	62			
11	<i>stati</i> 'countries, states'	56			

**Table 13.** The twenty most frequent content words of the 2030 Agenda Corpus (Italian).

The cultural keywords that can be spotlighted in the Italian document match the corresponding English and Hungarian cultural keywords. As in the English and Hungarian cases, also the political, social and cultural background of the Italian Agenda is moulded by the lexemes *sviluppo* 'development', *sostenibile* 'sustainable', *obiettivi* 'goals' and *globale* 'global'. These lexemes play a pivotal role in mirroring some of the most debated cultural and political issues of 2010s in the Italian society.

Once again as in the English and Hungarian cases, also in the 2030 Agenda (Italian) the remaining content words enrich the aboutness of the Italian resolution. Furthermore, the list of the most frequent content words of the Italian version of the 2030 Agenda shares the majority of its lexemes with the English and the Hungarian ones. All documents are characterised by *development-fejlődés/fejlődési-sviluppo*, *countries-országok-paesi/stati*, *sustainable-fenntartható-sostenibile*, *global-globális-globale*, *2030, agenda-agenda-agenda*, *goals-célok-obiettivi*, *international-nemzetközi-internazionale*, *developed-fejlett-sviluppati*, and *levels-szintű-livello*. These content words hint at the core of the document. They explicitly mention the UN's resolution, they point at its main practical outcome and they draw a line between world countries by separating a national from an international dimension and by distinguishing developed countries from developing countries. However, while *developed-fejlett-sviluppati* is common to all three languages, the pair *developing-fejlődő* is to be identified only in English and Hungarian. Nevertheless, this pair is reminded of even if not explicit also in Italian thanks to the lexeme *via*, that is always found in the pattern *in via di sviluppo* 'developing'.

The English and the Italian documents share also the high-frequency content words *access-accesso*, *nations-nazioni*, *promote-promuovere*, and *united-unite*. *nations-nazioni* and *united-unite* refer to the United Nations while the other two lexemes open to the opportunities that can be accessed (*access-accesso*) thanks to the promotion of the 2030 Agenda (*promote-promuovere*). Moreover, the Italian high-ranking content wordlist features also the lexemes *capacità* 'capacity; capabilities', *fine* (in *al fine di* 'to' and *porre fine* 'end'), *risorse* 'resources' and *particolare* 'particular'.

The noun *capacità* 'capacity; capabilities' stands for the ability to achieve goals for the sake of sustainable development. These capabilities are described as *PRODUTTIVO* 'productive' (f=3), *UMANO* 'human' (f=2), *adeguata*, literally 'adequate' (f=1), *commerciali* 'trade-related' (f=1), *effettiva* 'effective' (f=1), *fiscale* 'for tax' (f=1), *innovative* 'innovative' (f=1), *scientifiche* 'scientific' (f=1) and *tecnologiche* 'technologic' (f=1). Thus, capabilities are judged in terms of efficiency and they are mainly applied to an economic or productive area. In these areas, the capacity is one of planning and of putting these plans into practice. The entities possessing this capacity are countries or institutions. The lexeme *capacità* 'capacity; capabilities' is almost always introduced by verbs or by deverbal nouns that indicate an idea of improvement or bolstering expressed through the processes *RAFFORZARE* or *RINFORZARE* 'strengthen' (f=6), *AUMENTARE* 'raise' (f=3), *MIGLIORARE* 'enhance' (f=3), *consolidare* 'strengthening' (f=1),

*incrementando* 'increasing' (f=1), *supportare* 'support' (f=1), and *sviluppare* 'develop' (f=1).

The lexeme *fine* is used in the collocational patterns *al fine di* 'to' (f=23) and *porre fine* 'end' (f=9). The former is used as a prepositional phrase introducing an infinitive and it corresponds to the English *to* or *in order to*. The latter presents some of the points put forward by the agenda. In particular, *porre fine* 'end' encourages the end of harmful states and practices for humans: *alla fame* 'hunger' (f=4), *alla povertà* 'poverty' (f=3), *alla povertà e alla fame* 'poverty and hunger' (f=2), *ad ogni forma di povertà* 'poverty in all its forms' (f=2), *a tutte le morti che si possono prevenire* 'all such preventable deaths' (f=1), *all'abuso, allo sfruttamento, al traffico di bambini e a tutte le forme di violenza e tortura nei loro confronti* 'abuse, exploitation, trafficking and all forms of violence against and torture of children' (f=1), *alla defecazione all'aperto* 'open defecation' (f=1), *a ogni forma di discriminazione nei confronti di donne e ragazze* 'all forms of discrimination against all women and girls' (f=1), *alle epidemie di AIDS, tubercolosi, malaria e malattie tropicali trascurate* 'epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases' (f=1), *al lavoro minorile in ogni sua forma* 'child labour in all its forms' (f=1), *a tutte le forme di malnutrizione* 'all forms of malnutrition' (f=1), *alle morti prevenibili di neonati e bambini sotto i 5 anni di età* 'preventable deaths of newborns and children under 5 years of age' (f=1), *alla schiavitù moderna e alla tratta di esseri umani* 'modern slavery and human trafficking' (f=1). In a single case, also animals and plants are considered (i.e. *al bracconaggio e al traffico delle specie protette di flora e fauna* 'poaching and trafficking of protected species of flora and fauna').

The noun *risorse* 'resources' mainly refers to two kinds of elements: natural resources (*risorse naturali* 'natural resources', f=9, *risorse marine* 'marine resources', f=3, *risorse idriche* 'water resources', f=1) and economic resources (*risorse economiche* 'economic resources', f=5, *risorse finanziarie* 'financial resources', f=1). These resources are usually national (*domestiche* 'domestic', f=3, *interne* 'domestic', f=1). Resources are valued as *pubbliche* 'public' (f=2), *adeguate* 'adequately' (f=1), *disponibili* 'available' (f=2), *esistenti* 'existing' (f=1), and *supplementari* 'additional' (f=1).

*particolare* 'particular' is found in the prepositional phrase *in particolare* 'in particular' in 73% of its occurrences. This prepositional phrase functions as the Hungarian adverb *különösen* 'particularly' and it introduces the same distinctions between countries. In addition, it also precedes nouns referring to different human groups (e.g. *le donne* 'women', f=2, or *i più poveri e vulnerabili* 'the poor and the vulnerable', f=1).

Although most high-ranking content words of the Italian Agenda find correspondences in the lists of the most frequent content words of the English and Hungarian documents, the distinctiveness of the Italian document in terms of content word frequency adds hues of aboutness to the Italian resolution and it emphasises its style. The lexemes *capacità* ‘capacity; capabilities’ and *risorse* ‘resources’ hint at the commitment towards sustainability and at the assessment of the ability and of the means to reach it while the lexeme *fine* tells a story of struggling to end social inequalities; the lexeme *particolare* ‘particular’ reflects the style of the Italian document.

#### **4.1.2. The statistically significant lexemes**

The identification of the cultural keywords of the 2030 Agenda Corpus through the most frequent content words of the collection is bolstered through the extraction of the statistical keywords of the English, Hungarian and Italian versions of the resolution. In other terms, additional cultural keywords are spotlighted through the extraction of the statistically most salient lexemes of the collection. These statistically salient lexemes, in fact, highlight aspects of aboutness and style that stand out in the text or corpus under inquiry in a statistically significant way in comparison with another text or corpus (Scott and Tribble 2006).

Statistical keywords are computed with AntConc’s Keyword List function. Statistical keywords are found using the three versions of the Agenda as study corpora and the matched sections of the EUROPARL7 corpus as reference corpora. In the case of English, the English Agenda is compared with the English section of the EUROPARL7 corpus; for Hungarian, the Hungarian Agenda is contrasted with the Hungarian component of EUROPARL7; in the case of Italian, the Italian Agenda is matched with the Italian section of EUROPARL7. The study and reference corpora are compared with a combination of the Log Likelihood (4 term) statistical significance test ( $p < 0.0001$ ) together with the Difference Coefficient (Relative) effect-size test. The statistical threshold is set to 15.13 for the Log Likelihood statistical test. Only positive keywords (i.e. word types that appear in a statistically more significant way in the study corpus compared to the reference corpus) are considered. Negative keywords (i.e. word types that appear in a statistically less significant way in the study corpus compared to the reference corpus), in fact, are not deemed suitable for the retrieval of cultural keywords. The cultural keywords scattered in the statistical keywords’ list of the 2030 Agenda Corpus are expanded on further by reading their concordance lines and by noting the recurrent linguistic patterns they belonged to. As in the case of the

most frequent content words of the collection, the English Agenda is studied first and then it is adopted as a baseline for the analysis of the statistical keywords of the Hungarian and Italian Agendas.

**4.1.2.1. English**

Thus, the retrieval of the cultural keywords of the 2030 Agenda continues with the analysis of the most salient statistical keywords of the English version of the document.

The comparison between the English Agenda and the English section of the EUROPARL7 corpus results in 272 positive statistical keywords. The complete keyword list enriched with statistical and frequency data can be found in Appendix (Table 44), while Table 14 showcases the twenty most significant keywords.

<b>Rank</b>	<b>Keyword</b>		
1	<i>sustainable</i>	11	<i>developed</i>
2	<i>and</i>	12	<i>agenda</i>
3	<i>2030</i>	13	<i>nations</i>
4	<i>development</i>	14	<i>goal</i>
5	<i>including</i>	15	<i>targets</i>
6	<i>developing</i>	16	<i>forum</i>
7	<i>countries</i>	17	<i>landlocked</i>
8	<i>goals</i>	18	<i>all</i>
9	<i>inclusive</i>	19	<i>levels</i>
10	<i>global</i>	20	<i>technology</i>

**Table 14.** The twenty most significant keywords of the 2030 Agenda Corpus (English).

More than half of the most significant positive keywords of the 2030 Agenda (English) features also in the list of the most frequent lexemes of the document (i.e. *sustainable, 2030, development, developing, countries, goals, global, developed, agenda, nations, levels*). The ones that appear only among the high-ranking keywords are *and, including, inclusive, goal, targets, forum, landlocked, all, technology*.

Among these statistical keywords, *goal, inclusive* and *technology* could be bestowed the role of cultural keywords for the 2030 Agenda together with the ones identified among the most frequent content words of the resolution (i.e. *sustainable, development, goals* and *global*). These additional cultural keywords offer a cross section of the most debated issues of the time when the resolution was released: the noun *goal* recalls the cultural keyword *goals* and it reminds of the seventeen Sustainable Development Goals that have been guiding Western political planning since 2015; the

adjective *inclusive* presents the all-encompassing attitude that is required nowadays in every sphere of life so that no being or society is excluded from progress; the noun *technology* embodies one of the human products that is devoted more efforts of improving in Western contemporary societies. These cultural keywords emphasise that sustainable development should be aimed focusing on the seventeen sustainable development goals elaborated on by the United Nations, they recognise technology as a leading source of sustainable development and inclusiveness as a leading quality of sustainable development.

*inclusive* belongs to adjective phrases that modify the noun phrases displayed in Table 15.

<u>Noun phrase</u>	<i>f</i>		
<i>economic growth</i>	10	<i>growth</i>	1
<i>societies</i>	6	<i>learning environments</i>	1
<i>industrialization</i>	3	<i>nature</i>	1
<i>quality education</i>	3	<i>process</i>	1
<i>institutions</i>	2	<i>regional processes</i>	1
<i>reviews</i>	2	<i>spaces</i>	1
<i>decision-making</i>	1	<i>trading-system</i>	1
<i>follow-up</i>	1	<i>urbanization</i>	1
		<i>world</i>	1

**Table 15.** Noun phrases modified by *inclusive* in the 2030 Agenda Corpus (English).

Inclusiveness is construed as a property deeply intertwined with sustainability, reliability, safety and peace and it is associated with economic matters (*economic growth; industrialization; trading system*) but also to social matters (*societies; growth* in relation to migration). Inclusiveness regards education (*quality education; learning environments*). It involves political actions (*institutions; reviews; decision-making; follow-up; nature; process; regional processes*) that can accompany the evolution of settlements (*spaces; urbanization; world*).

The adjective *inclusive* is frequently associated with other adjectives with the conjunction *and*. When *and* follows *inclusive*, the conjunction creates the adjective phrases *sustained, inclusive, and sustainable* (*f*=6), *inclusive and sustainable* (*f*=4) *inclusive and equitable* (*f*=3), *inclusive, safe, resilient and sustainable* (*f*=2), *inclusive and sustained* (*f*=1), *open, inclusive, participatory and transparent* (*f*=1), *responsive, inclusive, participatory and representative* (*f*=1), *safe, inclusive, and accessible* (*f*=1), *safe, non-violent, inclusive and effective* (*f*=1), *universal, rules-based, open, transparent, predictable, inclusive, non-discriminatory and equitable* (*f*=1). When *and* precedes *inclusive*, the conjunction contributes to the adjective phrases *peaceful and inclusive* (*f*=3), *peaceful, just and inclusive*

(*f=3*), *effective, accountable and inclusive (f=2), regular and inclusive (f=2), coherent, efficient and inclusive (f=1), sustained and inclusive (f=1)*.

*technology* is found in noun phrases constructed with the conjunction *and*. When *and* follows *technology*, the noun belongs to the groups *science, technology and innovation (f=13), communications technology and global interconnectedness (f=1), knowledge, expertise, technology and financial resources (f=1), mitigation, adaptation, finance, technology development and transfer and capacity building (f=1), new technology and financial services (f=1), rural infrastructure, agricultural research and extension services, technology development and plant and livestock gene banks (f=1) technology bank and science (f=1)*. When *and* precedes *technology*, the only noun phrases emerging is *research and technology (f=1)*.

The noun *technology* is modified by the noun and adjective phrases that can be seen in Table 16.

<u>Noun and adjective phrases</u>	<i>f</i>		
<i>information and communications</i>	5	<i>clean-energy</i>	1
<i>enabling</i>	2	<i>domestic</i>	1
<i>marine</i>	2	<i>fossil-fuel</i>	1
		<i>new</i>	1

**Table 16.** Noun and adjective phrases modifying *technology* in the 2030 Agenda Corpus (English).

*technology* modifies the noun phrases reproduced in Table 17.

<u>Noun phrases</u>	<i>f</i>		
<i>facilitation</i>	4	<i>bank</i>	1
<i>development</i>	2	<i>development and transfer</i>	1
		<i>needs and gaps</i>	1

**Table 17.** Noun phrases modified by *technology* (2030 Agenda Corpus-English).

Based on the noun and adjective phrases that modify and that are modified by *technology* and on the noun phrases it belongs to, *technology* is encouraged (*facilitation, development, development and transfer*) especially for the research and employment of natural and financial resources (*marine, clean-energy, fossil-fuel*) but also for the spread of communication (*information and communications*). Its evolution might be of use to fill the *needs and gaps* of the countries involved in sustainable development (*enabling, domestic*).

The remaining statistical keywords contribute to the aboutness and to the style of the resolution. As far as style is concerned, the function words *and, including* and *all* extend the meaning of a lexeme and they convey an idea of inclusiveness. In terms of aboutness, *targets* points at the core of the 2030 Agenda, namely at the seventeen

sustainable development goals and their targets; *forum* launches the institutional and official coverage of the political actions that need to be taken in order for sustainability to be achieved; *landlocked* describes the geographical characteristics of some of the countries that the resolution deals with.

#### 4.1.2.2. Hungarian

The detection of cultural keywords among the statistically significant lexemes of the 2030 Agenda Corpus was then enhanced with the study of the statistical keywords computed for the Hungarian version of the 2030 Agenda.

The statistical comparison between the 2030 Agenda (Hungarian) subcorpus and the Hungarian section of the EUROPARL7 corpus produces a set of 249 keywords. Table 18 displays the twenty strongest keywords of the list.

Rank	Keyword	Rank	Keyword
1	<i>fenntartható</i> 'sustainable'	12	<i>fejlődés</i> 'development'
2	<i>2030</i> '2030'	13	<i>országok</i> 'countries'
3	<i>agenda</i> 'agenda'	14	<i>globális</i> 'global'
4	<i>és</i> 'and'	15	<i>valamint</i> 'and'
5	<i>ig</i> 'by'	16	<i>szárazfölddel</i> (part of <i>szárazfölddel körülvett</i> 'landlocked')
6	<i>fejlődő</i> 'developing'	17	<i>elismerjük</i> '(we) recognize; (we) acknowledge'
7	<i>beleértve</i> 'including'	18	<i>forum</i> 'forum'
8	<i>fejlett</i> 'developed'	19	<i>körülvett</i> (part of <i>szárazfölddel körülvett</i> 'landlocked')
9	<i>fejlesztési</i> 'development'	20	<i>célok</i> 'goals'
10	<i>alcélok</i> 'targets'		
11	<i>legkevesebbé</i> 'least'		

**Table 18.** The twenty most significant keywords of the 2030 Agenda Corpus (Hungarian).

Half of the most significant keywords of the Hungarian Agenda belong to the list of the most frequent content words of the Hungarian document too (i.e. *fenntartható* 'sustainable', *2030* '2030', *agenda* 'agenda', *fejlődő* 'developing', *fejlett* 'developed', *fejlesztési* 'development', *fejlődés* 'development', *országok* 'countries', *globális* 'global', *célok* 'goals'). These lexemes correspond also to ten of the most frequent and statistically significant words of the English Agenda.

In the case of the Hungarian version of the resolution, the most salient keywords of the 2030 Agenda (Hungarian) subcorpus do not add any cultural keyword to the list that has been collected among the most frequent content words of the document. Thus, the cultural keywords identified for Hungarian remain

*fenntartható* 'sustainable', *fejlesztés/fejlődés* 'development', *globális* 'global' and *célok* 'goals'.

Among the keywords that do not belong to the list of the most frequent content words of the Hungarian Agenda, seven match keywords of the English Agenda (*és* 'and', *beleértve* 'including', *alcélok* 'targets', *valamint* 'and', *szárazfölddel* and *körülvevett*, which belong to the adjective phrase *szárazfölddel körülvevett* 'landlocked', *forum* 'forum'). The use of these keywords in the two versions of the resolution is comparable.

Three keywords are typical of the Hungarian keyword list: the suffix *ig* 'by', the function word *legkevesbé* 'least' and the content word *elismerjük* '(we) recognize; (we) acknowledge'. The suffix *ig* 'by' is a bound morpheme that marks the terminative case. It can express anterior-duration both in relation to future and to past events. In the Hungarian Agenda *ig* 'by' follows the numerals 2030 (in *2030-ig* 'by 2030',  $f=63$ ), 2020 (in *2020-ig* 'by 2020',  $f=22$ ), 2025 (in *2025-ig* 'by 2025',  $f=3$ ), and 2017 (in *2017-ig* 'by 2017',  $f=1$ ) to determine the deadline of certain goals' achievement. *2030-ig* 'by 2030' is surely the most meaningful deadline mentioned by the resolution, as it refers to the year when the achievements of the 2030 Agenda will need to be assessed.

The function word *legkevesbé* 'least' modifies the adjective *fejletl* 'developed' in the adjective phrase *legkevesbé fejletl* 'least developed' ( $f=44$ ). This adjective phrase always modifies the lemma *ORSZÁG* 'country' in the noun phrase *LEGKEVESBÉ FEJLETT ORSZÁG* 'least developed country', which refers to those countries that stand at the bottom of the developing-developed continuum. This continuum is frequently reproduced in the list of the most frequent and statistically significant lexemes of the Hungarian Agenda.

*elismerjük* '(we) recognize; (we) acknowledge' is a verb form used in the 2030 Agenda (Hungarian) subcorpus to clarify the starting point of the enterprise towards the achievement of sustainable development or the steps that need to be taken aiming at it.

These three words do not contribute to the aboutness of the Hungarian Agenda but they help form the style of the document. *ig* 'by' serves to punctuate the unfolding of the content of the Hungarian Agenda as it spells out the timings of the journey towards reaching sustainability. *legkevesbé* 'least' qualifies some of the countries involved in the process described by the 2030 Agenda. *elismerjük* '(we) recognize; (we) acknowledge' reflects the attitude of the United Nations in relation to their resolution; considering that no modal operators appear together with the verb, I would argue that primary tense is preferred over modality for the most significant word forms of the UN's resolution.

### 4.1.2.3. Italian

The pinpointing of cultural keywords among the statistically significant lexemes of the 2030 Agenda proceeds with the exploration of the statistical keywords calculated for the 2030 Agenda (Italian) subcorpus.

The statistical comparison between the 2030 Agenda (Italian) subcorpus and the Italian section of EUROPARL7 results in 251 keywords. Table 19 showcases the most salient statistical keywords of the Italian Agenda.

Rank	Keyword		
1	<i>sviluppo</i> 'development'	11	<i>forum</i> 'forum'
2	2030 '2030'	12	<i>globale</i> 'global'
3	<i>sostenibile</i> 'sustainable'	13	<i>Nazioni</i> 'nations'
4	<i>e</i> 'and'	14	<i>Unite</i> 'united'
5	<i>entro</i> 'by'	15	<i>obiettivi</i> 'goals'
6	<i>sviluppati</i> 'developed'	16	<i>sostenibili</i> 'sustainable'
7	<i>agenda</i> 'agenda'	17	<i>insulari</i> 'island'
8	<i>via</i> (part of <i>in via di sviluppo</i> 'developing')	18	<i>riconosciamo</i> '(we) recognize; (we) acknowledge'
9	<i>traguardi</i> 'targets'	19	<i>capacità</i> 'CAPACITY; capabilities'
10	<i>paesi</i> 'countries'	20	<i>accesso</i> 'access; facilitation'

**Table 19.** The twenty most significant keywords of the 2030 Agenda Corpus (Italian).

Most of the salient statistical keywords of the Italian Agenda are shared by the list of the most frequent content words of the document (i.e. *sviluppo* 'development', 2030 '2030', *sostenibile* 'sustainable', *sviluppati* 'developed', *agenda* 'agenda', *via*, which is part of *in via di sviluppo* 'developing', *paesi* 'countries', *globale* 'global', *Nazioni* 'Nations', *Unite* 'United', *obiettivi* 'goals', *capacità* 'CAPACITY; capabilities', *accesso* 'access; facilitation'). Others do not appear among the most frequent content words of the Italian Agenda, but they correspond to some of the most significant statistical keywords of the English Agenda (i.e. *e* 'and', *traguardi* 'targets', *forum* 'forum', *sostenibili* 'sustainable') and they are employed in Italian in the same way as their translational equivalents are used in English.

As in the Hungarian case, also the most salient statistical keywords of the 2030 Agenda (Italian) subcorpus do not provide additional cultural keywords to the set of cultural keywords that have already been identified for the Italian Agenda among the most frequent content words of the document.

Nevertheless, among the most significant statistical keywords of the Italian version of the Agenda, three lexemes appear anew: the function word *entro* 'by' and the content words *insulari* 'island' and *riconosciamo* '(we) recognize; (we) acknowledge'.

The preposition *entro* 'by' functions in a similar way compared to the Hungarian suffix *ig* 'by'. *entro* 'by' marks time specifications for an anterior-duration and it serves as head of prepositional phrases modified by the noun phrases *il 2030*, literally '(the) 2030' (in *entro il 2030* 'by 2030',  $f=64$ ), *il 2020*, literally '(the) 2020' (in *entro il 2020* 'by 2020',  $f=20$ ), *il 2025*, literally '(the) 2025' (in *entro il 2025* 'by 2025',  $f=1$ ) and *il 2017*, literally '(the) 2017' (in *entro il 2017* 'by 2017',  $f=1$ ). These prepositional phrases set the states, events and processes encoded in the sentences they modify in specific time-frames, specifying that these states, events and processes should happen before a certain time. In particular, *entro il 2030* 'by 2030' recurrently sets the pace that the actions towards sustainable development should have.

The verb form *riconosciamo* '(we) recognize; (we) acknowledge' corresponds to the Hungarian *elismerjük* '(we) recognize; (we) acknowledge'. It is used to acknowledge with certainty the characteristics of the context on which the endeavour of acting in favour of sustainable development is rooted.

The adjective *insulari* 'island' modifies the noun phrase *piccoli stati* 'small states', thus contributing to the noun phrase *piccoli stati insulari* 'small island states' ( $f=23$ ). These small island states are written to be *in via di sviluppo* 'developing' and they are overall defined as *piccoli stati insulari in via di sviluppo* 'small island developing states' ( $f=21$ ).

*entro* 'by' and *riconosciamo* '(we) recognize; (we) acknowledge' enrich the style of the Italian Agenda: the former points out the deadlines required for the achievement of sustainable development; *riconosciamo* '(we) recognize; (we) acknowledge' introduces the statement of believes that the United Nations claim in the 2030 Agenda. On the other hand, *insulari* 'island' adds to the aboutness of the Italian Agenda as it boosts the way countries taking part in this enterprise can be described.

### 4.1.3. Economic, social and environmental lexemes

After the collection of the cultural keywords moulding the 2030 Agenda Corpus through the most frequent and statistically significant lexemes of its English, Hungarian and Italian sections, economic, social and environmental lexemes are

collected to study which dimension of sustainable development is given more emphasis in the 2030 Agenda in terms of lexical frequency and statistical salience.

The lexemes that refer to the economic, social and environmental dimensions of sustainable development are searched for only in the English version of the 2030 Agenda, which is deemed representative of the resolution for the purposes of this inquiry. Economic, social and environmental lexemes are listed by collecting the correlated most frequent and statistically most salient word types. In addition, I gather lexemes about economy, society and environment by reading through the whole frequency list of the document. This is feasible because of the dimensions of the corpus, that counts only 2,838 word types (on the features of the 2030 Agenda Corpus see § 2.1.2).

The economic dimension of sustainable development is mirrored in the 2030 Agenda both with lexemes that are solely economic and with others that are used also in non-economic contexts. The statistical keywords straightforwardly shaping the economic dimension of sustainable development are *affordable, debt, economic, finance, GNI, gross, income, poor, poverty, productivity, stakeholder, stakeholders*.

A perusal of the resolution's complete wordlist shows that not only are economic matters referred to with these 12 lexemes, but also with other 95 word types and a total number of 541 running tokens. Some of these word types can also be used in non-economic contexts but their economic uses are reported in the following paragraphs. In addition, these lexemes can either refer to economy in general or they can regard specific areas of economy or peculiar economic issues.

Economy in general is hinted at through the word types *economic* ( $f=60$ ), *growth* (part of the lexical item *economic growth*,  $f=13$ ), *industrialization* ( $f=3$ ), *development* (found in *economic development*,  $f=2$ ), *benefits* (in *economic benefits*,  $f=1$ ), *economies* ( $f=1$ ), *losses* (of *economic losses*,  $f=1$ ), and *spending* (belonging to *development spending*,  $f=1$ ).

Then, economic development can be driven by finance and banking, which are introduced through the lexemes *financial* ( $f=25$ ), *income* ( $f=13$ ), *finance* ( $f=8$ ), *financing* ( $f=8$ ), *markets* ( $f=6$ ), *services* (part of the lexical item *financial services*,  $f=6$ ), *flows* (in *financial flows*,  $f=1$ ), *relief* (belonging to *debt relief*,  $f=3$ ), *bank* (of *World Bank*,  $f=1$ ), *banking* ( $f=1$ ), *borrowing* (in *borrowing countries*,  $f=1$ ), *budgets* ( $f=1$ ), *capitalization* ( $f=1$ ), *credit* ( $f=1$ ), *creditors* ( $f=1$ ), *debtors* ( $f=1$ ), *indebted* ( $f=1$ ), *lenders* ( $f=1$ ), *microfinance* ( $f=1$ ), and *monetary* ( $f=1$ ).

The money exchange that characterizes finance and banking is also to be found in the trade system. The trade system is represented in the 2030 Agenda with the word

types *trade* (f=19), *markets* (f=6), *market* (f=4), *trading* (f=2), *commodity* and *derivatives* (both in *food commodity markets and their derivatives*, f=1), *duty* (of *duty-free*, f=1), *free* (part of the lexical items *duty-free*, f=1, and *quota-free*, f=1), *liberalization* (in *trade-liberalization*, f=1), *price* (f=1), and *retail* (f=1).

Economic development has also to do with management and business economics, which are referred to through the lexemes *production* (f=14), *stakeholders* (f=13), *productive* (f=11), *sector* (in *private sector*, f=9, and *business sector*, f=2), *stakeholder* (f=9), *business* (f=7), *industrial* (f=7), *enterprises* (f=4), *products* (f=3), *micro* (belonging to the lexical item *micro-enterprises*, f=3), *companies* (f=2), *cooperatives* (f=2), *costs* (f=2), *entrepreneurship* (f=2), *goods* (f=2), *multinationals* (f=2), *businesses* (f=1), *commodities* (f=1), *diversification* (of *industrial diversification*, f=1), *industries* (f=1), *industry* (f=1), *processes* (of *industrial processes*, f=1) and *producers* (f=1).

Also macroeconomics is touched upon by the resolution in relation to sustainable development thanks to the lexical items *work* (f=22), *employment* (f=13), *consumption* (f=12), *labour* (f=11), *ODA* (f=8), *gross* (f=7, of *gross domestic product* or *gross national income*), *investment* (f=6), *value* (f=6), *GNI* (f=5), *product* (of *gross domestic product*, f=5), *productivity* (f=5), *jobs* (f=4), *wealth* (f=3), *workers* (f=3), *exports* (f=2), *gains* (f=2), *job* (f=2), *providers* (in *ODA providers*, f=2), *unemployment* (f=2), *workforce* (f=2), *consume* (belonging to *produce and consume goods and services*, f=1), *consumer* (f=1), *export* (in *agricultural export subsidies*, f=1), *fiscal* (f=1), *imports* (f=1), *incomes* (f=1), *macroeconomic* (f=1), *produce* (part of *produce and consume goods and services*, f=1), *revenue* (f=1), *tax* (f=1), *taxation* (f=1), *unpaid* (f=1), *wage* (f=1), and *working* (of *working environments*, f=1).

The economic dimension of sustainable development is also said to be ruled by private law through lexemes like *debt* (f=20), *ownership* (f=6), *property* (f=5), *insurance* (f=1), and *provisions* (f=1), and it is also intertwined with political and social issues like welfare, which strives to provide basic forms of human wellbeing in order to ease social distress (in *poverty*, f=28, *affordable*, f=15, *poor*, f=9, *subsidies*, f=7, *poorest*, f=5, *prosperity*, f=5, and *prosperous*, f=1).

The social and environmental dimensions of sustainable development were identified in the 2030 Agenda thanks to the lexemes that mention human beings, animals, plants and natural elements. Among the keywords, the terms referring to human beings are *boys*, *child*, *generation*, *girls*, *migrants*, *peoples*, *persons*, *women*, *youth*. The only keyword that gives voice to animals and plants is *species* while the world's natural elements are represented by *biodiversity*, *climate*, *ecosystems*, *forests*, *island*, *land*, *oceans*, *planet*, *water* and *world*.

A thorough reading of the document's wordlist allows to add a wealth of other word types to the categories of humans, animals, plants and natural elements. The list of human beings counts 44 word types (among which 34 lemmas) and a total number of 274 tokens (Table 20). The word types representing animals are 6 and they occur 14 times overall (Table 21). Plants are reported with 7 word types in 14 tokens (Table 22). 26 word types (and 190 tokens) refer to natural elements (Table 23). Human beings, animals, plants, and natural elements appear overall 492 times in the form of 83 word types. Some of these are rather frequent in the document. Others are hapax legomena.

As it can be seen in Table 19, first of all, the list of human beings featuring in the 2030 Agenda includes generic labels (e.g. *humanity, individuals, people, peoples, person, persons, population*). Among these, the word-forms *people, peoples* and *persons* are included in different patterns and they display a slightly different meaning. *people* stands for a generic group of human beings, irrespective of their individual characteristics (as in *all people, f=8*). *peoples*, on the other hand, refers to groups of humans that stick together because of specific geopolitical reasons (e.g. *indigenous peoples, f=6*). As in the case of *peoples*, also *persons* is used for groups of humans that share common characteristics, although these are usually not geopolitical (e.g. *persons with disabilities, f=6*).

Rank	Word type	f			
1	<i>people</i>	64	22	<i>generations</i>	3
2	<i>women</i>	32	23	<i>refugees</i>	3
3	<i>girls</i>	15	24	<i>representatives</i>	3
4	<i>children</i>	12	25	<i>workers</i>	3
5	<i>persons</i>	12	26	<i>adults</i>	2
6	<i>peoples</i>	11	27	<i>individuals</i>	2
7	<i>child</i>	10	28	<i>leaders</i>	2
8	<i>youth</i>	10	29	<i>president</i>	2
9	<i>poor</i>	9	30	<i>workforce</i>	2
10	<i>communities</i>	8	31	<i>families</i>	1
11	<i>men</i>	8	32	<i>girl</i>	1
12	<i>generation</i>	6	33	<i>herders</i>	1
13	<i>migrants</i>	6	34	<i>infants</i>	1
14	<i>boys</i>	5	35	<i>learners</i>	1
15	<i>humanity</i>	5	36	<i>newborns</i>	1
16	<i>population</i>	5	37	<i>organizers</i>	1
17	<i>community</i>	4	38	<i>participants</i>	1
18	<i>family</i>	4	39	<i>person</i>	1
19	<i>citizens</i>	6	40	<i>producers</i>	1
20	<i>farmers</i>	3	41	<i>soldiers</i>	1
21	<i>fishers</i>	3	42	<i>teacher</i>	1
			43	<i>teachers</i>	1

**Table 20.** List of human beings of the 2030 Agenda Corpus (English).

A general reference to humans is in some cases boosted by hinting at their age. Human beings are thus classified as *adults, child, children, generation, generations, infants, newborns, and youth*. These age-groups are frequently paired and treated as if they were two sides of a coin through the use of the conjunction *and* (as in *youth and adults*,  $f=1$ ). The same applies also to those cases when human beings are mentioned in relation to their gender (e.g. *boys, girl, girls, men, woman, women*). For example, in *men and women* ( $f=2$ ) the conjunction *and* keeps together male adults and female adults and it introduces to the idea that they belong to the same group of human beings who deserve to obtain the same rights thanks to the advances suggested by the 2030 Agenda.

The human beings listed in Table 20 are distinguished also according to their wealth (*poor*). In 78% of its occurrences (7 out of 9), in fact, *poor* is used as a collective noun. It is introduced by the article *the* and it refers to a category of people living on a very low personal wealth.

Some word types belonging to this category classify human beings according to their socio-political status. The words *citizens, communities, families, family, leaders, migrants, organizers, participants, president, refugees, and representatives* introduce humans in discourse by foregrounding their socio-political role. For example, *communities* encompass smaller groups (e.g. *families*) and single *individuals* who are not depicted in their private sphere but in the public dimension of their belonging to the world as social beings. These communities (especially in the singular form *community*) can also be the gathering of individuals with a specific position in society, as in the cases of the *scientific community* ( $f=2$ ) or *academic community* ( $f=1$ ).

Human beings are also grouped according to their employment. They can be *farmers, fishers, herders, learners, producers, soldiers, teacher, teachers, workers, and workforce*. The most prominent kinds of employment that humans engage with or are encouraged to engage with in the Agenda are either generic (*workers* and *workforce*) or linked to the management of animals and plants (*farmers, fishers, herders, and producers*), to education (*learners, teacher, and teachers*), and to peace (*soldiers*). In fact, an educated life in peace and with the essential resources for nutrition is one of the main goals of the document.

Pronouns are not included in the list of human actors. For example, *we, our, ourselves* refers to the group of individuals and nations cooperating for the

implementation of the 2030 Agenda, but they are not counted in the list of the words referring to human beings. They are too generic and wide in coverage. Also words referring not to socially or politically engaged beings but to the socio-political environment where the human beings live (e.g. *countries* or *society*) are excluded from the count.

As Table 21 shows, the word types referring to animals include generic terms like *animals* as well as more scientifically sound words like *fauna*.

Rank	Word type	<i>f</i>			
1	<i>species</i>	7	4	<i>fauna</i>	1
2	<i>wildlife</i>	3	5	<i>fish</i>	1
3	<i>animals</i>	1	6	<i>livestock</i>	1

**Table 21.** List of animals of the 2030 Agenda Corpus (English).

The list mentions a specific species of animals, namely *fish*, but it also encompasses the variety of the animals spread throughout the world through the word type *species*. *species*, however, is used both for animals and plants. It is about animals in *wildlife and other living species* ( $f=1$ ) and *domesticated animals and their related wild species* ( $f=1$ ). It refers to both animals and plants in *protected species of flora and fauna* ( $f=1$ ) and in *threatened species* ( $f=1$ ). It is employed solely for plants in *invasive alien species* ( $f=1$ ) and in *priority species* ( $f=1$ ).

In addition, also the relationship that ties animals and human beings is accounted for in this group. In this sense, animals are called *wildlife* when they are not constrained by human rules and they are called *livestock* if, on the other hand, they are exploited as farming resources.

Pronouns were excluded also from the list of animal actors. In fact, even though *they* and *their* can refer to animals sometimes (as in *domesticated animals and their related wild species*,  $f=1$ ), they are used in too wide a variety of cases.

Even word types like *resources* and *products* are omitted as they refer to animals not as living beings but as means for economic development (as in *seas and marine resources*,  $f=2$ , or *illegal wildlife products*,  $f=1$ ). The same holds true also for the word *stocks* (in *fish stocks*,  $f=1$ ), that envisages fish as resources to pile up instead of treating them as living creatures whose life is worth on its own.

In the 2030 Agenda there are general references to plants (e.g. *plant* and *plants*) but also more abstract and scientific ways of referring to them (*flora*; see Table 22).

Rank	Word type	<i>f</i>			
			4	<i>forest</i>	1
1	<i>species</i>	7	5	<i>plants</i>	1
2	<i>plant</i>	2	6	<i>seed</i>	1
3	<i>flora</i>	1	7	<i>seeds</i>	1

**Table 22.** List of plants of the 2030 Agenda Corpus (English).

As in the case of animals, the word type *species* hints at the variety of plants living in our world. The same variety is also suggested by words like *forest* and by the elements from which plants originate, namely *seed* and *seeds*.

Table 23 presents the word types of the 2030 Agenda that concern natural elements.

Rank	Word type	<i>f</i>			
			14	<i>soil</i>	4
1	<i>climate</i>	26	15	<i>earth</i>	3
2	<i>island</i>	22	16	<i>mountains</i>	3
3	<i>water</i>	22	17	<i>ocean</i>	3
4	<i>environment</i>	14	18	<i>sea</i>	3
5	<i>land</i>	14	19	<i>drylands</i>	2
6	<i>planet</i>	13	20	<i>habitats</i>	2
7	<i>ecosystems</i>	11	21	<i>lakes</i>	2
8	<i>biodiversity</i>	10	22	<i>rivers</i>	2
9	<i>nature</i>	9	23	<i>wastewater</i>	2
10	<i>oceans</i>	7	24	<i>wetlands</i>	2
11	<i>air</i>	4	25	<i>mountain</i>	1
12	<i>freshwater</i>	4	26	<i>weather</i>	1
13	<i>seas</i>	4			

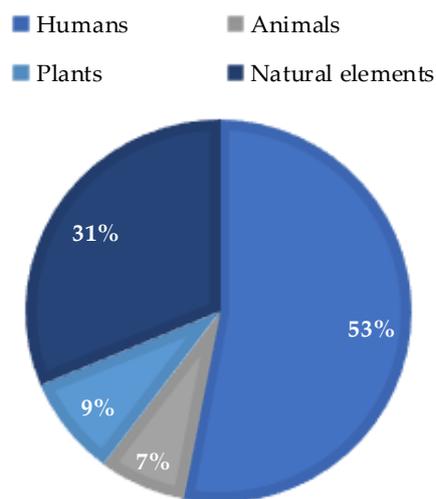
**Table 23.** List of natural elements of the 2030 Agenda Corpus (English).

These natural elements set the document and the actions that it recommends in our world (*earth, nature, and planet*). They describe different kinds of environment (apart from the very *environment*, also *drylands, island, lakes, mountain, mountains, ocean, oceans, rivers, sea, seas, wetlands*). They also mention the various elements that constitute these environments (*air, freshwater, land, soil, wastewater, water*) and their richness in terms of relationship between the elements (*biodiversity, ecosystems, habitats*). Also *climate* and *weather* can be regarded as natural elements, making up for the climatic conditions of the world.

*climate*, however, features as a modifier of *change* in the *climate change* expression in 77% of the occurrences. Consequently, apart from helping shape the list of natural

elements that the 2030 Agenda includes in its unfolding, it should also be deemed an instance of environmental concern.

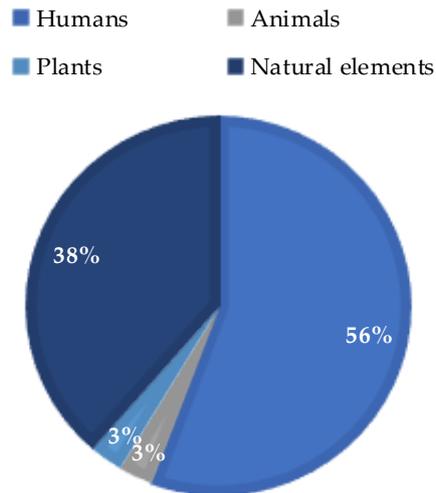
As it has already been stated, the word types referring to human beings are 44, those reproducing animals and plants are 6 and 7 respectively, and those reporting natural elements are 26. This accounts for a total number of 83 different lexemes moulding the ecological panorama of the human, animal, plant and natural beings worth mentioning according to the 2030 Agenda. Considering the relative number of the lexemes across the four groups, Graph 4 allows to visualise the distribution of the word types within the ecology of the Agenda.



**Graph 4.** Distribution of the word types belonging to the categories of humans, animals, plants and natural elements.

The graph shows that the most represented group is the one including human beings, which makes for more than half of the whole. This is followed in variety by the class of natural elements, 31% of the whole, and then by those of plants and animals.

Graph 5, on the other hand, displays the distribution of all single occurrences of the word types belonging to the groups of humans, animals, plants, and natural elements. These were calculated to be 274 for humans, 14 for animals, 14 for plants, and 190 for natural elements, with a total number of 492 running tokens.



**Graph 5.** Distribution of the tokens belonging to the categories of humans, animals, plants and natural elements.

As in Graph 4, also in this case the human beings stand out compared to animals, plants, and natural elements. However, the proportion of tokens included in the category of natural elements is higher than in the case of the word types. This depends on the high frequency of terms like *climate*, *island*, and *water*. Among these, *climate* is associated with *change* in 20 out of 26 cases and it increases thus the number of tokens associated with this group irrespective of its most peculiar nature. The word *island* is always found in the pattern *small island developing States* and so it drags the group of natural elements towards one of politics.

#### 4.2. Meaning by collocation of *sustainable* in the 2030 Agenda Corpus

It has already been observed that *sustainable* stars as one of the cultural keywords displayed by the 2030 Agenda Corpus. This comes as no surprise: the 2030 Agenda is a cornerstone of the international debate on sustainability and a tool that the United Nations have offered to frame urgent actions in favour of sustainable development. The specificity of the urgent actions recommended by the UN depends on the meaning that the institution has assigned to words like *sustainable*, *sustainability* and *sustainable development*.

The meaning that the adjective *SUSTAINABLE* (from now on simply *sustainable* because of its invariable form in English) has been bestowed in the 2030 Agenda is outlined in this section thanks to the analysis of its collocational patterns. The same is

done for the Hungarian and the Italian translational equivalents of *sustainable*, namely *FENNTARTHATÓ* and *SOSTENIBILE* respectively.

Within the realm of corpus semantics (Stubbs 2015), the meaning of a lexical item is sketched by studying its collocational patterns. Collocational patterns are analysed adopting Sinclair's (1991) method and enriching it with Partington's (2017) theory of evaluative prosody: collocates pinpoint the lexical preferences of the node, while colligates show its grammatical preferences; semantic preference identify the strongest semantic associations of the node; evaluative prosody explores the pragmatic function of the lexical item built by the node and its collocates and it frequently focuses on the connotation of the pattern (for more on this see the discussion in § 1.2). This analysis of collocational patterns allows to outline the meaning that the lexical items under inquiry have assumed in the 2030 Agenda Corpus thanks to their co-occurring together with a limited set of words.

Unlike in the investigation of the cultural keywords of the Agenda, where recurrent patterns are identified by means of frequency, the extraction of the collocational patterns of *sustainable* and its Hungarian and Italian translational equivalents is carried out in search for statistical significance and not for frequency of co-occurrence. This is done to highlight the most peculiar lexico-grammatical behaviour of the lexical items.

First of all, the meaning by collocation of the English *sustainable*, of the Hungarian equivalent *FENNTARTHATÓ* and of the Italian equivalent *SOSTENIBILE* is studied with the word sketches provided by the Sketch Engine platform. Word sketches automatically organise collocational patterns according to the grammatical relationships that bind node and collocates; they allow to observe the syntactic and semantic profile of the nodes at first glance. Thus, word sketches offer a snapshot of the meaning of the adjective: they represent the most meaningful grammatical and lexical patterns that surround it.

Word sketches are extracted for the English, Hungarian and Italian sections of the 2030 Agenda Corpus with the LogDice association measure and a statistical threshold of 6.0. The Word sketch Tool is also set with a frequency threshold of 5: according to this frequency threshold, collocates should co-occur at least five times with the node, and they should appear at least five times in the corpus independently of the node. Sketch Engine's word sketches are found by typing a simple query (i.e. *sustainable* for English, *fenntartható* for Hungarian and *sostenibile* for Italian) and by requiring the platform to consider the typed lexical item as a word.

Then, the meaning by collocation of the English *sustainable*, of the Hungarian equivalent *FENNTARTHATÓ* and of the Italian equivalent *SOSTENIBILE* is strengthened by analysing their collocation networks. Collocation networks enhance the collocational patterns highlighted by the word sketches with additional collocates and they arrange the complete list of a node's collocates in the form of a web. They allow to better visualise the nodes' collocational patterns and to organise them in terms of colligation, semantic preference and evaluative prosody.

The English, Hungarian and Italian sections of the 2030 Agenda Corpus are studied in search for collocation networks with the #LancsBox software. The collocation networks retrieved with #LancsBox are found with the Z association measure, a statistical threshold of 10.0, a frequency threshold of 5 for both the sheer collocate and for the co-occurrence of node and collocate. Collocational patterns are searched for within a collocation window of  $\pm 1$ ,  $\pm 3$ ,  $\pm 5$  and they are reproduced both in the form of tables and graphs. When needed, the collocation window is further expanded to comply with the morphological and syntactic characteristics of the languages. The English Agenda is queried with the string */sustainable/*, the Hungarian Agenda with the string */fenntartható\*/* and the Italian Agenda with the string */sostenibile|sostenibili/*.

The use of the wildcard for the search of the Hungarian */fenntartható\*/* is justified by the high morphological variability of the lemmas in this language. Once again due to morphological variability, the collocation network of the Italian adjective *SOSTENIBILE* 'sustainable' is searched for with the string */sostenibile|sostenibili/*, which encompasses both the singular and the plural form of the adjective. The string differs from the one employed for the corresponding Hungarian lexeme because of the diverse nature of morphological variability in Hungarian and Italian. Hungarian is an agglutinative language: it tends to add or change the grammatical and functional characteristics of a word with the use of bound morphemes that bear a single value and that usually follow the word stem. Italian is a fusive language: multiple pieces of grammatical information cluster in a single bound morpheme that usually follows the word stem, while pieces of functional information are generally encoded through free morphemes (e.g. prepositions).

Collocation networks retrieved for Hungarian and Italian are studied in detail only when they differ from the collocational patterns of the corresponding English lexical items.

When extracting collocational patterns with Sketch Engine and with #LancsBox, the description of the features of the collocation networks looks like the following label: Z(10.0), 1L-1R, NC5-C5. The association measure is mentioned first, followed by the statistical threshold, by the dimension of the collocation window, and by the frequency thresholds. In reporting the collocational patterns, small caps are used for lemmas.

The collocational patterns of the English *sustainable* are deemed baseline for tracing the meaning by collocation of the lexical item. Thus, the collocational patterns of the Hungarian and of the Italian translational equivalents are studied in comparison with the English collocational patterns and they are devoted an in-depth analysis only in case they prove different from the baseline.

The analysis of the collocational patterns retrieved with the word sketches and with the collocation networks is then boosted with a closer reading of the collocation's concordance lines. This implies an additional inclusion of frequency as an association measure for the organization of the collocates. Concordances are extracted with AntConc's Concordance Tool.

#### **4.2.1. The English *SUSTAINABLE***

The study of the meaning by collocation of *sustainable* in the 2030 Agenda Corpus starts with the exploration of the lexeme's collocational patterns in the English version of the 2030 Agenda.

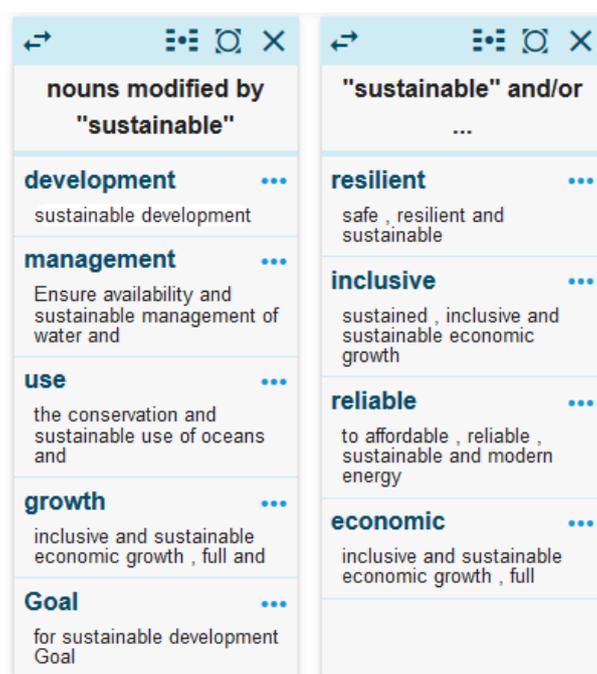
The adjective *sustainable* appears 177 times in the 2030 Agenda (English). It functions as a premodifier of the noun *development* in 108 instances, namely 61% of its occurrences. In these cases, the combination of adjective and noun results in the noun phrase *sustainable development*, whose meaning by collocation is detailed in § 4.4.1.

The extremely frequent co-occurrence of *sustainable* and *development* impacts on the extraction of the collocational patterns of *sustainable*: in fact, with the type of search employed in the present study, some of the collocates of *sustainable* are identified by the software not because of their co-occurring with the sole adjective but because they appear in a statistically significant way together with *sustainable development*. This is taken into account when analysing the collocational patterns of *sustainable*. Thus, in the description of the collocational patterns of *sustainable*, collocates and concordance lines having to do with the lexical item *sustainable development* and not with the sole

lexeme *sustainable* are left aside and treated in detail when analysing the meaning by collocation of *sustainable development* (see § 4.4.1).

#### 4.2.1.1. Word sketch

A glance at the meaning by collocation of the adjective *sustainable* is taken through the word sketch displayed in Figure 5. The word sketch of *sustainable* shows the nouns that the adjective modifies most significantly in the 2030 Agenda (English) and the adjectives that come together with *sustainable* the most.



**Figure 5.** Word sketch of *SUSTAINABLE* in the 2030 Agenda Corpus (English) – LogDice(6.0), NC5-C5.

The nouns meaningfully modified by *sustainable* are *development* ( $f=108$ ), *Goal* ( $f=33$ ), *growth* ( $f=6$ ), *management* ( $f=6$ ), and *use* ( $f=6$ ). These nouns provide the adjective *sustainable* with a semantic preference for material processes expressed through deverbal nouns (*development, growth, management, use*). Material processes report actions and events that are consciously initiated by animate actors and that tend to cause a change in the initial conditions of the events; material processes involve an actor (i.e. an acting being) and a goal (i.e. a receiving being or entity). The collocates *development* and *Goal* refer to the core of the 2030 Agenda, namely sustainable development and the Sustainable Development Goals. *growth, management, and use*

showcase the attitude that the Agenda has towards sustainability. Sustainable development, in fact, can be achieved in terms of *sustainable economic growth*. It can also be pursued with the *sustainable management* of natural elements or human activities that directly involve the environment. It can also engage the *sustainable use* of other natural elements.

The adjectives that appear most meaningfully together with *SUSTAINABLE* are *inclusive* (f=17), *economic* (f=10), *resilient* (f=9), and *reliable* (f=5). They suggest that patterns including *sustainable* might bear a positive connotation as they are frequently combined with the positively connotated adjectives *inclusive*, *reliable*, and *resilient*. *economic* refers to *sustainable economic growth*.

#### 4.2.1.2. Collocation networks

The meaning by collocation sketched for *sustainable* in the 2030 Agenda (English) thanks to the lexeme's word sketch is then improved with the analysis of the collocation networks of the adjective.

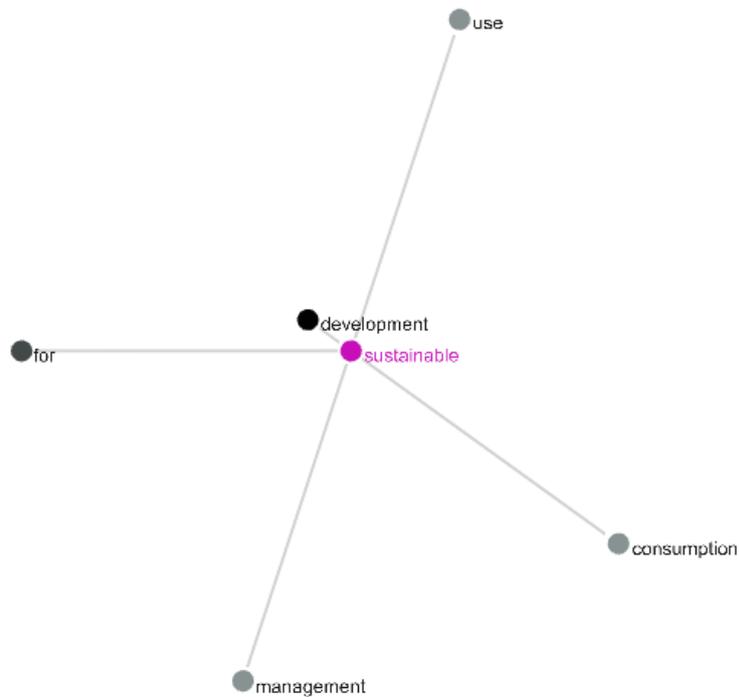
The collocation networks of the adjective *sustainable* with spans of  $\pm 1$ ,  $\pm 3$  and  $\pm 5$  words can be seen in Table 24. Table 24 displays only the lists of collocates with their rank, while Z-value, frequency of co-occurrence of node and collocate and frequency of occurrence of the collocate in the corpus can be read in Tables 46, 47 and 48 in the Appendix.

Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>development</i>	<i>development</i>	<i>development</i>
2	<i>for</i>	<i>goals</i>	<i>goals</i>
3	<i>consumption</i>	<i>sustained</i>	<i>and</i>
4	<i>management</i>	<i>for</i>	<i>for</i>
5	<i>use</i>	<i>and</i>	<i>goal</i>
6		<i>consumption</i>	<i>the</i>
7		<i>inclusive</i>	<i>sustained</i>
8		<i>modern</i>	<i>inclusive</i>
9		<i>resilient</i>	<i>consumption</i>
10		<i>partnership</i>	<i>resilient</i>
11		<i>production</i>	<i>agriculture</i>
12		<i>goal</i>	<i>production</i>
13		<i>growth</i>	<i>promote</i>
14		<i>reliable</i>	<i>of</i>
15		<i>management</i>	<i>to</i>
16		<i>on</i>	<i>patterns</i>

17	<i>energy</i>	<i>modern</i>
18	<i>promote</i>	<i>innovation</i>
19	<i>innovation</i>	<i>management</i>
20	<i>of</i>	<i>partnership</i>
21	<i>use</i>	<i>ensure</i>
22	<i>the</i>	<i>growth</i>
23	<i>affordable</i>	17
24	<i>to</i>	<i>on</i>
25		<i>be</i>
26		<i>reliable</i>
27		<i>are</i>
28		<i>use</i>
29		<i>peace</i>
30		<i>energy</i>
31		<i>we</i>
32		<i>targets</i>
33		<i>that</i>
34		<i>affordable</i>
35		<i>recognize</i>
36		<i>all</i>
37		<i>economic</i>
38		<i>including</i>
39		<i>technology</i>
40		<i>policies</i>
41		<i>is</i>
42		<i>enhance</i>
43		2030

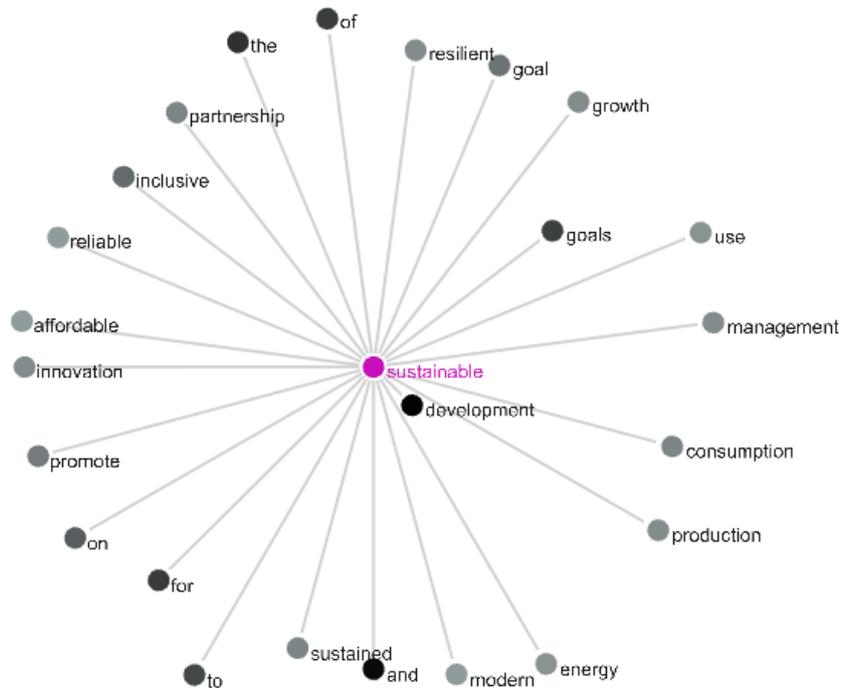
**Table 24.** List of collocates of *SUSTAINABLE* in the 2030 Agenda Corpus (English).

Figure 6 shows the collocation network of *sustainable* as it was created with the help of #LancsBox. As Figure 6 shows, between one word to the left and one word to the right, *sustainable* collocates with the content words *consumption*, *management*, and *use*, that co-occur with *sustainable* on its right-hand side, and with the function word *for*, that precedes the node.



**Figure 6.** Collocation network of *SUSTAINABLE* in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.

Between three words to the left and three words to the right of *sustainable*, the content words co-occurring with the adjective without being also collocates of *sustainable development* are *sustained, inclusive, modern, resilient, production, growth, reliable, energy, promote, innovation, and affordable*, and the function words are *and, on, of, the, and to*. Among the content words, *sustained, inclusive, reliable, promote, and affordable* tend to occur on the left-hand side of the node; *modern, resilient, production, growth, and energy* tend to follow it. Of the function words, *on, of, the, and to* tend to stay on the left-hand side of the node; *and* is found both before and after *sustainable*. The collocation patterns include also the already mentioned content words *consumption, management, use, and* the function word *for*. The collocates of *sustainable* that were found between three words to the left and three words to the right of the node can be seen in Figure 7.



**Figure 7.** Collocation network of *SUSTAINABLE* in the 2030 Agenda Corpus (English) – Z(10.0), 3L-3R, NC5-C5.

Within five words to the left and five words to the right, the adjective *sustainable* collocates with the same content and function words previously mentioned. It also collocates with the content words *agriculture*, *patterns*, *ensure*, *economic*, *recognize*, *policies*, and *enhance*. The collocates *ensure*, *policies*, and *enhance* tend to occur on the left-hand side of the node; the collocates *agriculture*, *patterns*, and *economic* tend to appear on the right-hand side of the node. Within the  $\pm 5$  span, the adjective *sustainable* collocates also with the function words *are*, *that*, *all*, *including*, *is*, and *2030*. The function words *are* and *2030* tend to collocate with *sustainable* on its left-hand side; the words *that*, *all*, *including*, and *is* tend to collocate with the node on its right-hand side. The lexemes that collocate with the adjective *sustainable* between five words to the left and five words to the right of the node can be seen in Figure 8 and the uses of the collocates gathered for *sustainable* with  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word spans in the 2030 Agenda (English) can be read in the following paragraphs.



*sustainable* collocates with *production* in the lexical item *sustainable consumption and production* in most of the cases and the only instance where *sustainable* collocates with *production* without involving *consumption* is in the noun phrase *sustainable food production systems* ( $f=1$ ).

Within one word to the left and one word to the right of the node, *sustainable management* consists in a thoughtful management of *water and sanitation for all* ( $f=2$ ), of *natural resources* ( $f=1$ ), of *our planet's natural resources* ( $f=1$ ), of *fisheries, aquaculture and tourism* ( $f=1$ ), and of *all types of forests* ( $f=1$ ), in order for *social and economic development* to happen. *sustainable management* stimulates a semantic preference for natural elements, which are treated as resources for human activities. The lexical item is associated with the *availability* and the *efficient use* of the same resources and it functions as the goal of material processes. In the 2030 Agenda (English), *sustainable management* co-occurs with material processes of transformation like ENSURE ( $f=2$ ), ACHIEVE ( $f=1$ ) and IMPLEMENT ( $f=1$ ). It seems that when *sustainable* qualifies *management*, i.e. when management is endowed with the property of being sustainable, the related sustainable action or state of being features as the goal of a material process. In other words, the sustainable action or state of being is the entity "to which the process is extended" (Halliday and Matthiessen 2004: 181). The collocational patterns that tie *sustainable* to *management* within three words to the left and three words to the right of the node are the same as those of the  $\pm 1$  span but for a case where *sustainable forest management* substitutes the *sustainable management of all types of forests*. When *sustainable* collocates with *management* within a  $\pm 5$  span, *management* tackles the issue of how urbanisation should evolve in a sustainable way so that natural resources could be used sustainably (in *sustainable urban development and management*,  $f=1$ , and in *sustainable human settlement planning and management*,  $f=1$ ). Therefore, not only do natural elements need to be sustainably managed according to the Agenda, but also cities and other forms of settlements should be sustainably planned.

Within a  $\pm 1$ -word span, *sustainable use* is paired with *conservation* and *restoration* in noun phrases like *the conservation and sustainable use* ( $f=2$ ) and *conservation, restoration and sustainable use* ( $f=1$ ). These noun phrases are modified respectively by the prepositional phrases of *oceans and their resources* ( $f=2$ ) and of *terrestrial and inland freshwater ecosystems* ( $f=1$ ). Natural elements like oceans and land are claimed to be resources due to be used, conserved, and restored sustainably. This is done through sentences where the noun phrases including *sustainable use* are the goals of the material processes *enhance* ( $f=1$ ) and *ensure* ( $f=1$ ) or when they belong to prepositional phrases

postmodifying a noun that plays the role of goal for the material process *provides* ( $f=1$ ). The same idea is encoded in sentences where *sustainable use* is equally employed as the goal of *protect, restore, and promote* ( $f=2$ ) without being associated to any other noun phrase. In this case it is *terrestrial ecosystems* ( $f=2$ ) to be bound to *sustainable use*. Enlarging the collocation window to five words to the left and five words to the right of the node, the adjective *sustainable* collocates with the noun *use* in a single extra sentence if compared to the instances found between one and three words to the left and to the right of the node. The new occurrence is *sustainable management and efficient use of natural resources* ( $f=1$ ). Unlike in the previous cases, *sustainable* refers to *management*, which is associated with *efficient use* by the conjunction *and*. Sustainability and efficiency are deemed to be hands in hands when it comes to the usage of natural resources.

Within a collocation window of three words to the left and three words to the right, *sustainable* collocates also with the nouns *production, growth, energy* and *innovation*.

*energy* can be found in the patterns *sustainable and modern energy* ( $f=3$ ) and *sustainable energy* ( $f=2$ ). The 2030 Agenda pleads for affordable, reliable, modern, and sustainable energy systems to be delivered to all, including developing countries, by expanding infrastructures and upgrading technologies.

The adjective *sustainable* collocates also with *growth*. In 6 out of 7 instances, *growth* is conceived in economic terms as *economic growth*. Economic growth is bound to commitments in the fields of employment (*full and productive employment*,  $f=2$ , and *decent work for all*,  $f=1$ ), of environment (*preserving the planet*,  $f=1$ ), of society (*eradicating poverty in all its forms and dimensions*,  $f=1$ ; *fostering social inclusion*,  $f=1$ ; *combating inequality within and among countries*,  $f=1$ ). This economic growth should be in line with the different needs and possibilities of the countries involved and *respect national policy space* ( $f=1$ ). At the same time, however, it should be aimed at fiercely because of its being *essential for prosperity* ( $f=1$ ). The only co-occurrence of *growth* with *sustainable* that does not refer to economic growth can be found in *inclusive growth and sustainable development* ( $f=1$ ). In this case, *growth* is used in relation to a general and comprehensive improvement of the world's countries that might take into consideration the role of migrants for and in sustainable development.

The adjective *sustainable* collocates with the noun *innovation* mainly through the noun phrase *sustainable development* (7 occurrences out of 10). In the remaining three instances, *innovation* is said to be a condition in need of fostering. The noun is found

in association with the fields of urbanisation and infrastructure. In fact, innovation fostering is written to be a prerequisite for resilient infrastructure to be built and to encourage the promotion of inclusive and sustainable industrialization.

The nouns that *sustainable* additionally collocates with between five words to the left and five words to the right are *agriculture*, *patterns* and *policies*.

One of the practices that can be regarded as *sustainable* is *agriculture*. *sustainable agriculture* is found in the goal position of material processes like PROMOTE ( $f=2$ ), DEVELOP ( $f=1$ ) and INCREASE ( $f=1$ ).

*sustainable* collocates with *policies* mainly through the sequence *sustainable development* (in 5 out of 7 occurrences). When *policies* is not related to *sustainable development*, the noun is either used in the complex prepositional phrase *in accordance with national policies and priorities* ( $f=1$ ) or as the goal of a material process that stimulates a causal-conditional expansion of purpose ( $f=1$ ), where *sustainable* belongs to the noun phrase *sustainable tourism*. Thus, policies are deemed as the necessary tool through which sustainable practices can be implemented.

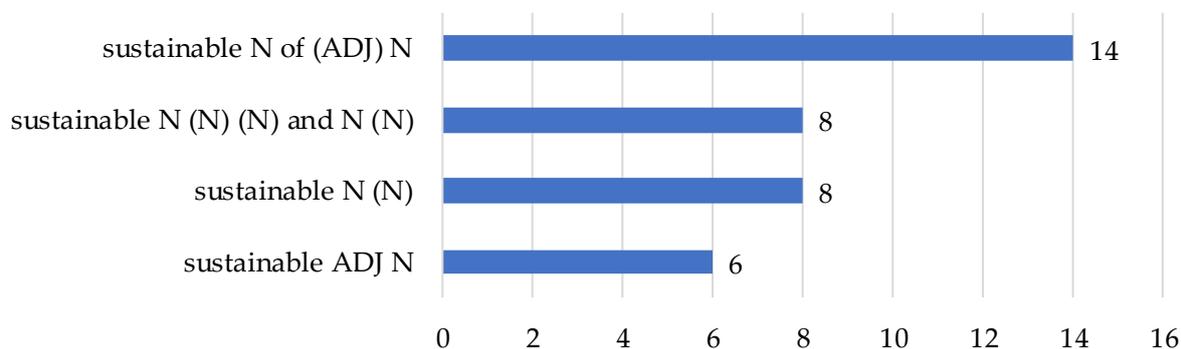
Within the  $\pm 5$ -word collocation window, *sustainable* collocates with the noun *development* in slightly different ways compared to the mainstream *sustainable development*. The 2030 Agenda, in fact, claims that sectors like agriculture, cities, fisheries, industries, infrastructures, and pastoralism should be sustainably developed through patterns like *sustainable urban development* ( $f=2$ ), *sustainable agriculture, pastoralist and fisheries development* ( $f=1$ ), *sustainable industrial development* ( $f=1$ ), *sustainable and resilient infrastructure development* ( $f=1$ ). In these patterns, the noun *development* operates as the head of a noun phrase that is modified by another noun phrase made of the adjective *sustainable* followed by one or more nouns.

When *sustainable* collocates with nouns, it modifies noun phrases in the following colligational patterns:

<i>sustainable</i>	N						
<i>sustainable</i>	N	(N)	(N)	<i>and</i>	N	(N)	
<i>sustainable</i>	ADJ	N					
<i>sustainable</i>	N	N					
<i>sustainable</i>	N	<i>of</i>	(ADJ)		N		

In these colligational patterns, N stands for “noun” and ADJ stands for “adjective”.

These colligational patterns are displayed also in Graph 6 together with their frequency.



**Graph 6.** The colligational patterns of *SUSTAINABLE* and nouns in the 2030 Agenda (English).

As it can be observed in Graph 6, the most frequent colligational pattern involving *sustainable* and nouns witnesses *sustainable* modifying a noun phrase that includes a noun and a prepositional phrase introduced by *of*. The second and third most frequent colligations witness *sustainable* modifying a noun phrase where two nouns or two compounds are kept together by the conjunction *and*. In the least frequent colligational pattern *sustainable* modifies a noun phrase made of an adjective and a noun.

These colligations are filled with nouns that cluster around the fields of change by depletion (*consumption, production, management, use*), of general change (*development, innovation, growth*) and of human activities (*agriculture*). These categories contribute to the adjective *sustainable* having a semantic preference for lexemes that refer to change caused by humans.

**Adjectives.** *sustainable* collocates with adjectives starting from the  $\pm 3$ -word span. The adjectives that *sustainable* collocates with within this span are *sustained, inclusive, modern, resilient, reliable, and affordable*.

*sustainable* collocates with the adjectives *affordable* and *reliable* in the adjective phrases *affordable, reliable, sustainable and modern* ( $f=3$ ) and *affordable, reliable and sustainable* ( $f=1$ ). They all refer to *energy* ( $f=3$ ) or *energy services* ( $f=1$ ). Moreover, *reliable* also co-occurs with *quality* and *resilient* in *quality, reliable, sustainable and resilient infrastructure* ( $f=1$ ), whereas *affordable* also appears with *safe* and *accessible* in *safe,*

*affordable, accessible and sustainable transport systems* (f=1). Thus, engagement towards sustainability tries to keep together trustworthiness and economic feasibility.

The general collocational tendencies of the adjective *inclusive* have already been described in § 4.1.2.1. When *inclusive* co-occurs with *sustainable*, the two adjectives appear in the pattern *sustained, inclusive and sustainable* (f=6; with its variant *sustainable, inclusive and sustained*, f=1). *sustainable* is also found solely with *inclusive* in the adjective phrase *inclusive and sustainable* (f=5). *sustained, inclusive and sustainable* always modifies *economic growth* while *inclusive and sustainable* modifies *industrialization* (f=3) and *urbanization* (f=1). Sustainable industrialization and urbanization are required to reach a wider population, both in developed and in developing countries. Inclusion needs also to reach human migrants in the form of an *inclusive growth and sustainable development* (f=1) that could assist them in safely settling in a new country and in the involved countries to manage the movement of people in a *coherent and comprehensive* way. In addition, inclusion should also characterize *societies* (f=2) that need to be *peaceful and inclusive* so that sustainable development can be achieved. Widening the collocation window to five words to the left and five words to the right of *sustainable*, *sustainable* and *inclusive* collocate also in the sequence *inclusive, safe, resilient and sustainable* (f=2), that modifies *cities and human settlements*. Cities and human settlements are thus asked to be welcoming towards inner and outer members, to safeguard them from an infrastructural point of view, being both safe and resilient, and to grow in a sustainable manner.

*sustainable* is combined with the adjective *modern* in modifying the lexical items *energy services* (f=3) and *energy* (f=2). This combination stresses the association between sustainable practices and innovation. It also highlights how sustainability cannot be achieved without the development of cutting-edge systems and technologies.

Being *sustainable* requires or prompts to be *resilient*. This combination of qualities is especially referred to infrastructures and settlements. The 2030 Agenda is committed to *shift the world on to a sustainable and resilient path* (f=1) and it wishes that *human settlements* (f=2) or *human habitats* (f=1) could be (*inclusive*), *safe, resilient and sustainable* (f=3). Consequently, resilience should characterise human habitats and settlements, but it should also concern *infrastructure* (f=1), *infrastructure development* (f=1) and, more specifically, *buildings* (f=1), especially when they are made *utilizing local materials*. In addition to being *resilient*, infrastructure should also be *quality* and *reliable*.

Among the adjectives collocating with *sustainable*, *sustained* belongs to the adjective phrase *sustained, inclusive and sustainable* (f=6) or to its variant *sustainable*,

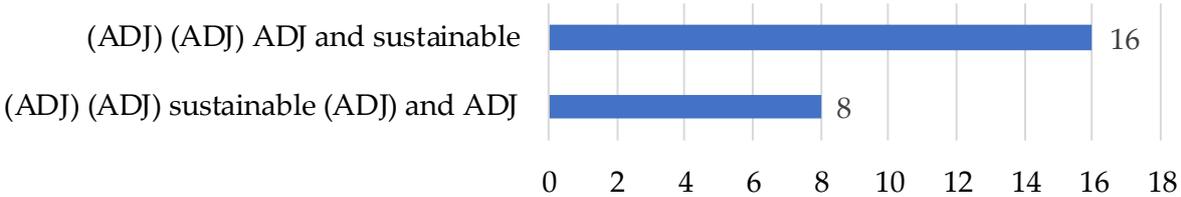
*inclusive and sustained* (f=1). This adjective phrase always modifies the lexical item *economic growth* and it offers an idea of the kind of economic development that the 2030 Agenda wishes could be realized in the path towards sustainability.

In addition to *sustained, inclusive, resilient, modern, reliable, and affordable*, the only adjective that *sustainable* collocates with within a ±5 span and not within ±3 and ±1 spans is *economic*. In most cases, *sustainable* collocates with the adjective *economic* when *economic* modifies the noun *growth*. The collocation *sustainable-economic growth* is found in the sequence *sustained, inclusive and sustainable economic growth* in more than half of its occurrences (6 out of 10), but it can also appear in the alternative sequence *sustainable, inclusive and sustained economic growth* (f=1). In two instances, *sustainable* co-occurs with *economic* with a looser discursive connection.

When *sustainable* collocates with adjectives, it contributes to shaping adjective phrases that reflect two main colligational patterns:

(ADJ) (ADJ) ADJ and sustainable  
 (ADJ) (ADJ) sustainable (ADJ) and ADJ

These colligational patterns are reproduced also in Graph 7 together with their frequency.



**Graph 7.** The colligational patterns of *SUSTAINABLE* and adjectives in the 2030 Agenda (English).

As it can be noted in Graph 7, the most frequent colligational pattern involving *sustainable* and adjectives witnesses *sustainable* at the end of an adjective phrase made of one or more adjectives, which are followed by the conjunction *and* and then by *sustainable*. On the other hand, the conjunction *and* follows *sustainable* in the second colligational pattern, where one or two optional adjectives can precede *sustainable*, which might be followed by an adjective, by the conjunction *and* and by another adjective.

These colligational patterns are occupied by adjectives that reproduce qualities of trustworthiness and inclusiveness (*inclusive, modern, reliable, affordable*) but also of long duration and adaptability (*sustained, resilient*). These adjectives suggest that *sustainable* has a semantic preference for lexemes of enduring inclusion, reliability and adaptability.

**Verbs.** Within a collocation window of three words to the left and three words to the right, *sustainable* collocates only with the verb *promote*. Most of the occurrences of *promote* in this context are in the imperative mood (e.g. *Promote inclusive and sustainable industrialization*; 7 occurrences out of 10). These imperative-mood occurrences of *promote* are used for the description of the seventeen goals of the Agenda and they encourage people to commit to the SDGs. The verb *promote* is used in the indicative mood in a single instance (i.e. *We are also determined to promote sustainable tourism*). In this sentence, the actor of the material process of promoting is *We*, that stands for the community of those who decide to work towards sustainability, and the patient of the process is *sustainable tourism* ( $f=2$ ). Other patients of the material process represented by the collocate *promote* are *sustainable industrialization* ( $f=3$ ), *sustainable agriculture* ( $f=2$ ), *sustainable use of terrestrial ecosystems* ( $f=2$ ), and *sustainable development* ( $f=1$ ), that encompasses all the others as a superordinate. When *promote* is related to the *use of terrestrial ecosystems*, it is also associated with the verbs *protect* and *restore*. The protection and the restoration of terrestrial ecosystems feature as the starting point of the process of promoting the same, as it is shown by the order with which the verbs are used. Within five words to the left and five words to the right, *sustainable* collocates with the verb *promote* in the cases that have already been discussed in relation to the  $\pm 3$  span, but it also holds *sustained, inclusive and sustainable economic growth* ( $f=1$ ) and *the implementation of sustainable management of all types of forests* ( $f=1$ ). In both cases the verb is in the imperative mood and its subject needs to be guessed among those unclear *we* that populate the document. The material process *promote* conveys an idea of supporting without certainty and what is promoted here is a form of economic growth and management of natural resources like forests that should be carried out with sustainability in mind and that functions as the goal of the process.

Apart from the already discussed *promote*, the verbs that *sustainable* collocates with within the  $\pm 5$  span are *ensure, recognize* and *enhance*.

As far as the collocational patterns of *sustainable* with *enhance* are concerned, only two occurrences of the *sustainable-enhance* collocation are independent of the

phrase *sustainable development*. In these two occurrences, the verb *enhance* is another example of material process that involves the adjective *sustainable* in the goal function. In both sentences the verbs are in the imperative mood and therefore the actor is apparently absent. *sustainable* belongs to the goal and it modifies *urbanization* in the first case and *use of oceans* in the second. What needs to be enhanced according to the Agenda, in fact, are human settlements and a responsible use of natural elements.

When *sustainable* co-occurs with the verb *ensure*, the verb performs a commitment to the success of the actions taken towards sustainability. *ensure* is always used in the imperative mood and it is employed to describe the agenda's goals. The adjective *sustainable* always features within the goal of the material process of ensuring. The agenda ensures thus: *access to affordable, reliable, sustainable and modern energy for all (f=2)*; *sustainable consumption and production patterns (f=2)*; *availability and sustainable management of water and sanitation for all (f=1)*; *sustainable food production systems (f=1)*; *sustainable withdrawals and supply of freshwater (f=1)*; *the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services (f=1)*.

In most cases, *sustainable* co-occurs with *recognize* when the verb is used in another sentence compared to the one that hosts the adjective (4 cases out of 6). In addition, one of the two remaining occurrences is not significant at this stage of analysis because it associates *recognize* to *sustainable* when the adjective modifies *development* in the noun phrase *sustainable development*. In the remaining sentence the mental process *recognize* includes a clause-like phenomenon where *sustainable* is associated with *urban development and management*. Mental processes reflect the perception that a sentient being has of the outer world; they involve a senser (i.e. the perceiving being) and a phenomenon (i.e. the perceived entity).

Within the  $\pm 5$ -word span, the adjective *sustainable* collocates also with two verb forms that function both as processes and as auxiliaries (i.e. *are* and *is*).

*sustainable* collocates with the verb form *are* independently of *development* in relational processes. Relational processes are employed to identify and classify; when they identify, they assign an identifier to an identified, whereas when they classify, they bestow an attribute on a carrier. In most cases, *are* bears an attributive function that associates the property of being sustainable to a wide range of carriers. In *human habitats are safe, resilient and sustainable (f=1)*, human settlements are said to be secure, able to adapt to changes and to develop sustainably. In *consumption and production patterns and use of all natural resources [...] are sustainable (f=1)*, sustainability is written

to characterise trends in manufacturing and use of resources. In *Promote public procurement practices that are sustainable* (f=1), practices of procurement are attributed the characteristic of being sustainable. In *sustainable urban development and management are crucial to the quality of life and or our people* (f=1), the sustainable development and governance of cities are claimed to be essential for wellbeing. In a single case, however, the verb form *are* does not associate to practices that are written to be sustainable, but it sticks to the *we* carrier (i.e. *We are also determined to promote sustainable tourism*). This carrier stands for the United Nations and the countries and the citizens involved for sustainability. Through the verb *are*, the subject *we* is described as being resolved to back sustainable actions like *sustainable tourism*.

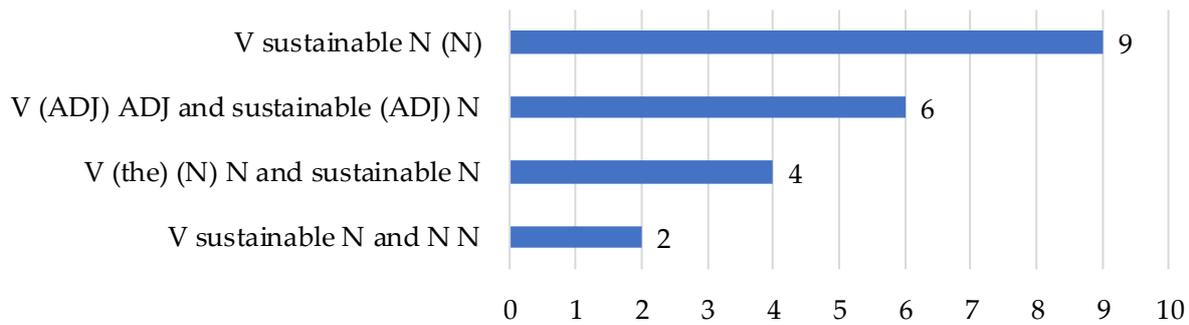
When *sustainable* collocates with the verb form *is*, which is mainly used in relational processes, only two occurrences involve the sole adjective *sustainable* without it modifying *development* and with the node and the verb form belonging to the same sentence. In the first sentence *sustainable economic growth* is qualified as paramount for *prosperity* through the attributive function of *is*; thus, sustainability is strongly associated with economic growth and wealth. In the second case, the identity function of *is* posits that keeping sustainable levels of debt corresponds to an obligation for *borrowing countries*.

The adjective *sustainable* and the verbs are mainly found in the following colligational patterns:

				V	<i>sustainable</i>	N	(N)		
	V	(ADJ)	ADJ	<i>and</i>	<i>sustainable</i>	(ADJ)	N		
	V	( <i>the</i> )	(N)	N	<i>and</i>	<i>sustainable</i>	N		
				V	<i>sustainable</i>	N	<i>and</i>	N	N

In these colligational patterns, v stands for “verb”.

These colligational patterns can be viewed also in Graph 8, which additionally reports the frequency of occurrence of these colligations.



**Graph 8.** The colligational patterns of *SUSTAINABLE* and verbs in the 2030 Agenda (English).

As it can be seen in Graph 8, the most frequent colligational pattern binding *sustainable* and verbs features a verb followed by a noun phrase which is modified by *sustainable*. In the second most frequent colligational pattern the verb is followed by a noun phrase including an adjective phrase with *sustainable*. In the third and in the fourth colligations the verb is followed by noun phrases including *sustainable*.

The slots available in these colligational patterns are enriched with verbs that represent material or mental processes of improving and supporting (*promote, ensure, recognize, enhance*). The verbs contribute to *sustainable*'s semantic preference for material processes of commitment to sustainability.

**Prepositions.** Among the function words that *sustainable* collocates with, within one word to the left and one word to the right, the adjective co-occurs with the preposition *for* in prepositional phrases followed by *sustainable development* but for two instances (26 occurrences out of 28). In the first case *for* is the head of a prepositional phrase that functions as a postmodifier of the noun *conditions*. What is more interesting about the prepositional phrase is its nominal content. The adjective *sustainable* is associated with the adjectives *inclusive* and *sustained* as a modifier of *economic growth, shared prosperity and decent work for all*. Economic growth, prosperity and work are demanded to be sustainable and sustained but at the same time they are asked to involve all the protagonists summoned by the Agenda. In the second case, *for* is the head of a prepositional phrase where the nominal component is made of *sustainable* and *tourism*. The prepositional phrase functions as a complement that encodes the recipient of the material process of developing and implementing tools. This recipient is *sustainable tourism*, which is said to contribute to employment and to the promotion of local traditions. Most occurrences of *sustainable* and *for* within the  $\pm 3$ -word span are the

same as those appearing within the  $\pm 1$ -word span. Among the new entries, most collocational patterns keeping together *sustainable* and *for* include *sustainable development* and not simply the sole adjective (7 out of 10). In the remaining three occurrences, *for* is the head of a prepositional phrase that modifies a noun phrase including *sustainable* (i.e. *sustainable development impacts for sustainable tourism*,  $f=1$ , *sustainable energy services for all*,  $f=1$ , and *sustainable transport systems for all*,  $f=1$ ). The *for all* prepositional phrase is particularly interesting as it suggests that the beneficiaries of *modern and sustainable energy services* and of *safe, affordable, accessible and sustainable transport systems* should be all peoples in the world. This stresses further the commitment for inclusivity that is expressed in the Agenda also through the collocational pair *sustainable-inclusive*. Within five words to the left and five words to the right, when the adjective *sustainable* collocates with the preposition *for* and it does not appear together with *development*, *for* is the head of prepositional phrases that can either include the adjective *sustainable* (7 occurrences overall, as in *for participatory, integrated and sustainable human settlement planning and management*,  $f=1$ , *for supplying modern and sustainable energy services*,  $f=1$ , and *for the conservation and sustainable use of oceans and their resources*,  $f=1$ ) or that witness *sustainable's* modifying the noun phrase that precedes the *for* phrase (9 occurrences overall; e.g. *affordable, reliable, sustainable and modern energy for all*). In the second case, the most frequent prepositional phrase introduced by *for* is *for all* ( $f=5$ ), which stresses that the recipient of sustainable practices should be the whole world population.

Within three words to the left and three words to the right, the adjective *sustainable* co-occurs with the prepositions *on*, *of* and *to*, in addition to the aforementioned *for*. When the preposition *on* co-occurs with the adjective *sustainable* and *sustainable* modifies nouns different from *development*, *sustainable* and *on* usually collocate in a prepositional phrase having *on* as head and including the adjective in the modifying noun phrase, as in *on Sustainable Consumption and Production* ( $f=2$ ), *on Sustainable Consumption and Production Patterns* ( $f=1$ ), *on Housing and Sustainable Urban Development* ( $f=1$ ), or *on the sustainable management* ( $f=1$ ). The prepositional phrases can be used either as a complement or as a postmodifier. In both cases they function as the matter of a discussion.

Most of the co-occurrences of *sustainable* with the preposition *of* are bound to the lexical item *sustainable development*. For the discussion of the collocational patterns that keep together *sustainable development* and *of* see § 4.4.1.2. When *sustainable* collocates with *of* without the adjective modifying *development*, *of* usually follows

*sustainable*. The noun phrases including *sustainable* and modified by an *of* prepositional phrase are *sustainable use* (f=6), *sustainable management* (f=5), and *sustainable patterns* (f=2). The prepositional phrase with *of* postmodifies the noun phrases by specifying their extent and it mainly assists in specifying the characteristics of what is labelled *sustainable*, namely *management*, *patterns*, and *use*. In other words, it clarifies what should be sustainably managed, organized, and used. Additionally, with a  $\pm 5$ -word span, the only cases where *sustainable* collocates with the preposition *of* in contexts dissimilar to those that have already been described for the  $\pm 3$  collocation window are those in which the preposition is the head of prepositional phrases that modify a noun phrase which includes the adjective *sustainable*. These prepositional phrases serve as a specification of the meaning of the preceding noun phrase. For example, *of natural resources* (f=1) modifies *sustainable management and efficient use* by detailing the patient of the act of managing and using; *of freshwater* (f=1) modifies *sustainable withdrawals and supply* by specifying that freshwater should be sustainably withdrawn and supplied to *reduce the number of people suffering from water scarcity*.

With a  $\pm 3$ -word span, the function word *to* is used either as a preposition or as the particle introducing infinitive verb forms. When *to* is not bound to the lexical item *sustainable development* and it is used as a preposition, it postmodifies a noun or it functions as a complement that conveys an idea of movement. When *to* is used as a particle and it flags infinitives, it introduces the verbs *to promote* (*sustainable tourism*) (f=2), *to adopt* (*sustainable practices*) (f=1), *to build* (*dynamic, sustainable, innovative and people-centred economies*) (f=1), *to finance* (*sustainable forest management*) (f=1), *to make* (*them sustainable*) (f=1), and *to pursue* (*sustainable livelihood opportunities*) (f=1). All these verbs represent material processes whose goals include some aspects of sustainability. The material processes are enforced by introductory verbs that stress the commitment to those processes. When *sustainable* and the function word *to* collocate within the  $\pm 5$ -word span, the function word is employed once again either as a preposition or as the introductory particle for infinitive verbs. When it encloses a prepositional phrase, it can either encompass the adjective *sustainable* or not. When *to* is used as the introductory particle for infinitive verbs, it is tied to the verb forms *to move* (*towards more sustainable patterns of consumption and production*) (f=2) and *to create* (*conditions for sustainable, inclusive and sustained economic growth*) (f=1). In the first clause, the material process is a transformative process of motion that turns an initial condition where patterns of consumption and production are unsustainable to one where they are

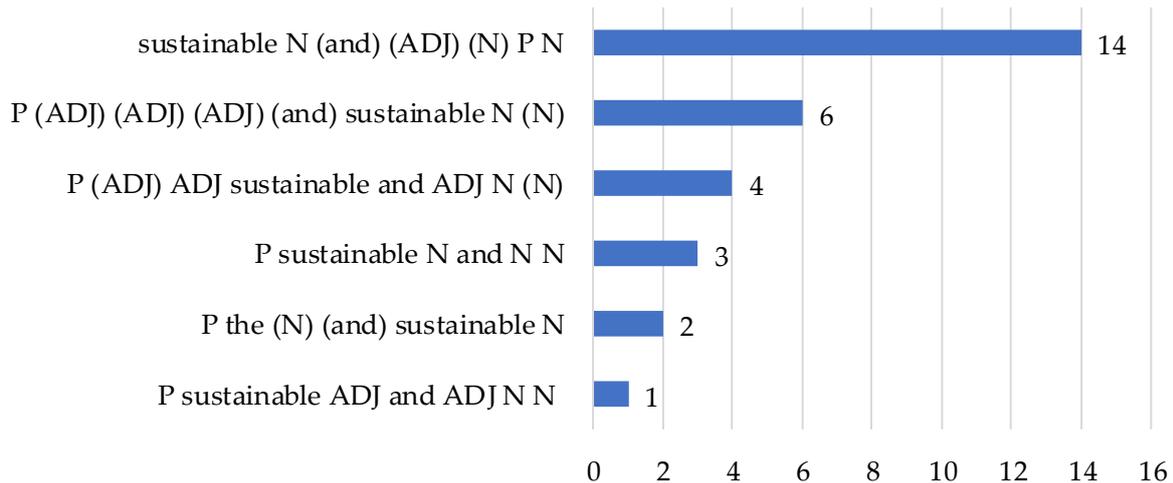
sustainable. The material process of the second clause is a general creative process that describes the generation of the right conditions for sustainable economic growth.

When *sustainable* collocates with prepositions, it mainly triggers the following colligational patterns:

P	<i>the</i>	(N)	( <i>and</i> )	<i>sustainable</i>	N				
			P	<i>sustainable</i>	N	<i>and</i>	N	N	
P	(ADJ)	(ADJ)	(ADJ)	( <i>and</i> )	<i>sustainable</i>	N	(N)		
	P	(ADJ)	ADJ	<i>sustainable</i>	<i>and</i>	ADJ	N	(N)	
			P	<i>sustainable</i>	ADJ	<i>and</i>	ADJ	N	N
				<i>sustainable</i>	N	( <i>and</i> )	(ADJ)	(N)	P N

In these colligational patterns, P stands for “preposition”.

These colligational patterns are reported with their frequency in Graph 9.



**Graph 9.** The colligational patterns of *SUSTAINABLE* and prepositions in the 2030 Agenda (English).

As it can be seen in Graph 9, in the most frequent colligational pattern keeping together *sustainable* and its collocating prepositions, the preposition is the head of a prepositional phrase that modifies a noun phrase including *sustainable*. In the other colligational patterns, the preposition is the head of a noun phrase where the adjective *sustainable* variably modifies a noun phrase.

**Conjunctions.** Within three words to the left and three words to the right of the node, *sustainable* collocates with a single conjunction (i.e. *and*). The co-occurrence of *sustainable* with the conjunction *and* is particularly insightful in determining the meaning of the adjective. *and* keeps together several noun phrases that feature the adjective and several verb phrases where the adjective appears in the object position. More interestingly, *and* associates *sustainable* to adjectives that contribute to its meaning. These adjectives are *accessible, affordable, dynamic, inclusive, innovative, modern, people-centred, reliable, resilient, safe, and sustained*. They participate in stressing that what is *sustainable* is up to date (*innovative* and *modern*) but at the same time energetic and open towards a possible change (*dynamic*); any change, at the same time, might make things spring back (*resilient*) and endure a long period of stability (*sustained*) in order to make what is *sustainable* also available for all people (*accessible, affordable, inclusive, and people-centred*) and secure (*reliable* and *safe*). Among these adjectives, *affordable, inclusive, modern, reliable, resilient, and sustained* do also collocate with *sustainable* within the  $\pm 3$  collocation window.

When the collocation window is widened to five words to the left and five words to the right of *sustainable*, the adjective happens to collocate also with the function word *that*. In most cases (8 out of 13), *sustainable* collocates with the function word *that* in contexts where it modifies the noun *development*. When *sustainable* does not belong to the noun phrase *sustainable development*, the function word *that* is either used as a pronoun that introduces a relative clause ( $f=6$ ) or as a conjunction that introduces a secondary clause ( $f=1$ ). When it is used as a conjunction, it encloses a secondary clause that encompasses also the adjective *sustainable* and the conjunction *that* introduces a sentence that functions as the phenomenon of the mental process *recognize*. The senser of the process is, also in this case, a general *we* that admits that a form of sustainable practice (i.e. *sustainable urban development*) is fundamental. When *that* is used as a relative pronoun, it can either modify a noun phrase that features the adjective *sustainable* (i.e. *sustainable tourism that creates jobs and promotes local culture and products, f=2*) or it can include *sustainable* (as in *public procurement practices that are sustainable, f=1*, and *levels that can produce maximum sustainable yield, f=1*).

**Determiners.** Within a  $\pm 3$ -word span, *sustainable* collocates with the determiner *the*. The determiner *the* is almost always found when the adjective *sustainable* modifies the noun *development*. In the other cases it aids in rendering the definiteness of concepts

such as *the conservation and sustainable use of oceans and their resources (f=2)*, *the implementation of sustainable management of all types of forests (f=1)*, *the sustainable management and efficient use of natural resources (f=1)*, *the sustainable management of our planet's natural resources (f=1)*, and *the sustainable use of marine resources (f=1)*.

**Other function words.** Within five words to the left and five words to the right, the adjective *sustainable* collocates with the numeral *2030* in the prepositional phrase *by 2030 (f=6)* and in the pattern *2030 Agenda for Sustainable Development (f=4)*. While the function of the collocation *sustainable-2030* in the second pattern is straightforward, the prepositional phrase co-occurring with *sustainable* operates as a circumstantial that places the UN's engagements towards sustainability in a specific time span and that highlights that the contents included in the agenda should become valid by 2030. The prepositional phrases, in fact, establish the deadline for the SDGs to be implemented. The same holds true for other sparse occurrences: *before 2030 (f=1)*, *between now and 2030 (f=1)*, *in 2030 (f=1)*, *through 2030 (f=1)*, *to 2030 (f=1)*. In all cases *2030* is signalled as the arriving point of a path that the United Nations stepped in in 2015 and that is grammatically marked by some of the prepositions that introduce the numeral (namely *before*, *by*, and *to*).

Within the  $\pm 5$ -word span, *sustainable* co-occurs with the indefinite pronoun *all*. When *sustainable* co-occurs with *all*, the function word is most significantly used in the prepositional phrase *for all* (5 occurrences out of 17 collocates), which makes *all* a beneficiary of sustainable practices. In the other cases, the word does not seem to contribute to shape the meaning of the adjective *sustainable*.

*sustainable* co-occurs also with the function word *including*. In most cases (10 out of 14), *sustainable* collocates with *including* when the function word modifies the noun *development*. In the four instances where *sustainable* and *including* co-occur without *sustainable* being part of the noun phrase *sustainable development*, *including* specifies the meaning of the adjective by mentioning superordinate or subordinate elements that could be included in a group of sustainable elements. The sustainable depletion of marine resources, for example, includes *sustainable management of fisheries, aquaculture and tourism (f=2)* while the protection of the planet from deterioration includes *sustainable consumption and production (f=1)*. At the same time, the development of *quality, reliable, sustainable and resilient infrastructure* includes *regional and transborder infrastructure (f=1)*.

Summing up, the collocational patterns identified for *sustainable* construct for the adjective a broad semantic profile. The property of being *sustainable* is associated with other qualities that contribute to inclusiveness and trustworthiness, as well as to feasibility. These properties qualify activities and states that especially regard the 2030 Agenda but also economic and social development. Economic and social development are said to be *sustainable* whenever they are the goals of material or mental processes of improvement and promotion and if these processes of improvement and promotion are achieved through cooperation.

The semantic profile of *sustainable* lies first in its most significant colligational patterns. *sustainable* plays an important role in adjective phrases where it is associated with other adjectives either with commas or with the conjunction *and*. These adjective phrases modify noun phrases. Noun phrases can be either used as arguments of verbs (typically goals of material processes or phenomena of mental processes) or they can function as modifiers of prepositional phrases.

These colligates are filled with collocates that can be organised into a semantic preference for lexemes related to the 2030 Agenda (*development, goals, goal, targets*), to qualities of inclusiveness and trustworthiness (*sustained, inclusive, resilient, modern, reliable, affordable*), to material nominalised processes that imply change (*development, innovation, growth*) through depletion (*consumption, production, management, use*), to other material or mental processes of improving and supporting (*promote, ensure, recognize, enhance*) that can pursue human activities and products (*agriculture, economic, technology, policies*).

The evaluative prosody associated with these collocational patterns can be positively judged in ecolinguistic terms. This positivity is due both to the overwhelming presence of positively connotated lexemes but also to the grammatical patterns that have been detailed during the analysis. For instance, the evaluative prosody of the adjective *sustainable* gathers an aura of positivity thanks to the fact that when the adjective happens to be included in the goal participant of a material process, the process is boosted by semantic or syntactic features of positive certainty. Sustainable practices seem thus to be strongly encouraged. In addition, the evaluative prosody of these patterns is positive because it depicts an inclusive attitude towards sustainable practices, and this is in line with the ecological framework adopted for the current research.

## 4.2.2. The Hungarian *FENNTARTHATÓ*

The study of the meaning by collocation of the adjective *sustainable* in the 2030 Agenda Corpus resumes with the analysis of the collocational patterns of the Hungarian translational equivalent *FENNTARTHATÓ* ‘sustainable’.

The adjective *FENNTARTHATÓ* ‘sustainable’ appears 187 times in the Hungarian version of the 2030 Agenda and it is followed by the noun *FEJLŐDÉS* ‘development’ 104 times. Thus, 56% of the occurrences of the adjective belong to the noun phrase *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’. As in the English case, the high frequency of this noun phrase influences the outcome of the extraction of the collocational patterns of *FENNTARTHATÓ* ‘sustainable’. This is considered when analysing the collocational patterns of the Hungarian adjective in search for its meaning by collocation.

### 4.2.2.1. Word sketch

The meaning by collocation of *FENNTARTHATÓ* ‘sustainable’ was first delineated thanks to the reading of the word sketch reproduced in Figure 9.

nouns modified by "fenntartható"	"fenntartható" and/or ...	usage patterns
<b>fejlődés</b> ... a fenntartható fejlődés	<b>befogadó</b> ... a tartós , befogadó és fenntartható gazdasági növekedés	<b>in nominative</b> ...
<b>cél</b> ... a Fenntartható Fejlesztési Célok	<b>ellenállóképes</b> ... biztonságossá , ellenállóképessé és fenntarthatóvá tétele	
<b>elérés</b> ... Fenntartható Fejlesztési Célok elérését támogató		
<b>érdek</b> ... a fenntartható fejlődés érdekében		
<b>használat</b> ... megőrzése és fenntartható használata		
<b>támogatás</b> ... fenntartható mezőgazdaság támogatása		
<b>megvalósítás</b> ... a fenntartható fejlődés megvalósítása		
<b>támogató</b> ... a Fenntartható Fejlesztési Célok elérését támogató , tudományos		
<b>növekedés</b> ... befogadó és fenntartható gazdasági növekedés , teljes és		
<b>előmozdítás</b> ... a fenntartható fejlődés előmozdításához		
<b>elősegítés</b> ... fenntartható turizmus elősegítése		
<b>kihívás</b> ... fenntartható fejlődés hatalmas kihívásai		

**Figure 9.** Word sketch of *FENNTARHATÓ* 'sustainable' in the 2030 Agenda Corpus (Hungarian) – LogDice(6.0), NC5-C5.

As it can be noted in Figure 9, the word sketch extracted for the Hungarian *FENNTARHATÓ* 'sustainable' resembles the one retrieved for the English *sustainable* (Figure 5).

The Hungarian *FENNTARHATÓ* 'sustainable' most significantly modifies the nouns *FEJLŐDÉS* 'development', *CÉL* 'goal', *ELÉRÉS* 'achieve', *ÉRDEK* 'interest', *HASZNÁLAT* 'use', *TÁMOGATÁS* 'support', *MEGVALÓSÍTÁS* 'implementation; achievement', *TAMOGATÓ* 'enabling', *NÖVEKEDÉS* 'growth', *ELŐMOZDÍTÁS* 'promote', *ELŐSEGÍTÉS* 'promote; facilitate', and *KIHÍVÁS* 'challenge'.

Among these, the only nouns that play a role in determining the meaning of *FENNTARHATÓ* 'sustainable' independently of the expression *FENNTARHATÓ FEJLŐDÉS*

'sustainable development' are *ELŐSEGÍTÉS* 'promote; facilitate', *HASZNÁLAT* 'use', *NÖVEKEDÉS* 'growth', and *TÁMOGATÁS* 'support' (for the word sketch of *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' see Figure 23).

As in the English case, these nouns suggest that sustainability should be associated with material processes like *NÖVEKEDÉS* 'growth' ( $f=13$ ) or *HASZNÁLAT* 'use' and that these and other processes and practices should be supported and promoted (*ELŐSEGÍTÉS* 'promote; facilitate' and *TÁMOGATÁS* 'support').

The word sketch adds that the adjective is usually found in the nominative case, namely in the form *fenntartható* 'sustainable', and that it is combined with the adjectives *BEFOGADÓ* 'inclusive' and *ELLENÁLLOKÉPES* 'resilient', like in the collocational tendencies of the English *sustainable*. As in the case of English, inclusivity and sustainability are asked to characterise enterprises like economic growth, industrialisation, and urbanisation so that they might encompass the needs of all societies and human beings. Furthermore, sustainability and resilience go in pair when describing architecture and the quality of cities.

#### 4.2.2.2. Collocation networks

The study of the meaning by collocation of the Hungarian *FENNTARTHATÓ* 'sustainable' is then enriched with the analysis of the adjective's collocation networks.

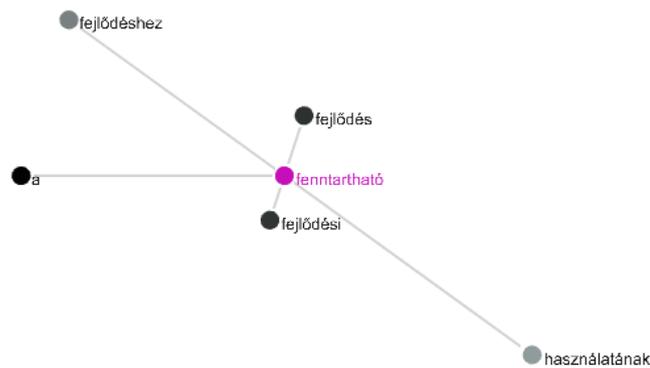
The collocational patterns retrieved for the adjective *FENNTARTHATÓ* 'sustainable' with  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word spans in the 2030 Agenda (Hungarian) can be seen in Table 25. Tables 49, 50, 51 in the Appendix present the same collocates' lists endowed with collocate's position in relation to the node, collocation's Z-value, frequency of co-occurrence of node and collocate and frequency of occurrence of the collocate in the corpus.

Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>fejlődési</i> 'development'	<i>fejlődési</i> 'development'	<i>fejlődés</i> 'development'
2	<i>fejlődés</i> 'development'	<i>fejlődés</i> 'development'	<i>fejlődési</i> 'development'
3	<i>a</i> 'the'	<i>a</i> 'the'	<i>a</i> 'the'
4	<i>fejlődéshez</i> 'for development'	<i>célok</i> 'goals'	<i>és</i> 'and'
5	<i>használatának</i> 'use'	<i>elérését</i> 'achieve'	<i>célok</i> 'goals'
6		<i>és</i> 'and'	<i>erdőgazdálkodás</i> 'forest management'
7		<i>használatának</i> 'use'	<i>érdekében</i> 'for'

8	<i>fejlődéshez</i> ‘for development’	<i>tartós</i> ‘sustained’
9	<i>támogatása</i> ‘support’	<i>elérését</i> ‘achieve’
10	<i>fogyasztás</i> ‘consumption’	<i>nélkülözhetetlen</i> ‘essential’
11	<i>termelés</i> ‘production’	<i>termelés</i> ‘production’
12	<i>befogadó</i> ‘inclusive’	<i>használatának</i> ‘use’
13	<i>használata</i> ‘use’	<i>fejlődéshez</i> ‘for development’
14	<i>tartós</i> ‘sustained’	<i>befogadó</i> ‘inclusive’
15	<i>érdekében</i> ‘for’	<i>támogatása</i> ‘support’
16	<i>megbízható</i> ‘reliable’	<i>megőrzése</i> ‘conserve’
17	<i>növekedés</i> ‘growth’	<i>fogyasztás</i> ‘consumption’
18	<i>foglalkozó</i> ‘-’	<i>fogyasztási</i> ‘consumption’
19	<i>megfizethető</i> ‘affordable’	<i>módok</i> ‘patterns’
20		<i>támogató</i> ‘enabling’
21		<i>cél</i> ‘goal’
22		<i>nélkül</i> ‘without’
23		<i>használata</i> ‘use’
24		<i>megbízható</i> ‘reliable’
25		<i>való</i> ‘-’
26		<i>foglalkozó</i> ‘-’
27		<i>megfizethető</i> ‘affordable’
28		<i>béke</i> ‘peace’
29		<i>elhatározásunk</i> ‘we are determined’
30		<i>partnerség</i> ‘partnership’
31		<i>termelési</i> ‘production’
32		<i>szakpolitikák</i> ‘policies’
33		<i>növekedés</i> ‘growth’
34		<i>beleértve</i> ‘including’
35		<i>globális</i> ‘global’
36		<i>fenntartható</i> ‘sustainable’
37		<i>az</i> ‘the’
38		<i>gazdasági</i> ‘economic’

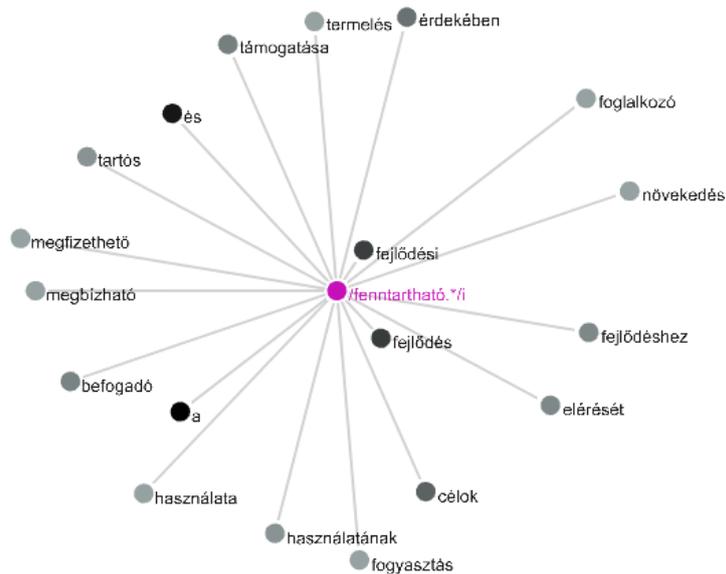
**Table 25.** List of collocates of *FENNTARTHATÓ* ‘sustainable’ in the 2030 Agenda Corpus (Hungarian).

Within one word to the left and one word to the right, the Hungarian adjective *FENNTARTHATÓ* ‘sustainable’ collocates with the content words *fejlődési* ‘development’, *fejlődés* ‘development’, *fejlődéshez* ‘for development’, *használatának* ‘use’ and with the function word *a* ‘the’. As it can be seen in Figure 10, the collocates *fejlődési* ‘development’, *fejlődés* ‘development’, *fejlődéshez* ‘for development’, *használatának* ‘use’ tend to occur on the right-hand side of *FENNTARTHATÓ* ‘sustainable’ while the collocate *a* ‘the’ tends to appear on its left-hand side.



**Figure 10.** Collocation network of *FENNTARTHATÓ* ‘sustainable’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.

When expanding the collocation window to three words to the left and three words to the right, *FENNTARTHATÓ* ‘sustainable’ collocates with other content and function words. Some of these content and function words are not collocates of the sole *FENNTARTHATÓ* ‘sustainable’ but of the lexical item *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’. When the adjective occurs independently of the lemma *FEJLŐDÉS* ‘development’, it collocates with the content words *támogatása* ‘support’, *fogyasztás* ‘consumption’, *termelés* ‘production’, *befogadó* ‘inclusive’, *használat* ‘use’, *tartós* ‘sustained’, *megbízható* ‘reliable’, *növekedés* ‘growth’, *megfizethető* ‘affordable’ and with the function words *és* ‘and’, *érdekében* ‘for’. These collocates are added to the ones already mentioned for the  $\pm 1$ -word span. As it can be seen in Figure 11, the collocates *befogadó* ‘inclusive’, *és* ‘and’, *tartós* ‘sustained’, *megbízható* ‘reliable’, *megfizethető* ‘affordable’ tend to appear on the left-hand side of the node, whereas the collocates *támogatása* ‘support’, *fogyasztás* ‘consumption’, *termelés* ‘production’, *használat* ‘use’, *érdekében* ‘for’, *növekedés* ‘growth’ tend to occur on the right-hand side of the node.



**Figure 11.** Collocation network of *FENNTARTHATÓ* 'sustainable' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.

Within five words to the left and five words to the right of the node, *FENNTARTHATÓ* 'sustainable' collocates with the content words *erdőgazdálkodás* 'forest management', *megőrzése* 'conserve', *fogyasztási* 'consumption', *módok* 'patterns', *elhatározásunk* 'we are determined', *termelési* 'production', *gazdasági* 'economic' and the function word *az* 'the'. These collocates are added to the list of the collocates found with smaller collocation windows. As it can be seen in Figure 12, the collocates *megőrzése* 'conserve', *elhatározásunk* 'we are determined' tend to stand on the left-hand side of the node, while the collocates *erdőgazdálkodás* 'forest management', *fogyasztási* 'consumption', *módok* 'patterns', *termelési* 'production', *gazdasági* 'economic', *az* 'the' tend to occur on the right-hand side of the node.



phrase *fenntartható fejlődés* into the adjective phrase *fenntartható fejlődési*, literally ‘of sustainable development’, with the use of the derivative suffix *-i*, that turns the noun *fejlődés* ‘development’ into an adjective that specifies the noun it is associated with with a general meaning of “belonging to/from/in/...N” (Kenesei et al. 1998: 364). The prominence of the collocate *development* and of its translational equivalents does not come as a surprise.

The other collocate shared by the Hungarian and English collocates’ lists with a  $\pm 1$ -word span, namely the noun *használatának* ‘use’, is the dative form of the lexeme *használata*, which is itself the result of the suffixation of the singular nominative form *használat* ‘use’ with the possessive suffix of third person singular. As in English, *használatának* ‘use’ is always preceded by the adjective *fenntartható* ‘sustainable’, constituting the noun phrase *fenntartható használatának*, literally ‘to sustainable use’, that functions as the singular possessor modifying the nouns *támogatása* ‘promote’ ( $f=2$ ), *biztosítása* ‘ensure’ ( $f=1$ ), and *erősítése* ‘enhance’ ( $f=1$ ).

Within one word to the left and one word to the right of *FENNTARTHATÓ* ‘sustainable’, the only collocate that appears in the Hungarian collocation list but that is absent from the English is the definite article *a* ‘the’. The definite article appears in English with a span of  $\pm 3$  words, but it is used in Hungarian and English in a comparable way. The determiner features among the collocates because of its preceding the adjective *FENNTARTHATÓ* ‘sustainable’ in roughly 63% of its occurrences (118 out of 187). Its presence is a strong indicator of the definiteness of the phrases including the adjective. It seems thus that the property of sustainability is so well-known and widespread that anything that can be defined as sustainable is also ontologically definite. This is true for the lemma *FEJLŐDÉS* ‘development’, which makes up for 70% of the nouns modified by *FENNTARTHATÓ* ‘sustainable’, but also for the other nouns modified by the adjective. These nouns correspond to the ones observed for English.

In addition, the adjective *FENNTARTHATÓ* ‘sustainable’ also takes the comparative form *fenntarthatóbb* ‘more sustainable’ ( $f=2$ ) when it modifies the noun phrase *fogyasztási és termelési módok* ‘patterns of consumption and production’. The use of the comparative *fenntarthatóbb* ‘more sustainable’ to qualify *fogyasztási és termelési módok* ‘patterns of consumption and production’ implies that the path towards sustainability for these consumption and production patterns does not start from scratch but that it begins from an already assessed level of sustainable development.

The same applies also to the English equivalent *more sustainable patterns of consumption and production*.

Among the collocation networks retrieved within three words to the left and three words to the right of the Hungarian *FENNTARTHATÓ* 'sustainable' and its English equivalent *sustainable*, the 2030 Agenda (Hungarian) and the 2030 Agenda (English) share the following collocates: the content words *sustained-tartós*, *consumption-fogyasztás*, *inclusive-befogadó*, *production-termelés*, *patterns-módok*, *growth-növekedés*, *reliable-megbízható*, *affordable-megfizethető*, and the function words *for-érdekében*, *and-és*, and *the-a*. The use of these lexemes in the Hungarian and English versions of the 2030 Agenda corresponds. Within three words to the left and three words to the right, the collocation network of the Hungarian adjective also includes the lexeme *támogatása* 'support', which is employed as its translational equivalent.

Between five words to the left and five words to the right of *FENNTARTHATÓ* 'sustainable', the Hungarian and English Agendas share some other collocates. Among the content words, the collocates in common are *use-használata*, *használatának* and *economic-gazdasági*. Among the function words, the two collocate lists include *the-a*, *az*. In addition to these shared collocates and to the collocates that have already been discussed in relation to the narrower collocation windows, the collocational pattern of the Hungarian *FENNTARTHATÓ* 'sustainable' also includes some peculiar content words (*erdőgazdálkodás* 'forest management', *elhatározásunk* 'we are determined', *megőrzése* 'conserve' and *fenntartható* 'sustainable').

The noun *elhatározásunk*, which functions as the translational equivalent of the English 'we are determined', introduces some opinions expressed in the 2030 Agenda in relation to sustainable development. It consists of the singular noun *elhatározás*, literally 'decision', followed by the first person plural possessive suffix *-unk*. The noun *elhatározás* 'decision' derives from the verb stem *elhatároz-* 'decide' and from the nominalizing suffix *-ás*, whose back vowel depends on the back-vowel harmony of the verb. In the sentences including *elhatározásunk* 'we are determined', the adjective *FENNTARTHATÓ* 'sustainable' appears in subordinate clauses that specify what has been decided and thus sustainability characterizes what has been agreed on.

The noun *erdőgazdálkodás* 'forest management' is a compound including the nouns *erdő* 'forest' and *gazdálkodás* 'management'. It is directly modified by the adjective *fenntartható* 'sustainable' in the noun phrase *fenntartható erdőgazdálkodás* 'sustainable forest management'. The sustainable management of forests is associated with other actions in favour of the environment and, furthermore, the document writes

that sustainable forest management is to be promoted and funded (*elősegítésére* 'to advance',  $f=1$ , *finanszírozására* 'to finance',  $f=1$ ).

The adjective *FENNTARTHATÓ* 'sustainable' collocates with the very *fenntartható* 'sustainable' because of noun phrases where two noun phrases including *FENNTARTHATÓ* 'sustainable' are coordinated or because of the proximity of independent noun phrases modified by *FENNTARTHATÓ* 'sustainable'.

The lexeme *megőrzése* 'conserve' consists of the singular noun *megőrzés*, literally 'conservation', and of the third person singular possessive suffix *-e*. The noun is almost always found in noun phrases associated by coordination with another noun phrase that features the adjective *FENNTARTHATÓ* 'sustainable'. In alternative, the noun phrase containing *megőrzése* 'conserve' can include another noun phrase featuring *FENNTARTHATÓ* 'sustainable'.

The meaning by collocation of the Hungarian adjective *FENNTARTHATÓ* 'sustainable' in the 2030 Agenda (Hungarian) is comparable to the meaning by collocation that the English adjective *sustainable* features in the United Nations' resolution. This depends on the correspondence between the Hungarian and English collocates' lists.

The correspondence between the two collocates' lists generates a correspondence between semantic preferences. The Hungarian adjective *FENNTARTHATÓ* 'sustainable' tends to co-occur first with lexemes having to do with the 2030 Agenda and its sustainable development goals. This semantic preference is typical of the adjective *FENNTARTHATÓ* 'sustainable' when it modifies the noun *FEJLŐDÉS* 'development' in the lexical item *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development'. The general semantic preference of the Hungarian adjective leans over material and mental processes of backing actions (*támogatása* 'support', *elhatározásunk* 'we are determined'). It is also associated with human activities that should be carried on under the light of sustainability according to the UN's document (*erdőgazdálkodás* 'forest management', *növekedés* 'growth', *gazdasági* 'economic'), especially when they involve depletion or protection of resources (*termelés* 'production', *használatának* 'use', *megőrzése* 'conserve', *fogyasztás* 'consumption', *fogyasztási* 'consumption', *módok* 'patterns', *használata* 'use', *termelési* 'production'). Furthermore, sustainability is combined with qualities of reliability, feasibility and inclusivity (*tartós* 'sustained', *befogadó* 'inclusive', *megbízható* 'reliable', *megfizethető* 'affordable').

Differences between the meaning by collocation of the Hungarian *FENNTARTHATÓ* 'sustainable' and the English *sustainable* arise at a grammatical level

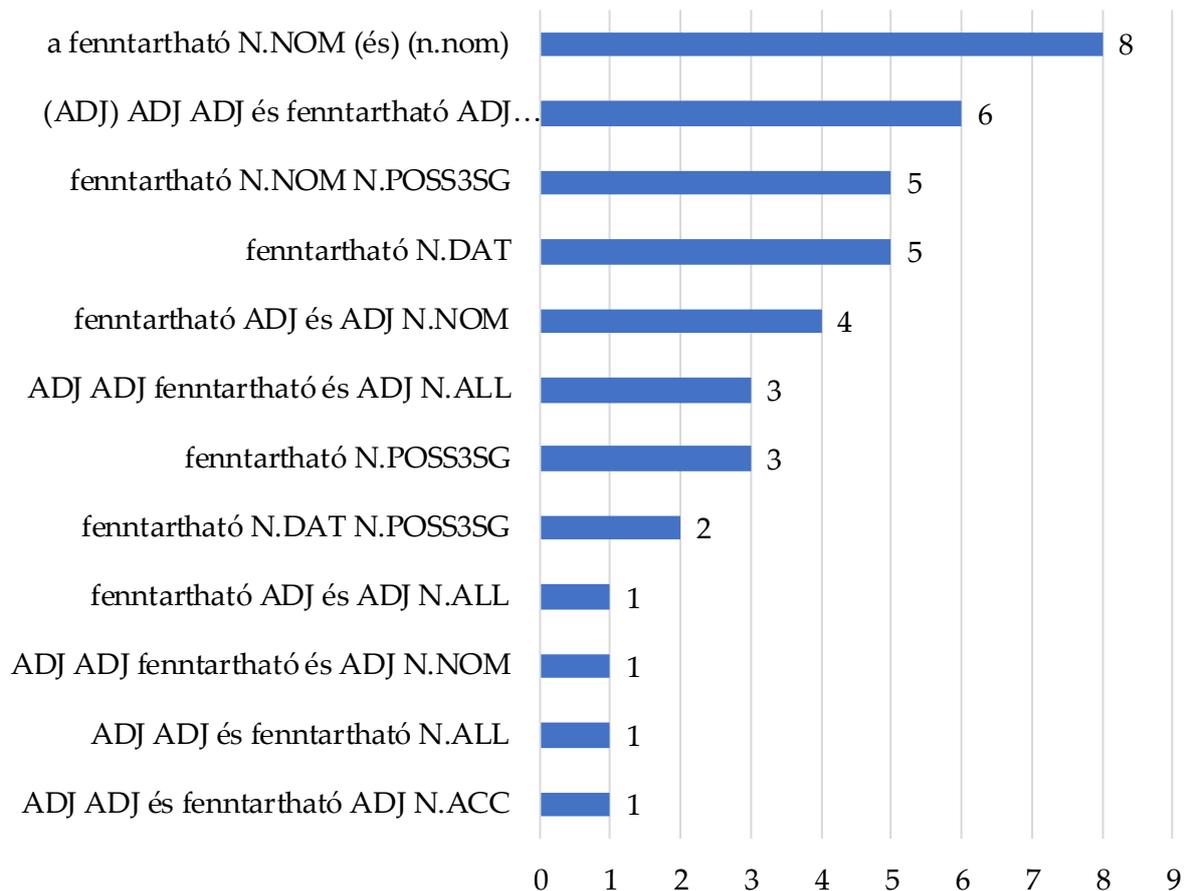
and they are reflected in the most frequent colligational patterns of the Hungarian adjective:

	<i>a</i>		<i>fenntartható</i>	N.NOM	(és)		(N.NOM)
			<i>fenntartható</i>	N.DAT			
			<i>fenntartható</i>	N.POSS3SG			
			<i>fenntartható</i>	N.NOM		N.POSS3SG	
			<i>fenntartható</i>	N.DAT		N.POSS3SG	
	ADJ	ADJ	és	<i>fenntartható</i>	N.ALL		
(ADJ)	ADJ	ADJ	és	<i>fenntartható</i>	ADJ	N.NOM	
	ADJ	ADJ	és	<i>fenntartható</i>	ADJ	N.ACC	
		ADJ	ADJ	<i>fenntartható</i>	és	ADJ	N.NOM
		ADJ	ADJ	<i>fenntartható</i>	és	ADJ	N.ALL
			<i>fenntartható</i>	ADJ	és	ADJ	N.NOM
			<i>fenntartható</i>	ADJ	és	ADJ	N.ALL

In these colligational patterns, NOM stands for “nominative”, DAT stands for “dative”, ACC stands for “accusative”, ALL stands for allative, POSS3SG stands for “third person singular possessive”.

The most frequent colligational patterns of *FENNTARTHATÓ* ‘sustainable’ reflect morphosyntactic patterns where the adjective belongs to an adjective phrase, which modifies a noun phrase ending with various cases. The main cases reported by these colligational patterns are the nominative, the dative, the allative and the accusative.

The colligational patterns of the adjective *FENNTARTHATÓ* ‘sustainable’ can be seen also in Graph 10. Graph 10 reproduces these colligations together with their frequency in the 2030 Agenda (Hungarian).



**Graph 10.** The colligational patterns of *FENNTARTHATÓ* ‘sustainable’ in the 2030 Agenda (Hungarian).

As in the English case, thanks to the semantic preference and the colligational patterns of *FENNTARTHATÓ* ‘sustainable’, the evaluative prosody of the adjective can be regarded as overall positive.

#### 4.2.3. The Italian *SOSTENIBILE*

The examination of the meaning by collocation of *sustainable* is then bolstered with the study of the meaning by collocation of the Italian adjective *SOSTENIBILE* ‘sustainable’.

The adjective *SOSTENIBILE* ‘sustainable’ occurs 180 times in the Italian version of the 2030 Agenda. It modifies the noun *sviluppo* ‘development’ 106 times in the noun phrase *sviluppo sostenibile* ‘sustainable development’. Thus, the adjective is found in this noun phrase in 59% of its occurrences. As in the English and Hungarian cases, also the extraction of collocational patterns for the Italian translational equivalent of

*sustainable* is skewed by the high frequency of the noun phrase and this was taken into consideration when exploring the collocation of the adjective.

#### 4.2.3.1. Word sketch

The meaning by collocation of *SOSTENIBILE* ‘sustainable’ is outlined first thanks to the word sketch that can be seen in Figure 13.



**Figure 13.** Word sketch of *SOSTENIBILE* ‘sustainable’ in the 2030 Agenda Corpus (Italian) – LogDice(6.0), NC5-C5.

The word sketch of *SOSTENIBILE* ‘sustainable’ is less elaborate than the word sketches of the English and Hungarian corresponding adjectives (Figures 5 and 9 respectively). This might depend on the features of the different annotation schemas used for English, Hungarian and Italian in Sketch Engine. Different annotation schemas, in fact, might change the way corresponding lexemes are labelled and can be retrieved cross-linguistically.

The word sketch of the Italian adjective stresses that the nouns that are most significantly modified by the adjective are *SVILUPPO* ‘development’ and *GESTIONE* ‘management’.

*SOSTENIBILE* ‘sustainable’ and *SVILUPPO* ‘development’ are always found together in the key noun phrase *sviluppo sostenibile* ‘sustainable development’, which is described in some detail in § 4.4.3. *SOSTENIBILE* ‘sustainable’ and *GESTIONE* ‘management’ collocate within the phrase *gestione sostenibile* ‘sustainable management’.

### 4.2.3.2. Collocation networks

The collocational tendencies highlighted for *SOSTENIBILE* 'sustainable' through its word sketch are further assessed with the retrieval of the adjective's collocation networks.

The collocational patterns extracted for the adjective *SOSTENIBILE* 'sustainable' with  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word spans in the 2030 Agenda (Italian) can be seen in Table 26. The lists of the collocates extracted within the three collocation windows are included also in Tables 52, 53, 54 in the Appendix. In these tables, collocates are reported together with their position relative to the node, with the collocation's Z-value, with the frequency of co-occurrence of node and collocate and with the frequency of occurrence of the sole collocate in the corpus.

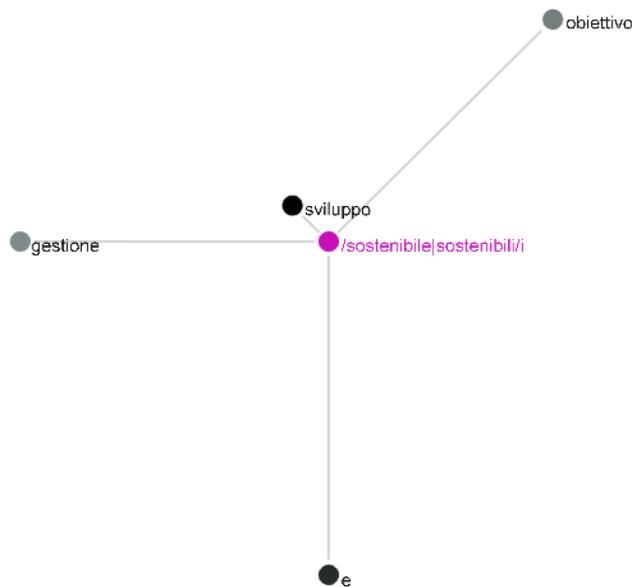
Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>sviluppo</i> 'development'	<i>sviluppo</i> 'development'	<i>sviluppo</i> 'development'
2	<i>gestione</i> 'management'	<i>lo</i> 'the'	<i>lo</i> 'the'
3	<i>obiettivo</i> 'goal'	<i>uno</i> 'a'	<i>obiettivi</i> 'goals'
4	<i>e</i> 'and'	<i>dello</i> 'of (the)'	<i>e</i> 'and'
5		<i>obiettivi</i> 'goals'	<i>per</i> 'for'
6		<i>per</i> 'for'	<i>uno</i> 'a'
7		<i>e</i> 'and'	<i>dello</i> 'of (the)'
8		<i>sullo</i> 'on (the)'	<i>degli</i> 'of (the)'
9		<i>inclusiva</i> 'inclusive'	<i>raggiungimento</i> 'achievement'
10		<i>gestione</i> 'management'	<i>duratura</i> 'sustained'
11		<i>obiettivo</i> 'goal'	<i>di</i> 'of'
12		<i>produzione</i> 'production'	<i>obiettivo</i> 'goal'
13		<i>oceani</i> 'oceans'	<i>produzione</i> 'production'
14		<i>di</i> 'of; to'	<i>consumo</i> 'consumption'
15		<i>promuovere</i> 'promote; foster'	<i>inclusiva</i> 'inclusive'
16			<i>sullo</i> 'on (the)'
17			<i>gestione</i> 'management'
18			<i>affidabili</i> 'reliable'
19			<i>la</i> 'the'
20			<i>promuovere</i> 'promote; foster'
21			<i>infrastrutture</i> 'infrastructure'
22			<i>oceani</i> 'oceans'
23			<i>economica</i> 'economic'
24			<i>che</i> 'that'
25			<i>gli</i> 'the'
26			<i>crescita</i> 'growth'

27	<i>il</i> 'the'
28	<i>sistemi</i> 'systems'
29	<i>garantire</i> 'provide'
30	<i>risorse</i> 'resources; sources'
31	<i>a</i> 'to; in; on'
32	<i>un</i> 'a'
33	<i>tutti</i> 'all'
34	<i>traguardi</i> 'targets'
35	<i>una</i> 'a'
36	<i>ad</i> 'to; in; on'

**Table 26.** List of collocates of *SOSTENIBILE* 'sustainable' in the 2030 Agenda Corpus (Italian).

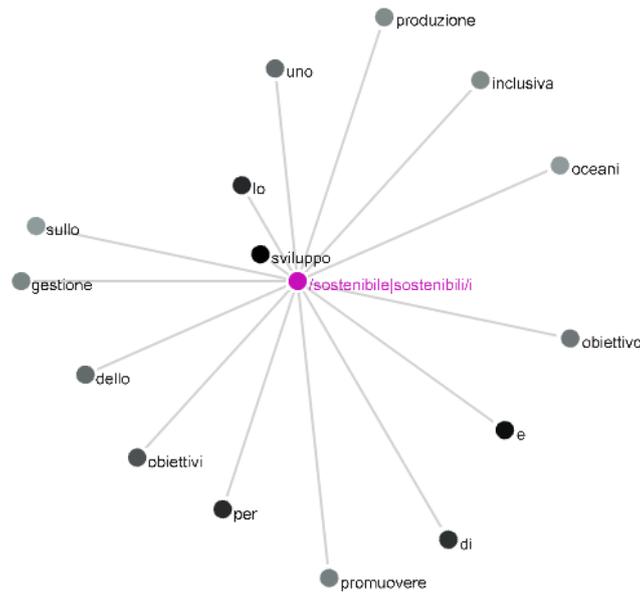
As it can be noted when comparing Table 26 with Figure 13, the collocates of *SOSTENIBILE* 'sustainable' identified with Sketch Engine's word sketch are all included in the collocation networks extracted with #LancsBox. The two collocates highlighted by the word sketch, namely *sviluppo* 'development' and *gestione* 'management', lay the foundations for the meaning by collocation of the adjective. In fact, they recognise that sustainability is first and foremost a property of development and management of resources. The collocation networks enrich this observation with a wider variety of collocational patterns.

Within one word to the left and one word to the right of *SOSTENIBILE* 'sustainable', the adjective collocates both with content words and with a function word. The only words that co-occur with *SOSTENIBILE* 'sustainable' without it modifying also *sviluppo* 'development' are the content word *gestione* 'management' and the function word *e* 'and'. As it can be seen in Figure 14, *gestione* 'management' tends to appear on the left-hand side of the adjective, whereas *e* 'and' tends to occur on its right-hand side.



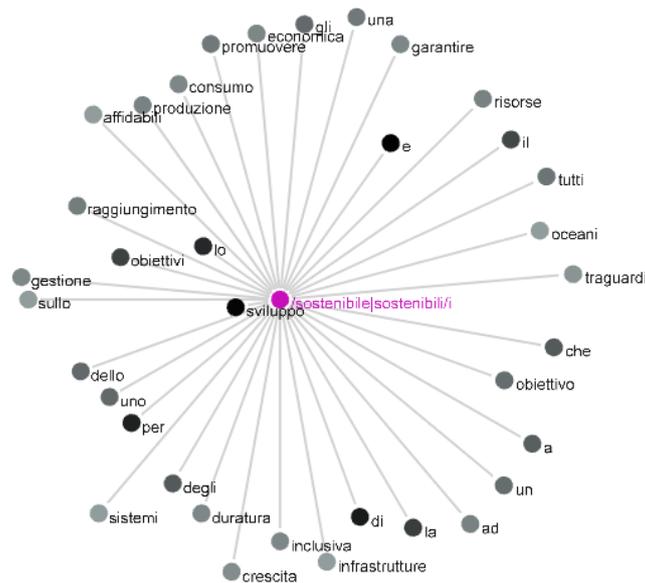
**Figure 14.** Collocation network of *SOSTENIBILE* ‘sustainable’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.

With a  $\pm 3$ -word span, *SOSTENIBILE* ‘sustainable’ collocates with the content words *inclusiva* ‘inclusive’, *produzione* ‘production’, *oceani* ‘oceans’, *promuovere* ‘promote; foster’ and with the function words *per* ‘for’, *di* ‘of’. These lexemes co-occur with *SOSTENIBILE* ‘sustainable’ in addition to the ones already mentioned for the  $\pm 1$ -word span and without the adjective modifying the noun *sviluppo* ‘development’. As it can be seen in Figure 15, the collocates *inclusiva* ‘inclusive’, *produzione* ‘production’, *di* ‘of’, *promuovere* ‘promote; foster’ tend to stand on the left-hand side of the node, whereas the collocates *oceani* ‘oceans’ tend to appear on its right-hand side.



**Figure 15.** Collocation network of *SOSTENIBILE* in the 2030 Agenda Corpus (Italian) – Z(10.0), 3L-3R, NC5-C5.

Within five words to the left and five words to the right, *sostenibile* ‘sustainable’ collocates also with the content words *duratura* ‘sustained’, *consumo* ‘consumption’, *affidabili* ‘reliable’, *infrastrutture* ‘infrastructures’, *economica* ‘economic’, *crescita* ‘growth’, *sistemi* ‘systems’, *garantire* ‘provide’, *risorse* ‘resources’, *tutti* ‘all’ and with the function words *la* ‘the’, *che* ‘that’, *gli* ‘the’, *il* ‘the’, *a* ‘to; in; on’, *un* ‘a’, *una* ‘a’. *SOSTENIBILE* ‘sustainable’ collocates with these lexemes when the adjective modifies a word other than *sviluppo* ‘development’. As Figure 16 shows, the collocates *duratura* ‘sustained’, *consumo* ‘consumption’, *affidabili* ‘reliable’, *infrastrutture* ‘infrastructures’, *economica* ‘economic’, *gli* ‘the’, *crescita* ‘growth’, *sistemi* ‘systems’, *una* ‘a’ tend to occur on the left-hand side of the node, while the collocates *la* ‘the’, *che* ‘that’, *il* ‘the’, *risorse* ‘resources; sources’, *a* ‘to; in; on’, *un* ‘a’, *tutti* ‘all’ tend to stand on the right-hand side of the node. *garantire* ‘provide’ stands both on the left- and on the right-hand side of the node.



**Figure 16.** Collocation network of *SOSTENIBILE* ‘sustainable’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 5L-5R, NC5-C5.

Comparing the collocation networks extracted for the Italian adjective *SOSTENIBILE* ‘sustainable’ (Table 26) with the collocation networks extracted for the English adjective *sustainable* (Table 24) and for the Hungarian adjective *FENNTARTHATÓ* (Table 25), several overlaps can be highlighted.

Within one word to the left and one word to the right, the three collocates’ lists share the lexeme *development* and its translational equivalents, namely *sviluppo* for Italian and *fejlesztési, fejlődés, fejlődéshez* for Hungarian. The lexemes are used in the same way in the three languages. In addition, the English collocate *management* is paired in Italian by the collocate *gestione*. The English *management* and its Italian counterpart *gestione* can be found in equivalent patterns in the two versions of the document. While the Hungarian translational equivalents of *development* appear for the first time with the  $\pm 1$ -word span, the translational equivalents of *management* do not appear among the collocates of *FENNTARTHATÓ* ‘sustainable’, not even when enlarging the collocation window up to ten words to the left and ten words to the right of the node. This depends on the fact that, while the English and the Italian documents use a single word type in these contexts, the Hungarian resolution employs various translational equivalents: *gazdálkodás* ( $f=1$ ), *gazdálkodástól* (with the ablative suffix *-tól*,  $f=1$ ), *kezelés* ( $f=1$ ), the compound *vízgazdálkodás* for the English ‘management of water’ ( $f=1$ ), or even the omission of the word.

In Italian, the  $\pm 1$ -word collocate list of *SOSTENIBILE* 'sustainable' also includes the conjunction *e* 'and'. The English and the Hungarian equivalents of *e*, namely *and* and *és*, can be found within three words to the left and three words to the right of the word. The cross-linguistic difference in distance between node and collocate depends on syntactic reasons but the use of the lexemes is the same.

Expanding the collocation window to three words to the left and three words to the right, the collocates' lists of the English *sustainable*, of the Italian *SOSTENIBILE* 'sustainable', and of the Hungarian *FENNTARTHATÓ* 'sustainable' share the following lexemes: the content words *inclusive-inclusiva-befogadó*, *production-produzione-termelés*, and the function word *for-per-érdekében*. In addition, the English and the Italian collocates' lists also share the lexemes *promote-promuovere* and *of-di*. All groups or pairs of equivalent lexemes function in equivalent ways in the three languages. In particular, in relation to the *production-produzione-termelés* correspondence, as it has been noted for Hungarian, also in Italian the adjective *sostenibile* 'sustainable' can be found in its comparative form in the pattern *modelli di consumo e produzione più sostenibili* 'more sustainable consumption and production patterns' ( $f=1$ ) with the alternative *modelli di consumo e di produzione più sostenibili* 'more sustainable patterns of consumption and production' ( $f=1$ ).

While with the  $\pm 3$ -word span the collocational patterns of the English *sustainable* and of the Hungarian *FENNTARTHATÓ* overlap with the content words *affordable-megfizethető*, *consumption-fogyasztás*, *growth-növekedés*, *reliable-megbízható*, and *sustained-tartós*, the Italian translational equivalents *affidabili* 'reliable', *consumo* 'consumption', *crescita* 'growth', and *duratura* 'sustained' appear with a span of  $\pm 5$  mainly because of syntactic reasons. In addition, the English adjective *affordable* does not feature among the collocates of the Italian *SOSTENIBILE* 'sostenibile', not even when enlarging the collocation window to  $\pm 10$  words to the left and to the right of the node. This depends on the wide variety of the adjective's translational equivalents. *affordable*, in fact, can be rendered with *economici* ( $f=3$ ), *conveniente* ( $f=1$ ), and *economicamente accessibile* ( $f=1$ ).

The only collocate that stands out in the Italian collocational pattern of *SOSTENIBILE* 'sustainable' in comparison with English and Hungarian is *oceani* 'oceans'. The English noun *oceans* is included in the collocates' list of *sustainable* when the collocation window is widened to  $\pm 7$  words, while it appears among the collocates of the Hungarian *FENNTARTHATÓ* 'sustainable' with a span of  $\pm 9$  words. Oceans are one of the natural elements that the 2030 Agenda encourages to treat sustainably. The lexeme is not directly modified by *SOSTENIBILE* 'sustainable', though. In this collocational

pattern, in fact, *SOSTENIBILE* 'sustainable' completes clauses and phrases like *utilizzare in modo sostenibile* 'sustainably use' or *utilizzo sostenibile* 'sustainable use', which are in turn modified by *oceani* 'oceans'.

Between five words to the left and five words to the right, the collocates' lists of the English *sustainable*, of the Italian *SOSTENIBILE* 'sustainable', and of the Hungarian *FENNTARTHATÓ* 'sustainable' share also the collocates *sustained-duratura-tartós*, *consumption-consumo-fogyasztás,fogyasztási*, *growth-crescita-növekedés*, *reliable-affidabili-megbízható*, and *economic-economica-gazdasági*. Among the function words, all collocate lists include *the-il,la,lo-a,az*. The English and the Italian collocation networks also share the function words *all-tutti* and *that-che*. The pairs or groups of lexemes behave similarly in the three languages.

Within the  $\pm 5$ -word collocation window, the collocational patterns of the English *sustainable* and of the Hungarian *FENNTARTHATÓ* display also the translational pairs *use-használata,használatának* and *affordable-megfizethető*. As it has already been observed for *affordable*, the Italian translational equivalent of *use* does not feature among the collocates of *SOSTENIBILE* 'sustainable', not even with a span of  $\pm 10$  words. This is mainly due to the fact that the English *use* is translated in Italian with a variety of word types, whose low frequency cuts them out of the frequency threshold set for the extraction of the collocates.

Within five words to the left and five words to the right, the Italian *SOSTENIBILE* 'sustainable' collocates also with the content words *infrastrutture* 'infrastructures', *sistemi* 'systems', *garantire* 'provide', *risorse* 'resources; sources', and with the function words *un* 'a' and *uno* 'a'.

Of these, *SOSTENIBILE* 'sustainable' directly modifies the noun *sistemi* 'systems' in the noun phrases *sistemi di energia economici, affidabili, sostenibili e moderni* 'affordable, reliable, sustainable and modern energy systems' ( $f=2$ ), *sistemi di produzione alimentare sostenibili* 'sustainable food production systems' ( $f=1$ ), *sistemi di trasporto sostenibili* 'sustainable transport systems' ( $f=1$ ). These occurrences testify that the Agenda begs for sustainability in the systematic management of energy, food production, and transport. The same sustainable management is required also for natural and marine resources.

The collocate *risorse* 'resources; sources', in fact, triggers the use of noun phrases or prepositional phrases that refer to a sustainable management of nature and seas, as in *della gestione sostenibile delle risorse naturali* 'on the sustainable management of our planet's natural resources' ( $f=1$ ), *la gestione sostenibile e l'utilizzo efficiente delle risorse*

*naturali* ‘the sustainable management and efficient use of natural resources’ ( $f=1$ ), and *un utilizzo più sostenibile delle risorse marine* ‘the sustainable use of marine resources’ ( $f=1$ ).

The meaning by collocation of the Italian adjective *SOSTENIBILE* ‘sustainable’ in the 2030 Agenda Corpus parallels the meaning by collocation of its English and Hungarian equivalent lexemes. This ensues from the almost thorough equivalence of the collocational patterns of the three adjectives.

The equivalence of the adjectives’ collocational patterns is first an equivalence of semantic preferences. As in the English and Hungarian cases, the collocational patterns of *SOSTENIBILE* ‘sustainable’ are characterised by a semantic preference for qualities of inclusiveness and trustworthiness (*duratura* ‘sustained’, *inclusiva* ‘inclusive’, *affidabili* ‘reliable’), for material processes of production, consumption and managing (*produzione* ‘production’, *consumo* ‘consumption’, *gestione* ‘management’, *crescita* ‘growth’), for mental processes of boosting (*promuovere* ‘promote; foster’, *garantire* ‘provide’), but also for human activities and products (*infrastrutture* ‘infrastructures’, *economica* ‘economic’, *sistemi* ‘systems’), as well as for natural elements (*oceani* ‘oceans’, *risorse* ‘resources; sources’).

The differences in the meanings by collocation of the Italian, English and Hungarian adjectives mainly stem from the colligational patterns that involve the adjectives. The most significant colligational patterns of the Italian *SOSTENIBILE* ‘sustainable’ are:

				DET	N		<i>sostenibile</i>				
				N	ADJ		<i>sostenibile</i>	(ADJ)	<i>e</i>	ADJ	
	N	P ( <i>di</i> )	N	(ADJ)	(ADJ)	(ADJ)	<i>sostenibile</i>	<i>e</i>		ADJ	
	(DET)	N		ADJ	(ADJ)	(ADJ)	<i>e</i>		<i>sostenibile</i>		
	N	P ( <i>di</i> )	(DET)	N	<i>e</i>	N	<i>non</i>		<i>sostenibile</i>		
	N	P ( <i>di</i> )	(DET)	N	<i>e</i>	N	<i>più</i>		<i>sostenibile</i>		
				DET	N		<i>più</i>	<i>sostenibile</i>	P ( <i>di</i> )	N	ADJ

(P) (DET) N *sostenibile* P (di) N (e) (P (N) (di))

(P (DET) N *e* (P (DET) N *sostenibile* (a))

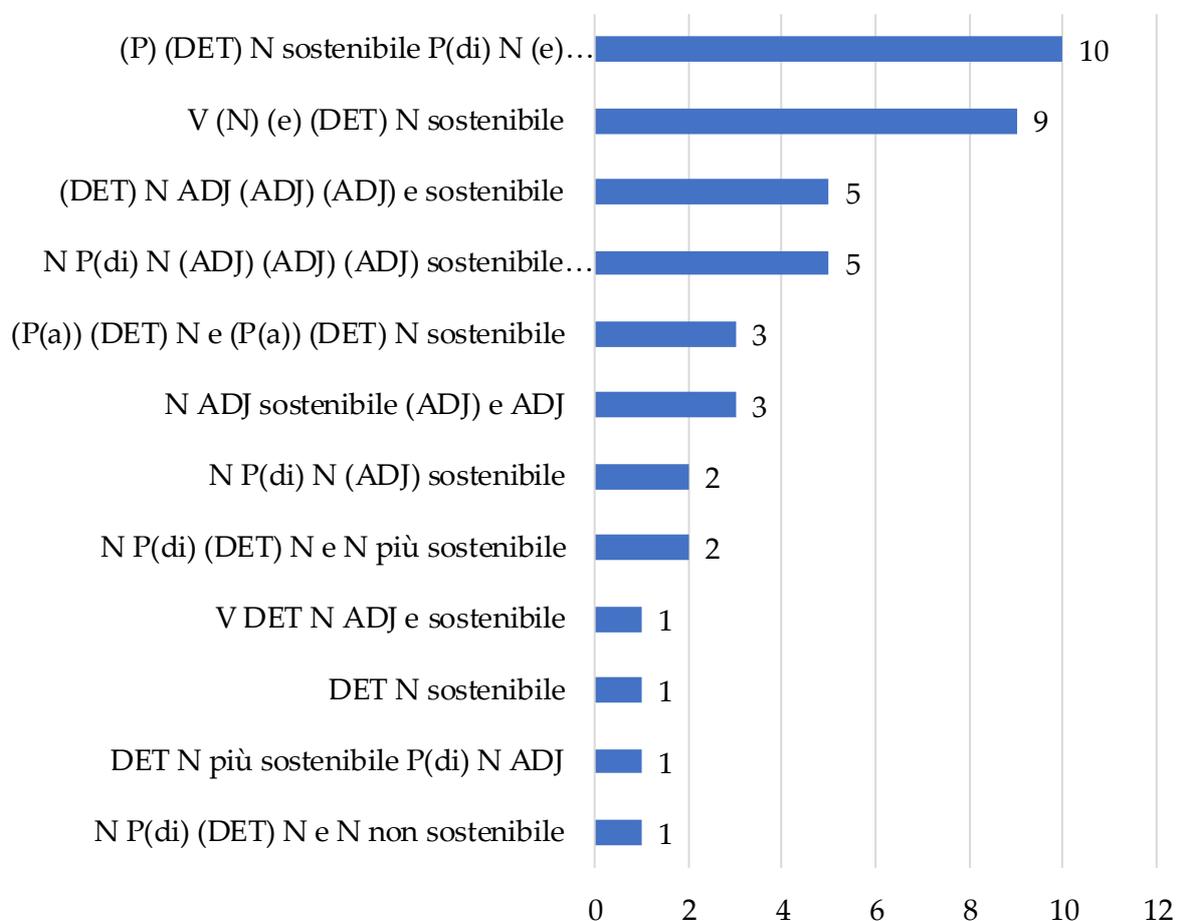
N P (di) N (ADJ) *sostenibile*

V (N) (e) (DET) N *sostenibile*

V DET N ADJ *e* *sostenibile*

In these colligational patterns, DET stands for “determiner”.

These colligational patterns are represented also in Graph 11 together with their frequency in the 2030 Agenda (Italian).



**Graph 11.** The colligational patterns of *SOSTENIBILE* ‘sustainable’ in the 2030 Agenda (Italian).

Thanks to semantic preference and to the connotation acquired by the colligational patterns that involve *SOSTENIBILE* ‘sustainable’, the Italian adjective acquires a positive evaluative prosody in the 2030 Agenda Corpus, as do the English *sustainable* and the Hungarian *FENNTARTHATÓ* ‘sustainable’.

### 4.3. Meaning by collocation of *sustainability* in the 2030 Agenda Corpus

After establishing the meaning by collocation of the English adjective *sustainable* and of its Hungarian and Italian translational equivalents, the study of the discursive construction of sustainable development advances with the analysis of the meaning by collocation of the English noun *SUSTAINABILITY* and of its Hungarian and Italian translational equivalents.

The meaning by collocation of the noun *SUSTAINABILITY* (from now on simply *sustainability* because of its invariable form), a cognate of the adjective *sustainable*, and of its Hungarian and Italian translational equivalents (i.e. *FENNTARTHATÓSÁG* and *SOSTENIBILITÀ* respectively) is traced in this section thanks to the analysis of the lexemes’ word sketches.

Word sketches are built with the Sketch Engine platform, adopting the same methodological choices already mentioned for the extraction of word sketches for the adjective *sustainable* and its Hungarian and Italian translational equivalents. Word sketches are computed by typing a simple query (i.e. *sustainability* for English, *fenntarthatóság* for Hungarian and *sostenibilità* for Italian); the typed lexical items were considered as a word by the Sketch Engine platform.

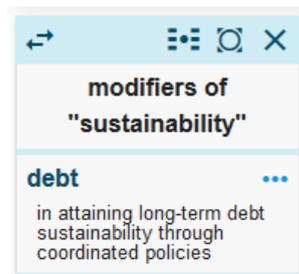
Collocation networks are not employed to find the meaning by collocation of *sustainability* and its Hungarian and Italian counterparts because of the low frequency of the words: the English *sustainability* appears six times in the English version of the 2030 Agenda, the Hungarian *FENNTARTHATÓSÁG* ‘sustainability’ occurs seven times in the Hungarian version of the 2030 Agenda, the Italian *SOSTENIBILITÀ* ‘sustainability’ is counted five times in the Italian version of the 2030 Agenda. This low frequency would make the extraction of collocation networks pointless.

The collocational patterns of the English *sustainability* are deemed baseline for the retrieval of the meaning by collocation of the noun in the 2030 Agenda Corpus. Therefore, the Hungarian and Italian collocational patterns are devoted an in-depth exploration only when they differ from the collocational patterns of the English equivalent.

### 4.3.1. The English *SUSTAINABILITY*

The investigation of the meaning by collocation of *sustainability* begins with the analysis of the lexeme's collocational patterns in the 2030 Agenda (English).

The noun *sustainability* occurs only six times in the 2030 Agenda (English). Figure 17 reproduces the word sketch of the noun.



**Figure 17.** Word sketch of *SUSTAINABILITY* in the 2030 Agenda Corpus (English) – LogDice(6.0), NC5-C5.

As Figure 17 displays, the only modifier of *sustainability* is *debt*, which proves that the noun is mainly used in economic terms. It refers to the affordability of debts for the world countries.

The compound *debt sustainability*, in fact, appears in five out of six occurrences of the noun. *debt sustainability* is explained under the light of the *Addis Ababa Action Agenda* through the relational verb *relates* ( $f=1$ ). The verb *relates* creates a relationship between the Addis Ababa Action Agenda and a wealth of issues including *debt sustainability*.

In addition, *debt sustainability* is the patient of a material process expressed by the nominalized verb *attaining* ( $f=2$ ) and the actor of the material process of attaining is *developing countries*. *developing countries*, as well as *small island developing States* and *some developed countries*, are also the beneficiaries of the action of backing debt sustainability. In similar cases, *debt sustainability* is the patient of the nominalized material process *maintenance*. The actor of this maintaining a sustainable debt are *those countries that have received debt relief and achieved sustainable debt levels*. These countries (namely developing countries together with small island developing States and with some developed countries) are the out-group of a *We* pronoun that opens the sentence. In fact, *We* seems to include only those developed countries that can offer their financial and organizational assistance to debt sustainability.

*debt sustainability*, however, can also be threatened whenever lending countries do not check on the feasibility of the debt. In this case, *a country's debt sustainability* is the patient of the material process of undermining, whose actor is the way lenders lend money to the indebted countries.

In the only occurrence of *sustainability* without *debt*, *sustainability* modifies *information* in the noun phrase *sustainability information*. The noun phrase plays the role of patient of the material process *to integrate*. The actor of the process is *companies*, who are asked to monitor sustainability and report on it.

### 4.3.2. The Hungarian *FENNTARTHATÓSÁG*

The exploration of the meaning by collocation of *sustainability* continues with the study of the collocational patterns of the Hungarian noun *FENNTARTHATÓSÁG* ‘sustainability’ in the 2030 Agenda (Hungarian).

In the 2030 Agenda (Hungarian), the lemma *FENNTARTHATÓSÁG* ‘sustainability’ appears only 7 times. As the word sketch in Figure 18 shows, the lemma is always used in the singular form.



**Figure 18.** Word sketch of *FENNTARTHATÓSÁG* ‘sustainability’ in the 2030 Agenda Corpus (Hungarian) – LogDice(6.0), NC5-C5.

In general, the Hungarian translational equivalent of *sustainability*, namely *FENNTARTHATÓSÁG*, is used exactly in the same contexts and with the same meaning of its English counterpart.

In 6 out of 7 instances *FENNTARTHATÓSÁG* ‘sustainability’ is used in the singular possessive form *fenntarthatósága*, literally ‘its sustainability’, with the noun stem *fenntarthatóság* ‘sustainability’ and the suffix marking a third person singular possessor *-a*. In almost half of these occurrences (3 out of 7), the lexical item *fenntarthatósága* ‘(its)

sustainability' is marked by the dative suffix *-nak* (*fenntarthatóságának*, literally 'to its sustainability'), which signals a possessor in possessive chains of the following sort:

<i>ország-ok</i>	<i>adósságállomány-Ø-a-</i>	<i>fenntarthatóság-Ø-á-nak</i>	<i>megőrzés-Ø-é-t</i>
	Ø		
country- PL	debt-SG-POSS.3PS- NOM	sustainability-SG-POSS.3PS- DAT	maintenance-SG-POSS.3PS- ACC

'the maintenance of debt sustainability of those countries'

As in the English case, sustainability is almost always found in relation to the welfare of developing countries and, in particular, to their possibility and difficulties in managing the debt they owe to developed countries. In fact, *FENNTARTHATÓSÁG* 'sustainability' is modified in five cases by 'debt'. In two cases it is preceded by the noun *ADÓSSÁG* 'debt' and in three cases by the compound *ADÓSSÁGÁLLOMÁNY* 'debt'. In addition, *FENNTARTHATÓSÁG* 'sustainability' modifies the lexical items *biztosítása* 'to ensure' ( $f=1$ ), *hosszú távú megvalósításában* 'in attaining long-term' ( $f=1$ ), *kérdéseikhez* '-', literally 'to its questions' ( $f=1$ ), and *megőrzését* 'the maintenance' ( $f=1$ ). The Agenda requires developed countries and especially creditors to help developing countries in guaranteeing debt sustainability.

Sustainability is almost exclusively associated with developing countries' debts and it triggers actions and reflections for its realisation. In the two instances where *FENNTARTHATÓSÁG* 'sustainability' does not collocate with *ADÓSSÁG* or *ADÓSSÁGÁLLOMÁNY* 'debt', the word seems nevertheless to refer to economic matters and to encourage the countries to check and keep track of their advances towards sustainability (as in the noun phrase *fenntarthatósággal kapcsolatos információk* 'sustainability information',  $f=1$ , or in *Az eddig elért eredmények fenntarthatóságának biztosítása érdekében* 'In order to ensure that achievements made to date are sustained',  $f=1$ ).

### 4.3.3. The Italian *SOSTENIBILITÀ*

The outline of the meaning by collocation of *sustainability* is then enhanced with the analysis of the collocational patterns of the Italian *SOSTENIBILITÀ* 'sustainability' in the 2030 Agenda (Italian).

In the 2030 Agenda (Italian), the Italian *SOSTENIBILITÀ* ‘sustainability’ occurs only five times. As in the case of its English and Hungarian counterparts, also in Italian the lexeme is tied to the concept of debt in more than half of its occurrences (3 out of 5). In these cases, the noun is followed by the prepositional phrase *del debito* ‘debt’. The noun phrases *sostenibilità del debito* ‘debt sustainability’ (with its variant *sostenibilità a lungo termine del debito* ‘long-term debt sustainability’) functions as the goal of material processes like reaching (in *affinché raggiungano la sostenibilità a lungo termine del debito* ‘in attaining long-term debt sustainability’), undermining (in *in modo da non indebolire la sostenibilità del debito di un paese* ‘in a way that does not undermine a country’s debt sustainability’), and maintaining (in *affinché [...] possano mantenere la sostenibilità del debito raggiunta* ‘the maintenance of debt sustainability’).

In the remaining two cases sustainability is equated to sustainable practices (in *ad adottare pratiche sostenibili e ad integrare le informazioni sulla sostenibilità nei loro resoconti annuali* ‘to adopt sustainable practices and to integrate sustainability information into their reporting cycle’) and it is combined with resilience (in *sulla strada della sostenibilità e della resilienza* ‘on to a sustainable and resilient path’).

#### **4.4. Meaning by collocation of *sustainable development* in the 2030 Agenda Corpus**

The discursive construction of sustainable development in the 2030 Agenda Corpus is addressed with the search for the meaning by collocation of the English adjective *sustainable* and of its Hungarian and Italian translational equivalents and then with the description of the meaning by collocation of the English noun *sustainability* and of its Hungarian and Italian translational equivalents. Finally, it is tackled through the gathering of the meaning by collocation of the English two-gram *SUSTAINABLE DEVELOPMENT* and its Hungarian and Italian translational equivalents. The details of the meaning by collocation of the English *SUSTAINABLE DEVELOPMENT* and of its Hungarian and Italian counterparts are reported in the following paragraphs.

The meaning by collocation of the English *SUSTAINABLE DEVELOPMENT* (from now on simply *sustainable development* because of its invariable form) and of its Hungarian and Italian translational equivalents (i.e. *FENNTARTHATÓ FEJLŐDÉS* and *SVILUPPO SOSTENIBILE* respectively) is outlined in this section through the description of word sketches and collocation networks.

Word sketches and collocation networks are extracted and studied with the same methodology already accounted for in relation to the extraction of the word sketches and of the collocation networks of the adjectives *sustainable*, *FENNTARTHATÓ* and *SOSTENIBILE* (see § 4.2 on this point). Sketch Engine's word sketches are retrieved by typing a simple query (i.e. *sustainable development* for English, *fenntartható fejlődés* for Hungarian and *sviluppo sostenibile* for Italian) and by demanding the platform to consider the typed lexical item as a word. #LancsBox's collocation networks are extracted by querying the software with the following strings: */sustainable development/* for English, */fenntartható fejlődés\*/* for Hungarian and */sviluppo sostenibile/* for Italian. The Hungarian lexical item is searched for with a wildcard so as to include its morphological variability, while the English and the Italian lexical items are searched for in their basic form because of their being unchangeable.

The meaning by collocation of the English *sustainable development* is esteemed baseline in the 2030 Agenda Corpus. As a consequence, the collocational patterns of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' and of the Italian *SVILUPPO SOSTENIBILE* 'sustainable development' are detailed only when they clash with the collocational patterns of the English *sustainable development*.

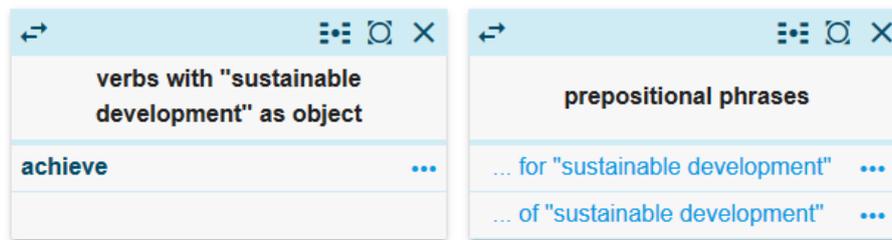
#### **4.4.1. The English *SUSTAINABLE DEVELOPMENT***

The meaning by collocation of *sustainable development* in the 2030 Agenda Corpus is first sought in the English version of the resolution.

The lexical item *sustainable development* appears 108 times in the 2030 Agenda (English). It is the second most frequent two-gram of the corpus and it follows only the grammatical two-gram *of the*, whose high frequency is due to the inherent structural characteristics of the English language.

##### **4.4.1.1. Word sketch**

The study of the meaning by collocation of *sustainable development* begins with the exploration of the word sketch of the lexical item (Figure 19).



**Figure 19.** Word sketch of *SUSTAINABLE DEVELOPMENT* in the 2030 Agenda Corpus (English) – LogDice(6.0), NC5-C5.

The word sketch of *sustainable development* shows that the node appears most significantly as the object of the verb *achieve*. This stresses the idea that sustainable development is viewed as a positive goal to be reached with some effort. In addition, *sustainable development* prominently features as the noun phrase of the prepositional phrases *for sustainable development* and *of sustainable development*.

The prepositional phrase *for sustainable development* modifies nouns like *requirement* ( $f=2$ ), *awareness* ( $f=1$ ), *challenge* ( $f=1$ ), *education* ( $f=1$ ), *foundation* ( $f=1$ ), and the adjective *essential* ( $f=2$ ). Through the nouns it modifies, the prepositional phrase suggests that sustainability is not an easy goal to achieve (*challenge*) and that it needs prerequisites to be met (*foundation, requirement*) and an enhanced level of sensitivity from the part of human beings (*awareness, education*). The prepositional phrase also functions as a circumstantial augment. Also in these cases, sustainability is meant as a target that can be advanced thanks to the promotion of inclusive and peaceful societies, thanks to the use and preservation of the seas and of their elements, thanks to the adoption of coherent policies.

On its closest left side, the prepositional phrase *of sustainable development* modifies the nouns *dimensions* ( $f=3$ ), *pursuit* ( $f=2$ ), *enabler* ( $f=1$ ), *enablers* ( $f=1$ ), and *promotion* ( $f=1$ ). This prepositional phrase testimonies for the complexity of sustainable development that should cover economic, social and environmental dimensions. It also highlights once again that sustainable development is a desired condition that has not been met yet and that should be worked for (*enabler, enablers, promotion, pursuit*).

#### 4.4.1.2. Collocation networks

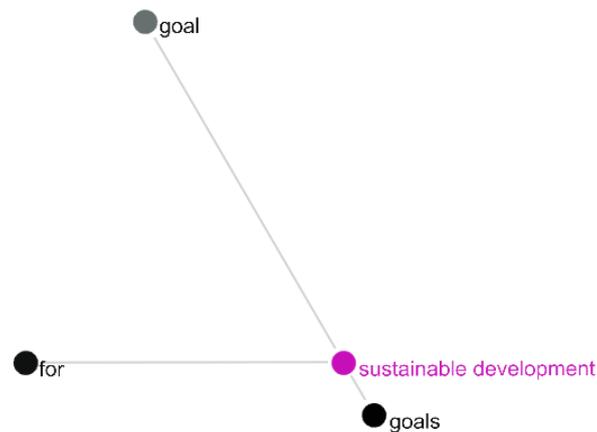
The meaning by collocation gathered for *sustainable development* through Sketch Engine's word sketch is then refined thanks to the collocation networks extracted with #LancsBox.

The collocation networks of sustainable development retrieved with a  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word collocation window can be seen in Table 27. The three collocates' lists endowed with Z-values, frequency of co-occurrence between node and collocate and frequency of occurrence of the sole collocate within the whole corpus can be read in Tables 55, 56, 57 in the Appendix.

<b>Rank</b>	<b>Collocate (1L-1R)</b>	<b>Collocate (3L-3R)</b>	<b>Collocate (5L-5R)</b>
1	<i>goals</i>	<i>goals</i>	<i>goals</i>
2	<i>for</i>	<i>for</i>	<i>for</i>
3	<i>goal</i>	<i>partnership</i>	<i>the</i>
4		<i>the</i>	<i>peace</i>
5		<i>innovation</i>	<i>achieving</i>
6		<i>goal</i>	<i>partnership</i>
7		<i>on</i>	<i>and</i>
8		<i>global</i>	<i>17</i>
9		<i>targets</i>	<i>targets</i>
10		<i>agenda</i>	<i>of</i>
11		<i>we</i>	<i>be</i>
12		<i>be</i>	<i>innovation</i>
13		<i>and</i>	<i>goal</i>
14			<i>we</i>
15			<i>on</i>
16			<i>to</i>
17			<i>that</i>
18			<i>policies</i>
19			<i>agenda</i>
20			<i>poverty</i>
21			<i>challenges</i>
22			<i>an</i>
23			<i>global</i>
24			<i>technology</i>
25			<i>a</i>
26			<i>recognize</i>
27			<i>are</i>
28			<i>in</i>
29			<i>will</i>
30			<i>relevant</i>
31			<i>including</i>

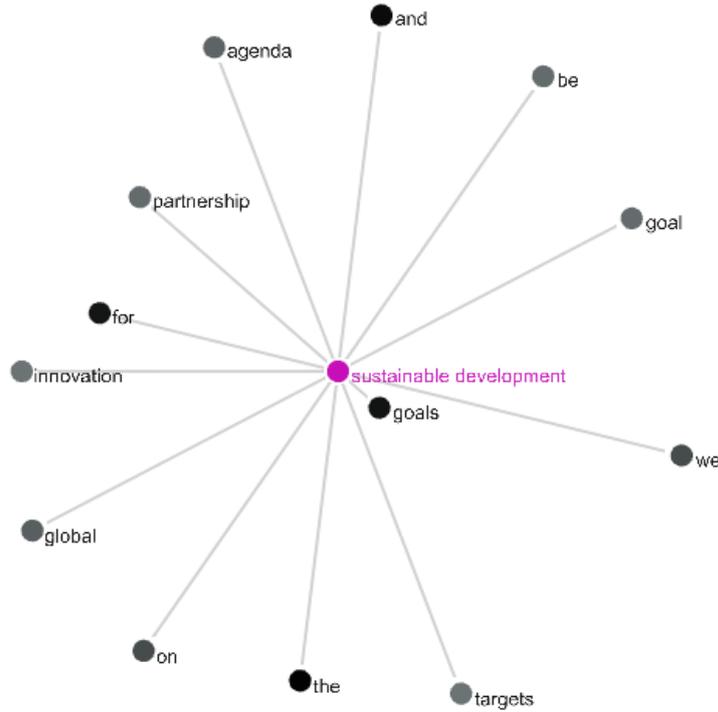
**Table 27.** List of collocates of *SUSTAINABLE DEVELOPMENT* in the 2030 Agenda Corpus (English).

Within a collocation window of one word to the left and one word to the right, the collocates of *sustainable development* are the content words *goal* and *goals* (on the right) and the function word *for* (on the left), as it can be seen in Table 27 and in Figure 20.



**Figure 20.** Collocation network of *SUSTAINABLE DEVELOPMENT* in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.

Widening the collocation window from one to three words to the left and to the right of *sustainable development*, the number of content and of function words included in the collocates' list increases, as it can be seen in Table 27 and in Figure 21. The content words that co-occur with *sustainable development* within this span are *partnership*, *innovation*, *global*, *targets*, *agenda*, and *be*, in addition to *goals* and *goal*, which collocate with *sustainable development* both with a  $\pm 1$ - and with a  $\pm 3$ -word spans. *goals* is the strongest collocate, followed by the function word *for*, which is the second strongest collocate of this set. The content words *targets* and *be* usually collocate with *sustainable development* on the right, while *partnership*, *innovation*, *global*, and *agenda* preferably co-occur on the left-hand side of the node. The function words collocating with *sustainable development* within three words to the left and three words to the right of the node are *the*, *on*, *we*, and *and*, added to the already mentioned *for*, which appears also in the collocation network found with a  $\pm 1$ -word span. *on* and *the* collocate with *sustainable development* on its left-hand side, whereas *and* and *we* preferably appear on the node's right-hand side.



**Figure 21.** Collocation network of *SUSTAINABLE DEVELOPMENT* in the 2030 Agenda Corpus (English) – Z(10.0), 3L-3R, NC5-C5.

As it can be seen in Table 27 and in Figure 22, the content words collocating with *sustainable development* between five words to the left and five words to the right of the node and absent in the previous collocation networks are *peace, achieving, policies, poverty, challenges, technology, recognize, and relevant*. *peace, achieving, challenges, technology* tend to appear to the left of *sustainable development*; *policies, recognize, relevant* usually occur to the right of the node; *poverty* is found both on the left and on the right. These add to the previously mentioned *agenda, be, global, goal, goals, innovation, partnership, and targets*. The function words collocating with *sustainable development* within this collocation window and not within the  $\pm 1$ - and  $\pm 3$ -word spans are 17, *of, we, to, that, an, a, are, in, will, and including*. 17, *of, to, and an* tend to collocate on the left-hand side of the node; *that, in, will, and including* tend to collocate on the right-hand side of the node; *a* and *are* are equally found both on the left and on the right of *sustainable development*. The function words *and, for, on, the, we* collocate with *sustainable development* also within the current a collocation window.



cohesive plurality. In the second case, the pre-modifying *sustainable development goal* stands for *the achievement of sustainable development goals* and it adds information on what is measured by the *indicators*. In the remaining three instances, the lexeme *goal* can be hardly defined as a collocate of *sustainable development* because it appears in the sentence that follows the one including the node.

In both *sustainable development goals* and *sustainable development goal*, *sustainable development* functions as a noun-modifier. Within a  $\pm 1$ -word span, *sustainable development* functions as a noun-modifier in 44% of its occurrences. Similar collocational patterns can be observed also with a  $\pm 3$ -word span.

When widening the collocation window to three words to the left and three words to the right of *sustainable development*, the nouns that *sustainable development* collocates with are *partnership*, *innovation*, *targets*, and *agenda*, in addition to *goals* and *goal*.

The co-occurrence of *sustainable development* and *partnership* shows that, according to the agenda, the enterprise of sustainable development needs to be undertaken through a revitalized mutual commitment. Sustainability, in fact, can be advanced thanks to initiatives like a *global Partnership for Sustainable Development* ( $f=8$ ).

*sustainable development* co-occurs with *innovation* in the prepositional phrase *on science, technology and innovation* that is modified by the prepositional phrase *for the Sustainable Development Goals* ( $f=6$ ). Global mutual commitment in favour of sustainable development can involve teams of experts that research and act to guarantee *innovation*, but also scientific and technological growth. Even *creativity* is associated with *innovation* ( $f=1$ ).

Advances in science, technology, innovation, and creativity can also help to achieve the *targets* set by the 2030 Agenda. The collocate *targets* refers to 169 points associated with the sustainable development goals: in the agenda, every goal is operationalized in sub-goals that help clarify the specific issues that need to be tackled to advance towards sustainability. *targets* contribute to the meaning of *sustainable development* by being frequently used as head and sole constituent of noun phrases coordinated with the noun phrase *goals*, and modified altogether by *sustainable development*. The noun phrase *goals and targets* appears in the agenda 28 times, with *targets* occurring 41 times overall. The pattern is introduced by *the* in 10 cases, confirming that goals and targets are treated by the resolution as a whole set of renown scopes. In the other instances they are described as *new* (in *new goals and targets*,  $f=2$ ), *ambitious* (in *ambitious goals and targets*,  $f=1$ ), *transformative* ( $f=1$ ) and *universal* (in

*universal and transformative goals and targets* or simply in *universal goals and targets*, 3 occurrences overall). The goals and targets set by the agenda are acknowledged to be a novel and ambitious approach to change the current state of affairs for the achievement of sustainability. Their extension is declared to be universal, as with the *global sustainable development*.

According to the collocational pattern that can be observed for *sustainable development* within this span, these goals and targets should be aimed at within the framework of the *agenda*. In two instances, *sustainable development* and *agenda* collocate in the pattern *The Agenda, including the Sustainable Development Goals, [...]*, which describe the SDGs as a part of the agenda.

Within five words to the left and five words to the right, *sustainable development* collocates with the nouns *peace, policies, poverty, challenges, and technology* in addition to the aforementioned *agenda, goal, goals, innovation, partnership, and targets*.

The collocation of *sustainable development* and *peace* is particularly strong: *peace*, in fact, is ranked fourth in the list of the collocates of the node within this span. The semantics of this relationship is one of necessary co-existence because of the two lexical items being coordinated by the conjunction *and* in *durable peace and sustainable development* ( $f=2$ ). This coordination makes them a noun phrases and a conceptual indivisible whole that modifies the noun *achievement*. According to the agenda, sustainable development and peace cannot be reached independently but they need to be worked for contemporarily. In terms of the distribution of salient topics, however, the position of *sustainable development* compared to *durable peace* might suggest that sustainable development is assigned conceptual priority thanks to a lower givenness. In the information structure of Standard Average European, in fact, usually what comes first is already given information while what comes next is new for the reader or the listener (Heine and Kuteva 2006). Thus, I would suggest that since *sustainable development* is mentioned after *peace*, it is treated as a novel concept whose realization depends on the actualization of peace.

According to the 2030 Agenda, sustainable development can be achieved only through the respect of *policies*. *policies* are mainly found in the collocational patterns *policies for poverty eradication and sustainable development* ( $f=2$ ), *of sustainable development policies* ( $f=2$ ), or *policies for sustainable development* ( $f=1$ ).

*sustainable development* and *poverty* co-occur in patterns like *poverty eradication and sustainable development* ( $f=3$ ). *sustainable development* and *poverty* are usually found together with verbs belonging to the semantic field of finishing (e.g. *to eradicate, to end*),

as they are classified in Levin (1993), with *poverty* being the goal of material processes whereby somebody should cease poverty. The actor of this process of ceasing is often unspecified. In addition, *poverty eradication and sustainable development* is associated with material processes like *implement* (f=2) and *promoting* (f=1).

The co-occurrence of *sustainable development* and *challenges* hints at the problems and issues that need to be solved when aiming at sustainable development. In other terms, the appearance of *challenges* casts a doubt on the ease with which the path towards sustainable development can be walked by the actors involved in implementing sustainable practices. *sustainable development* and *challenges* collocate most frequently in the pattern *challenges to achieve sustainable development* (f=2). The co-occurrence of *sustainable development* and *challenges* can be found also in sequences which are grammatically different compared to this collocational pattern, but which display a very similar function. These sequences are *challenges to sustainable development* (f=1), *challenges in its pursuit of sustainable development* (f=1), and *to solving sustainable development challenges* (f=1). As in *challenges to achieve sustainable development*, also in these cases sustainable development can be regarded as the goal of a material process of achieving. This experience is encoded with qualifiers taking the form of prepositional phrases (e.g. *to sustainable development* or *in its pursuit of sustainable development*) or of a noun-modifying noun phrase (as in the latest usage of *sustainable development with challenges*).

*sustainable development* collocates with *technology* together with *science* and *innovation* in the sequences *multi-stakeholder forum on science, technology and innovation for the Sustainable Development Goals* (f=4) or *inter-agency task team on science, technology and innovation for the Sustainable Development Goals* (f=2). Thus, *technology* is paired to *science* and *innovation* as part of a global enterprise that is asked to *promote coordination, coherence and cooperation within the United Nations system on science, technology and innovation-related matters*.

When collocating nouns are concerned, the most frequent colligational patterns that tend to precede the lexical item *sustainable development* are the following:

(DET)	(ADJ)	N	<i>and</i>	( <i>the</i> )	<i>sustainable development</i>	(N)
(N)	(P ( <i>for</i> ))	N	N	<i>and</i>	<i>sustainable development</i>	

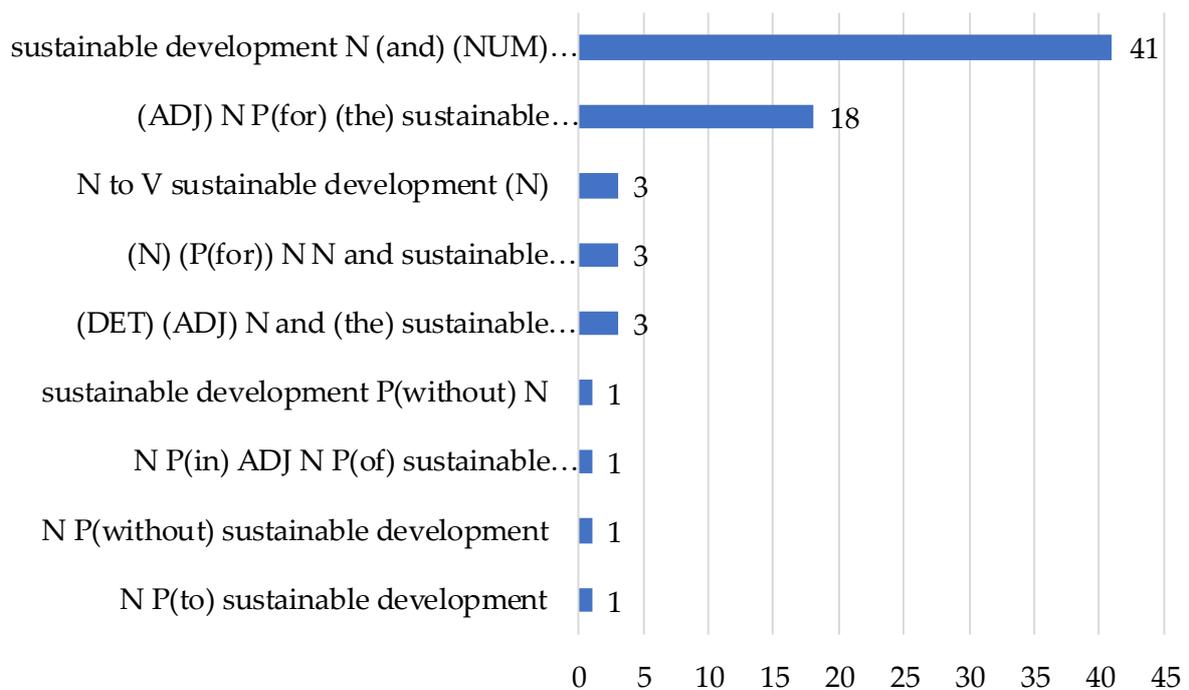
			N	P (to)	<i>sustainable development</i>	
			N	P (without)	<i>sustainable development</i>	
N	P (in)	ADJ	N	P (of)	<i>sustainable development</i>	
	(ADJ)	N	P (for)	(the)	<i>sustainable development</i>	
		N	to	V	<i>sustainable development</i>	N

The colligational patterns that follow the lexical item *sustainable development* are the following:

<i>sustainable development</i>	N	(and)	(NUM)	(N)
<i>sustainable development</i>	P (without)	N		

In these colligational patterns, NUM stands for 'numeral'.

These colligational patterns are reproduced with their frequency in the corpus in Graph 12.



**Graph 12.** The colligational patterns of *SUSTAINABLE DEVELOPMENT* and nouns in the 2030 Agenda (English).

The most frequent colligational pattern of *sustainable development* involves the lexical item in compounds, with *sustainable development* modifying a single noun (e.g. in *sustainable development goals*), or with *sustainable development* modifying a noun phrase (e.g. in *sustainable development goals and 169 targets*). The second most frequent colligational pattern of *sustainable development* consists of a noun followed by a prepositional phrase whose modifier includes *sustainable development* (e.g. in *policies for sustainable development*).

The N slots of all these colligational patterns are occupied by nouns that show a semantic preference for concepts related to the 2030 Agenda (*goals, targets, goal, agenda*), for political conditions involving collaboration (*peace, partnership*), for human products (*innovation, policies, technology*) and for difficulties of any kind (*poverty, challenges*).

**Adjectives.** With a collocation window of three words to the left and three words to the right, the only adjective that *sustainable development* collocates with is *global*. The co-occurrence of *sustainable development* with the adjective *global* suggests that the enterprise of reaching *sustainable development* is international, as in *Global Partnership for Sustainable Development* ( $f=7$ ), in *new global Sustainable Development Goals* ( $f=1$ ), or in *Global Sustainable Development Report* ( $f=1$ ). The adjective *global* appears most frequently with the lexeme *Partnership*; in this case, it modifies *sustainable development* only accidentally, being referred to *Partnership*. The adjective directly modifies *sustainable development* in the collocational patterns *global Sustainable Development Report* and by *new global Sustainable Development Goals*.

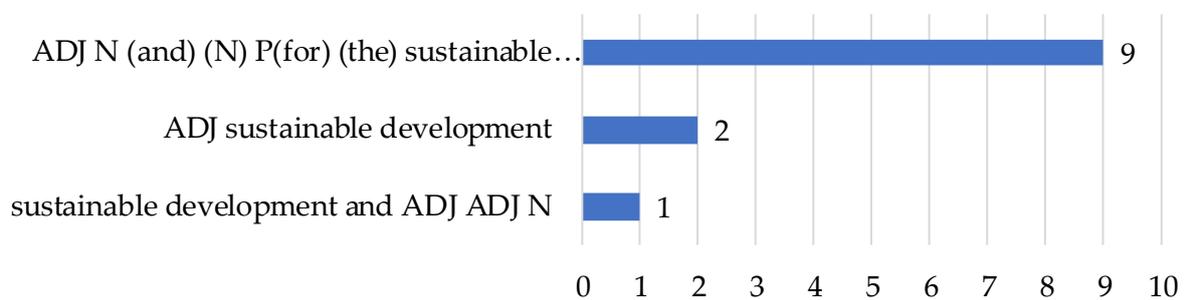
Within a  $\pm 5$ -word span, *sustainable development* collocates only with the adjective *relevant*, apart from the aforementioned *global*. The adjective *relevant* does not directly modify *sustainable development* but it is associated with noun phrases that are grammatically bound to the node. For instance, *sustainable development* is linked to *relevant ongoing processes in the economic, social and environmental fields* ( $f=1$ ). *relevant* qualifies *ongoing processes* that are taking place in the fields of economy, society, and environment for the sake of *sustainable development*. Furthermore, in order for *sustainable development* to be reached, it should also be assured that *relevant information and awareness* ( $f=1$ ) are diffused worldwide on the topic and that *relevant international rules and commitments* ( $f=1$ ) are respected by countries in spite of their autonomous attitude towards sustainability. At the same time, not only are countries

summoned for this challenge but also *relevant stakeholders*, who should *actively contribute in their area of expertise* ( $f=1$ ); these relevant stakeholders are also asked to assist in the *development, transfer and dissemination of relevant technologies* ( $f=1$ ).

*sustainable development* collocates with adjectives in the following colligational patterns:

ADJ N (*and*) (N) P (*for*) (*the*) *sustainable development*  
 ADJ *sustainable development*  
*sustainable development* *and* ADJ ADJ N

The frequency of occurrence of these colligational patterns is reproduced in Graph 13.



**Graph 13.** The colligational patterns of *SUSTAINABLE DEVELOPMENT* and adjectives in the 2030 Agenda (English).

The most frequent colligation involving *sustainable development* and adjectives witnesses the adjectives modifying a noun in a noun phrase that is itself modified by a prepositional phrase introduced by the preposition *for*.

This colligational pattern and the other two identified in the 2030 Agenda (English) are filled with adjectives that signal a semantic preference for internationality (*global*) and for importance (*relevant*).

**Verbs.** Within three words to the left and three words to the right of the node, the sole verb co-occurring with *sustainable development* is *be*. The verb *be* functions both as an auxiliary ( $f=6$ ) and as an existential verb ( $f=2$ ) when it collocates with *sustainable*

*development*. When it is used as an existential verb, *be* describes sustainability in relation to *peace*, as in the following sentence:

*There can be no sustainable development without peace and no peace without sustainable development.*

Sustainable development and peace are thus said to be mutually responsible for each other's existence. This is stressed by the syntactic structure of the sentence. In *no sustainable development without peace and no peace without sustainable development*, in fact, two parallel *no... without...* structures are coordinated. The place of the negated lexeme is taken on first by *sustainable development* and then by *peace*; the slot introduced by *without* is filled first by *peace* and then by *sustainable development*. The unshakable bond between *sustainable development* and *peace* is evident also when *be* is used as an auxiliary. In these cases, the realisation of sustainable practices is subordinated to the achievement of *peace* and *security*. The two lexemes, in fact, constitute a complex noun phrase and they belong to a prepositional phrase (*without peace and security*) that modifies the clause unfolding from the process *be realized*.

Broadening the collocation window to five words to the left and five words to the right, the only lexical verb forms that *sustainable development* collocates with are *achieving* and *recognize*, added to the previously identified *be*. The co-occurrence of *sustainable development* and *achieving*, i.e. the present participle of the verb ACHIEVE, sets sustainability as a goal to be reached. The actor of the process of *achieving* is the international community involved in reflecting and acting in favour of sustainability. The international community is linguistically recalled with the personal pronoun *we*. The personal pronoun refers to a general group of people and institutions that are required to take part in the endeavour of achieving sustainable development and *sustainable development* functions as the goal of the material process of achieving.

The verb form *recognize* seems to be less relevant for the meaning by collocation of *sustainable development* because it appears in another sentence in 80% of the cases. The only meaningful occurrence, however, confirms that sustainability is part of a web of interconnected endeavours: in this occurrence, the cognitive mental process *recognize* points at a phenomenon whereby sustainable development is linked to actions taking place in economic, social and environmental dimensions. This is extremely meaningful since the very core of the 2030 Agenda lies on the idea that sustainable development tackles economic, social and environmental issues.

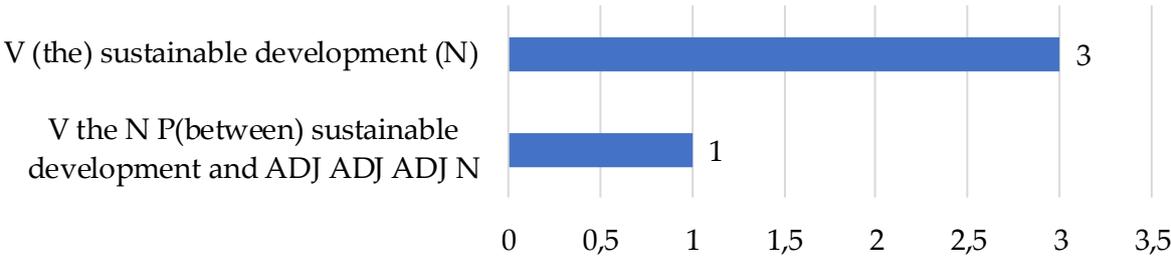
As far as auxiliary verbs are concerned, within five words to the left and five words to the right, *sustainable development* collocates also with the verb forms *are* and *will* in addition to *be*. *sustainable development* usually collocates with *are* in relational processes of the attributive intensive type (9 instances out of 10). *are* is the verbal component of these relational processes and *sustainable development* usually appears in the noun phrase that functions as an attribute. In other cases, *sustainable development* is a modifier of the noun phrase that functions as a carrier or it belongs to the carrier itself. Through this collocational pattern, sustainable development is associated with interconnected and inseparable actions that play a paramount role towards sustainability. *are* functions also as an auxiliary of other kinds of processes in one instance (i.e. *We are announcing today 17 Sustainable Development Goals*).

The collocation between *sustainable development* and *will* is looser than others within the group of auxiliary verbs. The auxiliary *will*, in fact, frequently occurs in the sentence following the one that *sustainable development* belongs to (in 5 out of 11 instances). When it belongs to the same sentence as *sustainable development*, however, it conveys an idea of future. This future meaning is applied to verbs like *continue* (f=1), *foster* (f=1), *promote* (f=1), that prove how the commitment for sustainable development does not only belong to the present but also to the future.

The lexical verbs co-occur with *sustainable development* in the following colligational patterns:

V	(the)		<i>sustainable development</i>		(N)				
V	the	N	P (between)	<i>sustainable development</i>	and	ADJ	ADJ	ADJ	N

The frequency of these colligational patterns can be seen in Graph 14.



**Graph 14.** The colligational patterns of *SUSTAINABLE DEVELOPMENT* and lexical verbs in the 2030 Agenda (English).

These colligational patterns are filled with verbs pointing at a semantic preference for the material process of achieving (*achieving*) and for the mental process of understanding and admitting (*recognize*).

**Prepositions.** Within one word to the left and one word to the right, *sustainable development* collocates with the preposition *for* in the prepositional phrase *for sustainable development*. This prepositional phrase functions either as a postmodifier ( $f=18$ , among which *2030 Agenda for Sustainable Development* or *Global Partnership for Sustainable Development*) or as a circumstantial complement ( $f=8$ ). In all cases, *for* encloses *sustainable development* in a prepositional phrase that reproduces a goal, an aim to be reached: sustainable development is the target of entities expressed through nouns like *Partnership* ( $f=7$ ) or *Agenda* ( $f=4$ ) but also of actions like enhancing or promoting something related to sustainability. Enlarging the collocation window to three words to the left and three words to the right of *sustainable development*, *for* appears in combination with the node in the same contexts that have been described for the  $\pm 1$ -word span, but also in the pattern *for the Sustainable Development Goals* ( $f=7$ ). *sustainable development* is directly preceded by the preposition when the node is used as an independent noun phrase; it is preceded also by a definite article when the node functions as a noun-modifier (as in *for the Sustainable Development Goals*). Within the  $\pm 5$ -word span, the proposition *for* belongs in prepositional phrases that function as goals for verbs or nouns like in the previous instances. However, additional occurrences of *for* pair *sustainable development* with *poverty eradication* ( $f=2$ ) and *inclusive growth* ( $f=1$ ). *for* is also the head of two prepositional phrases whose modifier is a nominalised verb conveying the idea of achieving (i.e. *implementation*,  $f=1$ , and *realization*,  $f=1$ ). In these cases, *sustainable development* functions as the goal of the material processes of *implementing* and *realizing*; however, it does not belong to a finite sentence, but it is posited in a prepositional phrase that modifies the noun phrases *implementation* and *realization*, in *for the realization of the Sustainable Development Goals* ( $f=1$ ) and *for the implementation of the Sustainable Development Goals* ( $f=1$ ). Moreover, in a single case, the preposition *for* does not modify *sustainable development* but it makes the monitoring of sustainable development impacts fundamental for the sake of *sustainable tourism*.

With a  $\pm 3$ -word span, *sustainable development* collocates with the preposition *on* in addition to the already discussed *for*.

The preposition *on* belongs to prepositional phrases including *sustainable development* with the function of postmodifier. Unlike the prepositional phrases constructed with *for* as head, prepositional phrases originating from *on* function as the matter of nominalized verbs like *report* ( $f=1$ ) or of nouns like *assembly* ( $f=2$ ) or *summit* ( $f=1$ ). When the prepositional phrase functions as a postmodifier it usually follows a noun referring to an official meeting or programme (e.g. *Assembly*,  $f=2$ , *forum*,  $f=2$ , *Group*,  $f=2$ , *progress*,  $f=2$ , *Conference*,  $f=1$ , *report*,  $f=1$ , and *Summit*,  $f=1$ ). When the prepositional phrase functions as a complement it completes verbs like *DECIDE* ( $f=1$ ). Sustainability is consequently discursively constructed as a matter worth debating.

Within the  $\pm 5$ -word span, *sustainable development* collocates also with the prepositions *of*, *to*, and *in*.

*sustainable development* meaningfully collocates with the preposition *of* when the preposition precedes it. When the co-occurring preposition follows *sustainable development*, it usually belongs to another sentence and therefore it is not relevant for the meaning by collocation of the node. The nouns modified by the prepositional phrases with *of* as their head can be common nouns (i.e. *dimensions*,  $f=3$ , *nature*,  $f=1$ ) or, more frequently, nominalizations (i.e. *implementation*,  $f=3$ , *achievement*,  $f=2$ , *pursuit*,  $f=2$ , *coordination*,  $f=1$ , *enabler*,  $f=1$ , *enablers*,  $f=1$ , *promotion*,  $f=1$ , *realization*,  $f=1$ , and *translation*,  $f=1$ ). Nominalisation turns verbs into nouns and by doing so it might conceal the actors of the related processes. This proves particularly helpful for ideologically charged discourses. In the case of nominalized verbs modified by a prepositional phrase including *sustainable development*, sustainable development is written to be the goal of material processes of achieving without specifying the agents of the processes. In 12 cases, *sustainable development* does not belong to an *of* prepositional phrase, but it belongs to another prepositional phrase modifying the *of* phrase. In these cases, the collocate *of* is believed to be less meaningful for the semantics of the node.

When collocating with *sustainable development*, the preposition *to* often introduces an infinitive or a participle, whose process features *sustainable development* as goal. The verb forms appearing in these collocational patterns are *achieve* ( $f=4$ ), *achieving* ( $f=2$ ), *implement* ( $f=1$ ), *monitor* ( $f=1$ ), *promote* ( $f=1$ ), *solving* ( $f=1$ ), *support* ( $f=1$ ). They are all material processes. Most of them are transformative (*achieve*, *achieving*, *monitor*, *promote*, *solving*, *support*) and one of them is creative (*implement*). They all imply an action of reaching to or of controlling advances towards sustainability. The same semantic preference is shared also by the prepositional phrases *to our common pursuit of sustainable development* ( $f=1$ ) and *to the promotion of sustainable development*

( $f=1$ ). In these prepositional phrases the nominalized forms *pursuit* and *promotion* witness *sustainable development* as a goal within a phrase-modifying prepositional phrase whose head is *of*. These nominalized forms reproduce material processes of transformation and, together with the previous instances, they show that sustainability is a stage of development that should be encouraged and aimed at although it also needs monitoring and mending in case conditions are not right for it to happen properly. In other occurrences, *to* introduces a prepositional phrase that retains a literal or figurative spatial content of path (as in *road to sustainable development*,  $f=1$ , or *transition to the Sustainable Development Goals*,  $f=1$ ). Through these patterns, the 2030 Agenda portrays the world's engagement towards sustainable development as a journey that needs to be taken from a stage where development is not sustainable for peoples and societies to one in which growth is sustainable. The start of the journey is not marked by any spatial indication. On the contrary, the end of the journey is. This end corresponds to a fulfilled sustainable development. During this journey, accidents might occur, and they may disrupt the connection between the start and the end of the race. The world's nations are asked to *monitor* and *solve* these disruptions in order to achieve their goal, namely sustainability. In this panorama, *to* also builds prepositional phrases including *sustainable development* and modifying nouns like *challenges* ( $f=1$ ; in *We are meeting at a time of immense challenges to sustainable development*) and *contribution* ( $f=1$ ; in *appreciation of cultural diversity and of culture's contribution to sustainable development*).

*sustainable development* and the preposition *in* co-occur in contexts that are significant for the meaning by collocation of the node when *in* is the head of prepositional phrases that clarify the scope of actions taken in favour of sustainability (17 occurrences). This can involve circumstantial prepositional phrases with a specific function of location, matter, etc. In two cases, *in* belongs to the complex preposition *in order to*. One of the occurrences of this preposition introduces an indefinite sentence that features *sustainable development* as the goal of a material process. Also in this case, sustainable development is a goal to be aimed at, to be achieved and the attitude for this achievement is positive.

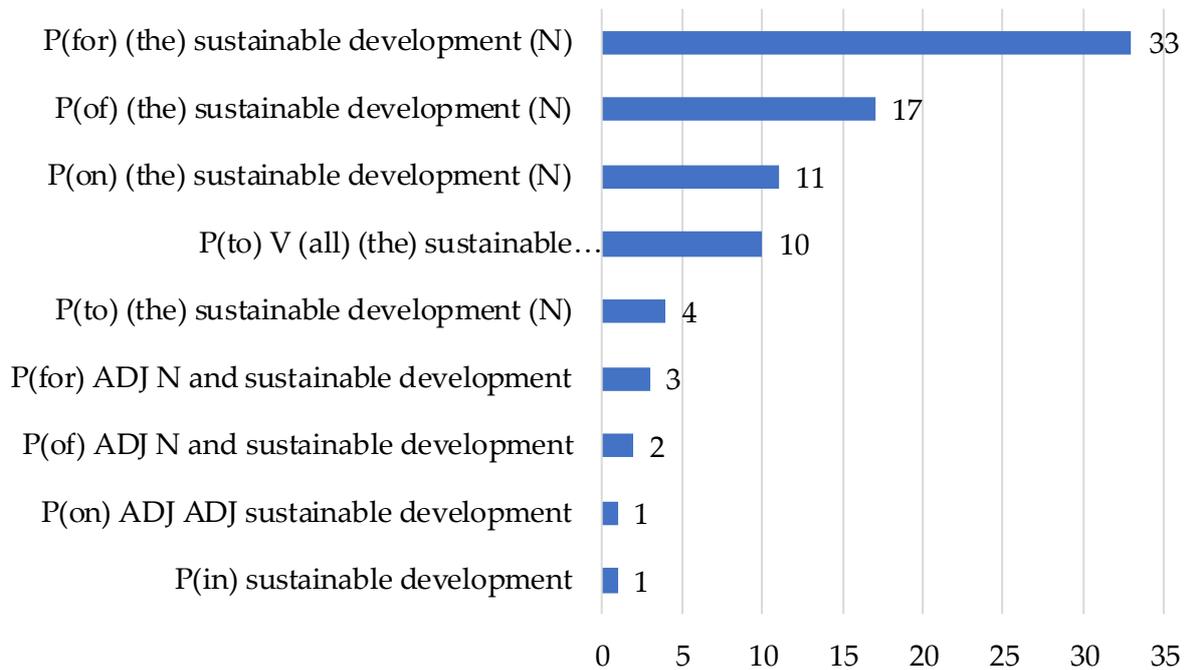
In addition, also the collocate *including* plays an important role in shaping the meaning of *sustainable development*. *including* is a non-finite verb form that functions as a relational process and that is used as a preposition. *including* can explicitly associate the SDGs to the Agenda through the patterns *including the Sustainable Development Goals* ( $f=2$ ) or *including achieving the Sustainable Development Goals* ( $f=1$ ). In addition,

*including* explains what sustainable development consists of and it allows to present some of the issues related to sustainability. Some of these are *sustained and inclusive economic growth, social development, environmental protection and the eradication of poverty and hunger* (f=1). At the same time, *including* suggests that *the knowledge and skills needed to promote sustainable development* should establish a relationship with *education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development* (f=1). It also explains that *international support* to developing countries should be meant as *North-South, South-South and triangular cooperation* (f=1). *including* tells that the *means of implementing sustainable development* are believed to be an essential part of the progress towards the effectiveness of the Agenda and of the SDGs (f=1). Finally, it highlights that *reviews of progress on the Sustainable Development Goals* monitor *cross-cutting issues* to tackle new and urgent problems as soon as they appear (f=1). Therefore, this non-finite verb used as a preposition clarifies the fields touched upon by sustainable development and it creates a semantic bridge between the node and social issues like *education, gender equality, human rights, poverty, and social development*. Also economic and environmental issues are hinted at but in a more limited way.

The collocational patterns involving *sustainable development* and prepositions can be schematised according to the following colligational patterns:

			P ( <i>for</i> )	( <i>the</i> )	<i>sustainable development</i>	(N)
			P ( <i>of</i> )	( <i>the</i> )	<i>sustainable development</i>	(N)
			P ( <i>on</i> )	( <i>the</i> )	<i>sustainable development</i>	(N)
			P ( <i>to</i> )	( <i>the</i> )	<i>sustainable development</i>	(N)
				P ( <i>in</i> )	<i>sustainable development</i>	
P ( <i>for</i> )	ADJ	N		<i>and</i>	<i>sustainable development</i>	
P ( <i>of</i> )	ADJ	N		<i>and</i>	<i>sustainable development</i>	
	P ( <i>on</i> )	ADJ	ADJ		<i>sustainable development</i>	
P ( <i>to</i> )	V	( <i>all</i> )	( <i>the</i> )		<i>sustainable development</i>	(N)

These colligational patterns occur in the 2030 Agenda (English) with the frequencies shown in Graph 15.



**Graph 15.** The colligational patterns of *SUSTAINABLE DEVELOPMENT* and prepositions in the 2030 Agenda (English).

The most frequent colligational patterns involving *sustainable development* and a preposition are prepositional phrases introduced by the prepositions *for*, *of* and *on* and followed by the sole noun phrase *sustainable development* or by a noun phrase modified by the determiner *the*, and having the noun phrase *sustainable development goals* as its head.

**Conjunctions.** Within three words to the left and three words to the right, *sustainable development* co-occurs with the conjunction *and*. The conjunction *and* is generally used to pair noun or prepositional phrases including *sustainable development* with other noun and prepositional phrases. Consequently, *and* extends the lexico-semantic relations of the node by introducing a positive addition. When *and* precedes *sustainable development*, the conjunction associates the node to lexical units like *poverty eradication* ( $f=3$ ), *durable peace* ( $f=2$ ), *human potential* ( $f=1$ ), *inclusive growth* ( $f=1$ ), *national development* ( $f=1$ ), and *Sustainable Development Goals* ( $f=1$ ). When it follows *sustainable development*,

the conjunction associates the node to lexical units like *lifestyles in harmony with nature* (f=1), *other relevant ongoing processes* (f=1), and *sustainable lifestyles* (f=1). These lexical units aid in discursively constructing sustainable development as an endeavour to be pursued within the goals and targets set by the agenda. Moreover, they bind sustainability with a kind of national and international growth that cannot neglect the fundamental need of eradicating poverty. Such national and international growth must take into account the potential of human beings and of their lifestyles but also the relationship between humans and nature. In 6 occurrences out of 40, however, the conjunction *and* occurs in the sentence preceding or following the one where *sustainable development* appears and it cannot be considered as a proper collocate. Within the  $\pm 5$  span, the conjunction *and* links noun phrases that are modified in some way by a prepositional phrase including *sustainable development*. For example, *and* links *global challenge* and *indispensable requirement* (f=2), which are together considered to be basic starting points towards sustainability. Other similar pairings are: *science, technology and innovation* (f=6); *oceans, seas and marine resources* (f=2); *peaceful societies and inclusive societies* (f=2); *interlinkages and integrated nature* (of the SDGs, f=1); *sustained economic growth and inclusive economic growth* (f=1); *knowledge and skills* (f=1); *relevant information and relevant awareness* (f=1); *develop and implement* (f=1); *non-discriminatory laws and non-discriminatory policies* (f=1); *policy coherence and enabling environment* (f=1); *creativity and innovation* (f=1); *universal, integrated and interrelated nature and three dimensions* (social, economic and environmental, f=1); *Inter-Agency group and Expert Group* (f=1); *system-wide coherence and system-wide coordination* (f=1). These pairs group elements that share semantic relations, being them of hyponymy (e.g. *seas* and *oceans*), of near-synonymy (e.g. *develop* and *implement*). All these pairings create a picture of sustainable development as a stage of development that ought to be worked for with the means of integrated, well-pondered and coherent scientific and technological innovation that could be profitable for societies and the environment. Cooperation and communal engagement stand out as a recurrent semantic preference in these combinations of noun phrases. They contribute to a positive semantic prosody since the ecological framework employed for the analysis recommends unity and mutual assistance.

Within five words to the left and five words to the right, *sustainable development* collocates with the conjunction *that*. In half of the cases, *that* belongs to the sentence that follows the one in which *sustainable development* can be found. In the remaining occurrences, *that* either functions as a subordinative conjunction, or as a relative pronoun. When it functions as a subordinative conjunction, it follows verbs expressing

mental processes like RECOGNIZE ( $f=1$ ) and REITERATE ( $f=1$ ). These mental processes respectively feature *sustainable development* and *we* as sensors. While *we* is naturally associated with mental activities, the use of *sustainable development* as a sensor of a mental process might be regarded as a form of grammatical metaphor. In addition, the recurrent appearance of these mental processes adds a category to the semantic preference of the lexeme: not only is sustainable development associated with material processes of achievement, but also to mental processes of believing and repeating. When *that* is used as a relative pronoun, it modifies the nouns *sustainable tourism* ( $f=1$ ), *benefits* ( $f=1$ ), *partnerships* ( $f=1$ ), *measurements of progress* ( $f=1$ ).

**Determiners.** Within three words to the left and three words to the right, *sustainable development* collocates with the determiner *the*. *sustainable development* is usually employed without determiners, but it is preceded by the definite article *the* in 23% of its occurrences. In all cases when *sustainable development* is introduced by the determiner *the*, the lexical item is modified by *goals* in the noun phrase *sustainable development goals*. When *sustainable development* is modified by *goals*, the determiner *the* precedes *sustainable development goals* in 76% of its occurrences. The SDGs, in fact, are introduced by the definite article since the very beginning of the document. This makes them a concept whose knowledge is presupposed by the authors. *the* can also specify not the very *sustainable development* but a lexeme that is related to it, as in *the road to sustainable development* ( $f=2$ ) or *the path towards sustainable development* ( $f=1$ ). These two instances represent sustainable development as the arriving point of a journey. The route of this journey is well-known, as it is suggested by the definite article introducing the noun *road*. The determiner *the* is also used before the nouns *link* ( $f=1$ ) and *promotion* ( $f=1$ ); in these cases, it associates sustainability to *other relevant ongoing processes in the economic, social and environmental fields* ( $f=1$ ). In addition, the definite article is used to introduce the *Global Sustainable Development Report* ( $f=1$ ).

Within five words to the left and five words to the right, *sustainable development* collocates also with the determiners *a* and *an*. While the definite article *the* ranks third in the list of the collocates of *sustainable development*, the two forms of the indefinite article (*a* and *an*) rank twenty-second and twenty-fifth respectively and they total overall 18 co-occurrences with the node, whereas *the* is seen to collocate with it 84 times. This might suggest that the sentences including *sustainable development* have a high degree of facticity.

**Pronouns.** *sustainable development* collocates with a single pronoun (*we*), which emerges in the collocates' list with a  $\pm 3$ -word span. When co-occurring with *sustainable development*, the pronoun *we* is almost always found in another sentence (in 86% of the cases). It stands for the group of politicians and representatives that participated in the preliminary meetings and in the discussion for the elaboration of the 2030 Agenda. These politicians and representatives play the role of first-person plural actors that commit themselves to a concrete action in favour of sustainability. They commit themselves through declarative sentences. It is interesting to observe that, once *sustainable development* is mentioned, the following sentence is often shaped as a declarative that engages political actors in some sort of obligation to sustainability. The same collocational tendencies can be observed within the  $\pm 5$ -word span.

**Numerals.** *sustainable development* co-occurs with a single numeral (17), which appears for the first time with a  $\pm 5$ -word span. The numeral 17 refers to Goal 17 (i.e. *Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development*) in 3 cases and more generally to the 17 Sustainable Development Goals in the remaining 2 cases. This collocate echoes the collocational tendency that was already suggested for *goals* and *goal*, namely the semantic preference for lexemes that have to do with the 2030 Agenda.

Summing up, the closest collocates of *sustainable development* shape its meaning by associating it to the semantics of actions aimed at the achievement of a goal, both in the case of the more explicit co-occurrence with *goals* and *goal* and in the case of the less explicit co-occurrence with *for*. Sustainable development, in fact, is acted for through various goals but, at the same time, it is itself an overarching goal. Looking further, the lexemes collocating with *sustainable development* contribute to shape the meaning of the lexical item as a common, worldwide enterprise that should be worked for through innovation by tackling not only the broader goals but also the more specific targets set by the 2030 Agenda. The collocational patterns of *sustainable development* suggest that sustainable development is a well-known concept, worked for in an international dimension that involves political forces and laymen, and that it is advanced towards thanks to its being the goal of action and thanks to its being talked of. In the end, the collocational patterns of *sustainable development* construct the semantic profile of the lexical item in the following way: sustainable development is a condition that needs to be achieved within the realm of the United Nations' 2030

Agenda as a sum of goals and targets; these goals and targets have to be dealt with both in theory and in practice through international cooperation; they impact on social issues like poverty eradication and peace and they require universal measures to be taken for the innovative improvement of technology and other enabling resources; this endeavour is challenging and it compels the global institutions to meaningfully rule and monitor development.

This semantic profile is first contributed to by the colligational patterns of *sustainable development*. *sustainable development* is most frequently included in noun phrases that modify prepositional phrases; however, noun phrases including *sustainable development* can be also the direct object of transitive verbs performing material or mental processes.

These colligational patterns are filled with content and function words that can be clustered according to a precise semantic preference. The semantic preference of *sustainable development* includes material processes of achieving and mental processes of understanding used either as verbs or as nominalizations (e.g. *achieving* and *recognize*) as well as lexemes bound to the 2030 Agenda (e.g. *agenda*, *goals*, and *targets*) and to international social matters (e.g. *global*, *partnership*, and *poverty*).

*sustainable development* tends to be surrounded by a lexical aura that can be positively appraised from the point of view of the ecological framework adopted in the present work. The lexical item is surrounded by lexemes that convey an idea of positive and firm engagement towards sustainability for the wellbeing of human societies, and even the colligational patterns of the lexical item stress that sustainable development is a condition that is strongly craved for. From a discursive point of view, this lexical and grammatical aura plays a part in knitting discourse so that the 2030 Agenda turns out to be a sequence of positive recommendations or solutions to identified problems.

This positive evaluative prosody might seem to be stained by negatively connoted collocates like *poverty* and *challenges*. On the contrary, the patterns that tie *sustainable development* and these two collocates prove positive in spite of the lexemes' semantic negativity. The evaluative prosody triggered by the combination of *sustainable development* with *poverty eradication* is a positive one because it involves the end of an undesired condition. At the same time, the presence of the negatively connoted *challenges* among the most salient collocates of *sustainable development* shows that not only is sustainable development associated with a positive aura of active

improvement, but it also proves a challenge for the global community. Thus, global community needs to overcome its limits and cooperate towards sustainability.

#### 4.4.2. The Hungarian *FENNTARTHATÓ FEJLŐDÉS*

The study of the meaning by collocation of *sustainable development* in the 2030 Agenda Corpus is enriched with the extraction of the collocational patterns of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’.

The Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ appears 104 times in the Hungarian version of the 2030 Agenda, roughly the same number of times of its English equivalent (see § 4.4.1).

##### 4.4.2.1. Word sketch

The meaning by collocation of the Hungarian noun phrase *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ is first portrayed through the lexical item’s word sketch (Figure 23).

determiners before "fenntartható fejlődés"	"fenntartható fejlődés" 's ...	nouns modified by "fenntartható fejlődés"
a ...	megvalósítás ...	érdek ...
	érdek ...	megvalósítás ...
		kihívás ...

**Figure 23.** Word sketch of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – LogDice(6.0), NC5-C5.

The word sketch of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ showcases the main traits of the lexical item’s semantics. As it can be seen from the first column, the determiner that precedes the lexeme the most is the definite article *a* ‘the’. Sustainable development, in fact, is written about as a well-known concept. And as the nouns modified by *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ show, sustainable development is primarily meant as an enterprise triggering a challenge (*KIHÍVÁS* ‘challenge’) or as a goal to be achieved (*MEGVALÓSÍTÁS* ‘realization’) or at least to be

aimed at (*ÉRDEK* ‘interest’, although the actual form used is *érdekében* ‘for’). These patterns equal the one that have been detailed for English (see § 4.4.2).

The challenge represented by the lexeme *KIHÍVÁS* can come from difficulties raised by the search for sustainable development, but it can also be the outcome of issues opposing the progress towards sustainability.

Sustainable development appears as a challenge or it is frequently associated with challenges. However, it is also a condition that is wished for and that should be committed to. This is stressed by the nominal, possessive pattern *A FENNTARTHATÓ FEJLŐDÉS MEGVALÓSÍTÁSA* ‘achieve sustainable development’ ( $f=8$ ). The pattern consists of the noun phrase *fenntartható fejlődés*, functioning as singular possessor, and by the noun phrase *a [...] megvalósítása*, literally ‘the achievement of’. In *megvalósítása* the noun *megvalósítás* ‘achievement’ is followed by the third person singular possessive suffix *-a* that marks the entity possessed in the pattern. This possessive pattern is found in its nominative form (i.e. *a fenntartható fejlődés megvalósítása*,  $f=3$ ) but it can also be modified by suffixes whose semantics is at least partially one of goal or of target. So, for example, in *a fenntartható fejlődés megvalósítására* ( $f=2$ ), where the noun phrase *a fenntartható fejlődés megvalósítása* ‘achieve sustainable development’ is turned into an indirect object through the sublative suffix *-ra*, or in *a fenntartható fejlődés megvalósításához* ( $f=1$ ), where the same happens with the allative suffix *-hoz*, the achievement of sustainable development is grammatically constructed as a goal to be reached. The pattern can also be used as the possessor in a possessive construction and, in this case, it is marked by the dative case *-nak* (in *a fenntartható fejlődés megvalósításának*, literally in this case ‘the achievement of sustainable development’,  $f=2$ ). A similar goal function applies also to the recurrent pattern *a fenntartható fejlődés érdekében* ‘for sustainable development’. While in *A FENNTARTHATÓ FEJLŐDÉS MEGVALÓSÍTÁSA* ‘achieve sustainable development’ sustainable development is clearly regarded as a goal, in *a fenntartható fejlődés érdekében* ‘for sustainable development’ this meaning is enshrined in the function word *érdekében* ‘for’.

#### 4.4.2.2. Collocation networks

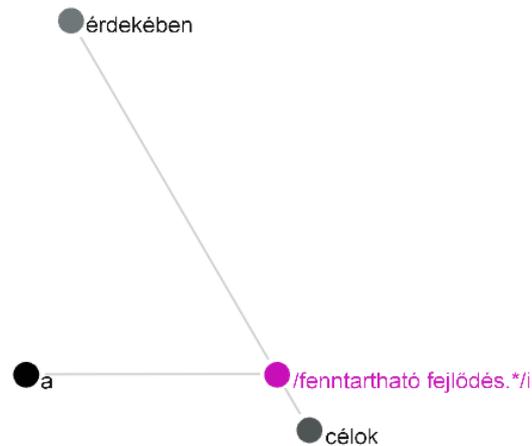
The meaning by collocation outlined through Sketch Engine’s word sketch is then boosted by analysing the collocation networks of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ with the #LancsBox software.

The collocation networks found for the Hungarian lexical item with the  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word spans are shown in Table 28. The collocates' lists can be seen in Tables 58, 59, 60 in the Appendix together with the collocates' position in relation to the node, their Z-value, their frequency of occurrence together with the node, and their frequency of occurrence in general in the 2030 Agenda (Hungarian).

Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>célok</i> 'goals'	<i>célok</i> 'goals'	<i>a</i> 'the'
2	<i>a</i> 'the'	<i>a</i> 'the'	<i>célok</i> 'goals'
3	<i>érdekében</i> 'for'	<i>elérését</i> 'achieve'	<i>elérését</i> 'achieve'
4		<i>támogató</i> 'enabling'	<i>megújított</i> 'revitalized'
5		<i>érdekében</i> 'for'	<i>nélkülözhetetlen</i> 'essential'
6		<i>béke</i> 'peace'	<i>nélkül</i> 'without'
7		<i>foglalkozó</i> '-'	<i>támogató</i> 'enabling'
8		<i>szakpolitikák</i> 'policies'	<i>és</i> 'and'
9		<i>és</i> 'and'	<i>béke</i> 'peace'
10		<i>globális</i> 'global'	<i>foglalkozó</i> '-'
11			<i>szegénység</i> 'poverty'
12			<i>érdekében</i> 'for'
13			<i>partnerség</i> 'partnership'
14			<i>globális</i> 'global'
15			<i>szakpolitikák</i> 'policies'
16			<i>beleértve</i> 'including'
17			<i>is</i> 'too'
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			

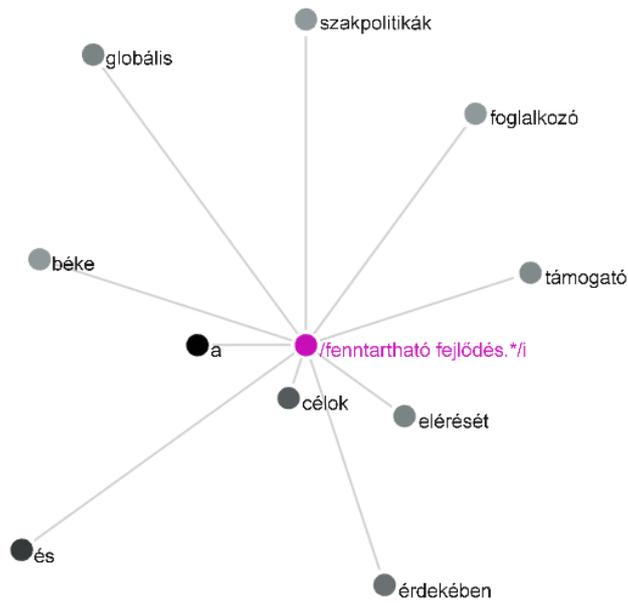
**Table 28.** List of collocates of *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' in the 2030 Agenda Corpus (Hungarian).

As it can be seen in Figure 24, within one word to the left and one word to the right, *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ collocates with the content words *célok* ‘goals’ and *érdekében* ‘for’, and with the function word *a* ‘the’. The content words *célok* ‘goals’ and *érdekében* ‘for’ tend to stand on the right-hand side of the node, while the function word *a* ‘the’ tends to occur on its left-hand side.



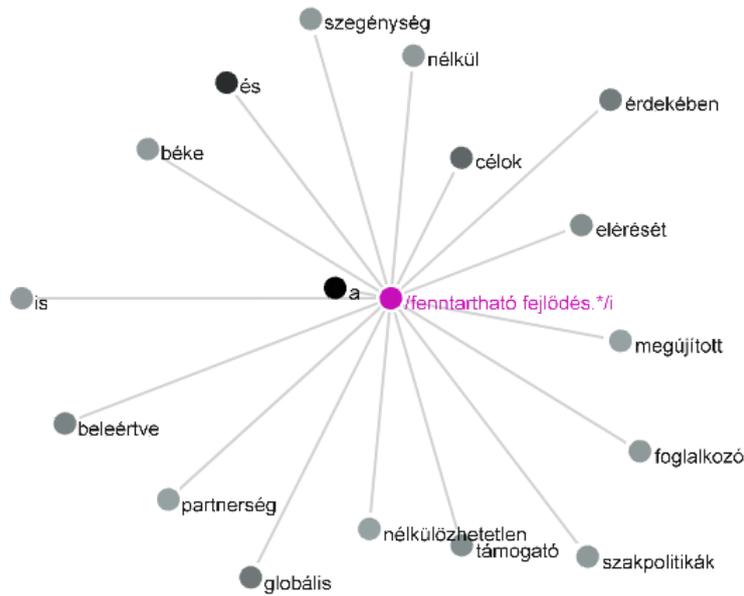
**Figure 24.** Collocation network of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.

Within three words to the left and three words to the right, *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ collocates with the content words *elérését* ‘achieve’, *támogató* ‘enabling’, *béke* ‘peace’, *foglalkozó* ‘-’, *szakpolitikák* ‘policies’, *globális* ‘global’ and with the function words *és* ‘and’ in addition to the already mentioned content and function words (Figure 25). The collocates *elérését* ‘achieve’, *támogató* ‘enabling’, *foglalkozó* ‘-’, *szakpolitikák* ‘policies’, *globális* ‘global’ tend to appear on the node’s left-hand side, whereas the collocates *béke* ‘peace’ and *és* ‘and’ tend to stand on its right-hand side.



**Figure 25.** Collocation network of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.

As it can be seen in Figure 26, within five words to the left and five words to the right, the collocation network of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ is complemented by other collocates in addition to the aforementioned. *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ collocates with the content words *megújított* ‘revitalized’, *nélkülözhetetlen* ‘essential’, *szegénység* ‘poverty’, *partnerség* ‘partnership’ and with the function words *nélkül* ‘without’, *beleértve* ‘including’, *is* ‘too’. The collocates *beleértve* ‘including’ and *is* ‘too’ tend to occur on the node’s left-hand side, while the collocates *megújított* ‘revitalized’, *nélkülözhetetlen* ‘essential’, *nélkül* ‘without’, *partnerség* ‘partnership’ tend to stand on its right-hand side. *szegénység* ‘poverty’ can be equally found on the left- and on the right-hand side of the node.



**Figure 26.** Collocation network of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 5L-5R, NC5-C5.

Comparing the collocation network extracted for the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ (Table 28) and for the English *sustainable development* (Table 27), it can be noticed that with a  $\pm 1$ -word span the collocational patterns of the two lexical items share the collocates *goals-célok* and *for-érdekében*. The lexemes are used in comparable ways in the two languages.

With the  $\pm 1$ -word span, the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ collocates with *célok* ‘goals’ only when goals are mentioned in their plural form (*célok*, consisting of the noun stem *cél* ‘goal’ and of the plural suffix *-ok*) in the phrase *Fenntartható Fejlődési Célok* ‘Sustainable Development Goals’ ( $f=18$ ).

The other shared collocate, namely *érdekében* ‘for’, consists of the noun stem *érdek* ‘interest’, followed by the suffix *-e* that indicates a singular possession held by a third person singular possessor; this suffix is followed by the inessive suffix *-ben* ‘in’. *érdekében* literally means ‘in the interest of [sth]’ but it is used in this context with a similar function to the English preposition *for*.

Within one word to the left and one word to the right, in Hungarian the other collocate of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ is the definite article *a* ‘the’, which appears in English with a span of  $\pm 2$  words. The determiner precedes *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ in 83 instances out of 104 (80% of the cases) and it posits thus the expression as a given, known concept. In addition, only 8 of the remaining 21 occurrences of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable

development’ are not introduced by a definite article: in the other 13 cases, the definite article precedes a noun phrase or an adjective phrase that modifies the node on the left.

Within three words to the left and three words to the right of the node, the collocational patterns of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ and of its English equivalent *sustainable development* share the collocates *the-a, and-és* and *global-globális* in addition to the already mentioned *goals-célok* and *for-érdekében*. These lexemes are used in a comparable way in the English and in the Hungarian versions of the resolution.

In addition, with a  $\pm 3$ -word span the collocation network of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ reveals in advance some of the content words that emerge in the English collocation networks only with larger spans (i.e. *elérését* ‘achieve’, *béke* ‘peace’, and *szakpolitikák* ‘policies’).

*elérését* and *béke* respectively correspond to the English *achieving* and *peace*, which are collocates of *sustainable development* with a  $\pm 5$ -word span. *szakpolitikák* corresponds to the English *policies*, which co-occurs with *sustainable development* within five words to the left and five words to the right of the node. The use of these collocates in the Hungarian document corresponds to the one described in detail for English.

The only content words that appear exclusively in the Hungarian collocates’ list are *támogató* ‘enabling’ and *foglalkozó* ‘-’. In 6 out of 7 co-occurrences of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ with *támogató* ‘enabling’, *támogató* is found together with *elérését* ‘achieve’ in the adjective phrase *Fenntartható Fejlődési Célok elérését támogató* ‘for the Sustainable Development Goals’, which can be analysed as follows

<i>Fenntartható-Ø</i> sustainable-NOM	<i>Fejlődési-Ø</i> developmental-NOM	<i>Cél-ok-Ø</i> goal-PL-NOM
<i>elérés-Ø-ét</i> achievement-SG-POSS.3SG-ACC		<i>támogató-Ø</i> enabling-NOM

‘for the sustainable development goals’

The present participle *támogató* ‘enabling’, which functions as an adjective, is the head of the adjective phrase. *támogató* is constituted by the verb stem *támogat-* ‘enable’ and by the suffix *-ó*, which obtains present participles from verb stems with back-vowel roots. In this case, the semantics of the present participle is completed by the noun

phrase *Fenntartható Fejlődési Célok elérését* ( $f=6$ ), literally ‘achievement of the Sustainable Development Goals’. This noun phrase consists in a possessive pattern where the singular possessed entity *elérés* ‘achievement’ is marked by the third person singular possessive suffix *-e-* and by the accusative suffix *-t*. The possessive suffix recalls the third person plural possessor *Fenntartható Fejlődési Célok* ‘Sustainable Development Goals’, which behaves as a third person singular possessor for structural reasons. This adjective phrase keeps together the SDGs and the urge and the support to achieve them. This urge and support for the achievement of the SDGs is written to characterise also the *több érdekcsoportot* ‘multi-stakeholder’ ( $f=3$ ), *ENSZ-munkacsoport* ‘United Nations inter-agency task team’ ( $f=2$ ), and *releváns technológiák* ‘relevant technologies’ ( $f=1$ ). In other words, the UN’s task team and other multi-stakeholders are summoned to back the achievement of the SDGs; this achievement needs also to rely on essential conditions for sustainability to be met (like relevant technologies).

The semantics of supporting the pursuit of sustainable development is contributed also by those collocational patterns where *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ collocates with the adjectival present participle *foglalkozó*. *foglalkozó* comes from the verb stem *foglalkoz-* ‘deal with’ and from the suffix *-ó*, which derives present participles from back-vowel verbs roots, as in the case of the aforementioned *támogató*. The verb *foglalkozik* ‘deal with’ (i.e. the dictionary entry of the verb stem *foglalkoz-*) is a material process that requires a patient expressed with the comitative suffix *-val/vel* for its meaning to be completed. *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ collocates with *foglalkozó* in two significant patterns: *fenntartható fejlődéssel foglalkozó*, literally ‘dealing with sustainable development’ ( $f=2$ ), and *Fenntartható Fejlődési Célokkal foglalkozó*, literally ‘dealing with the Sustainable Development Goals’ ( $f=2$ ). The first present participle modifies the noun phrase *Nyílt Munkacsoport* ‘Open Working Group’ ( $f=2$ ), while the second is associated with the noun phrase *Magas Szintű Politikai Fórum* ‘high-level political forum’ ( $f=2$ ). These noun phrases are in the position of actors of the material process of ‘dealing with’ and, thus, they present two actors engaged for sustainable development.

Within five words to the left and five words to the right, the collocation networks of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ and of the English *sustainable development* share the collocates *partnership-partnerség* (nominative singular in Hungarian), *poverty-szegénység* (nominative singular in Hungarian), and *including-beleértve* in addition to the ones that have been already

pointed at. These words are used in the same contexts in the two versions of the document.

The only differences that can be observed between the English and the Hungarian occurrences are of morphosyntactic nature: Hungarian prefers nominalized verbs to the finite or indefinite verb forms used in English in the corresponding cases, as in

<i>a</i>	<i>Globális-Ø</i>	<i>Partnerség-Ø</i>	<i>erősítés-Ø--e</i>
the;ART.DEF	Global-NOM.SG	Partnership-NOM.SG	enhancement-SG-POSS.3SG

'Enhance the Global Partnership'

or in

<i>a</i>	<i>szegénység-Ø</i>	<i>valamennyi-Ø</i>	<i>formá-Ø-já-nak</i>	<i>felszámolás-Ø-a</i>
the;ART.DEF	poverty-NOM.SG	all-NOM.SG	form-SG-POSS.3SG-DAT	end-SG-POSS.3SG

'end poverty in all its forms'

In the 2030 Agenda (Hungarian), *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' collocates with content and function words that do not appear in the English collocates' list with a  $\pm 5$ -word span. These content words are the adjectives *megújított* 'revitalized' and *nélkülözhetetlen* 'essential' and the function word is the postposition *nélkül* 'without'. The English *revitalized* appears with a  $\pm 10$ -word span and *without* with a  $\pm 6$ -word span.

The adjective *megújított* 'revitalized' modifies *GLOBÁLIS PARTNERSÉG* 'Global Partnership' in the noun phrases *a Fenntartható Fejlődés megújított Globális Partnerségének keretrendszerében* 'within the framework of a revitalized Global Partnership for Sustainable Development' ( $f=2$ ) and *a Fenntartható Fejlődés megújított Globális Partnersége útján* 'through a revitalized Global Partnership for Sustainable development' ( $f=1$ ). The adjective does not modify the very sustainable development and thus it is not directly responsible for the semantics of *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development'. Nonetheless, it adds information on the Global Partnership for Sustainable Development, which contributes to the sustainability enterprise.

On the contrary, *nélkülözhetetlen* 'essential' is strongly bound to *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development'. *FENNTARTHATÓ FEJLŐDÉS* participates in the adjective phrase *a fenntartható fejlődéshez nélkülözhetetlen* 'essential for sustainable

development' ( $f=2$ ) with its variant *nélkülözhetetlen a fenntartható fejlődéshez* ( $f=1$ ). *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' operates as a modifier of the adjective *nélkülözhetetlen* 'essential', which constrains the noun phrase with the allative suffix *-hez*. The difference in word order between *a fenntartható fejlődéshez nélkülözhetetlen* and *nélkülözhetetlen a fenntartható fejlődéshez* depends on pragmatic reasons. The first element to appear, being it either the noun phrase *a fenntartható fejlődéshez* or the adjective *nélkülözhetetlen*, is focused on in discourse. This distinction does not play a role in the meaning of the pattern, though. This adjectival pattern modifies nouns like the singular nominative *követelmény* 'requirement' ( $f=2$ ), the plural accusative *előnyöket* 'benefits' ( $f=1$ ), and the singular nominative *környezet* 'environment' ( $f=1$ ). A certain type of advantages, environment, and requirement are thus claimed to be essential for sustainable development.

The postposition *nélkül* 'without' is always found in clauses that refer to the relationship between sustainable development and peace and it follows the noun phrases *fenntartható fejlődés* 'sustainable development' (in *fenntartható fejlődés nélkül* 'without sustainable development',  $f=2$ ), *béke* 'peace' (in *béke nélkül* 'without peace',  $f=1$ ), and *béke és biztonság* 'peace and safety' (in *béke és biztonság nélkül* 'without peace and safety',  $f=1$ ).

The meaning by collocation of *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' in the Hungarian version of the 2030 Agenda is comparable to the meaning by collocation of *sustainable development* in the English version of the resolution.

This comparability depends first on the equivalent semantic preference of the two lexical items. As in the English case, also the Hungarian *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' collocates with lexemes that explicitly refer to the 2030 Agenda (*célok* 'goals'), with lexemes that encode material processes of achievement (*elérését* 'achieve', *támogató* 'enabling') and international political and social matters (*béke* 'peace', *szegénység* 'poverty', *partnerség* 'partnership', *globális* 'global', *szakpolitikák* 'policies'). Moreover, the Hungarian *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' collocates with qualities of novelty and of paramount importance (*megújított* 'revitalized', *nélkülözhetetlen* 'essential').

The differences between the English and the Hungarian meaning by collocation of *sustainable development* and *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' lie basically in the distinct colligational patterns that reproduce the grammatical relations binding the words that appear in the two lexical items' collocational patterns. The main

colligational patterns that can be identified when *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ is used in its nominative form (i.e. *fenntartható fejlődés*) are several. They can include *fenntartható fejlődés* in noun phrases as in

<i>a</i>	<i>fenntartható fejlődés</i>	ADJ	ADJ	N.POSS3SG.(NOM/DAT)
<i>a</i>	<i>fenntartható fejlődés</i>	ADJ	N.POSS3SG.DAT	N.POSS3SG.SUB
<i>a</i>	<i>fenntartható fejlődés</i>	N.POSS3SG.INE		

or in

<i>a</i>	N.NOM	N.POSS3SG.SUB	és	<i>a</i>	<i>fenntartható fejlődés</i>	N.POSS3SG.(SUB)
<i>a</i>	ADJ	N.NOM	és	<i>a</i>	<i>fenntartható fejlődés</i>	N.POSS3SG.(NOM/DAT)

In these colligational patterns, SUB stands for “sublative” and INE stands for “inessive”.

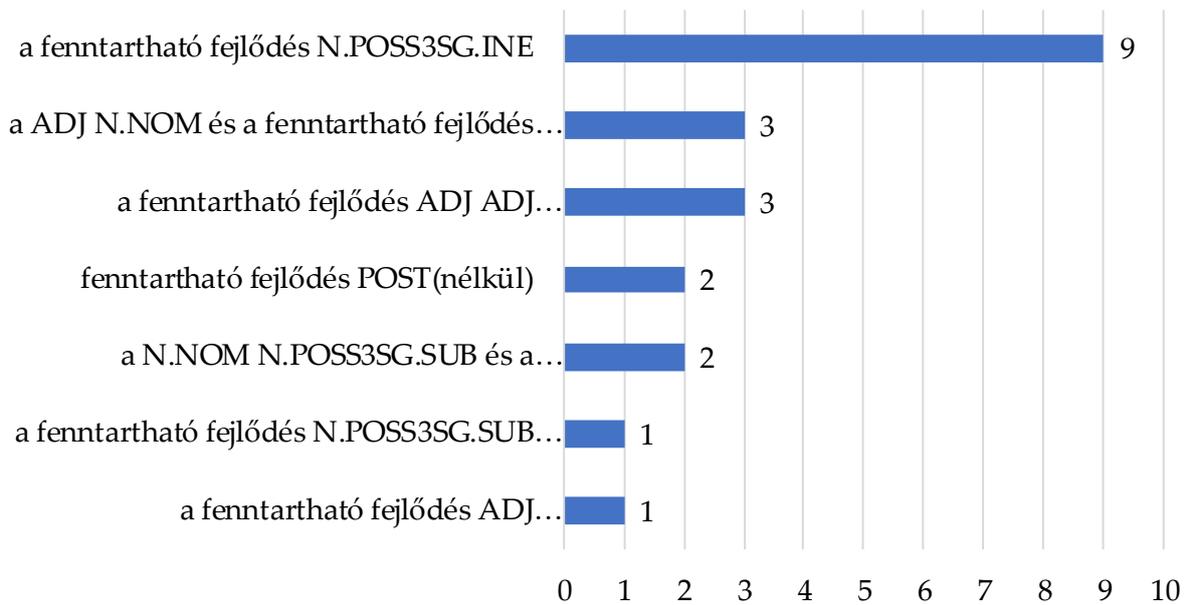
Colligational patterns can contain *fenntartható fejlődés* in adjective phrases like

<i>a</i>	<i>fenntartható fejlődés</i>	N.POSS3SG.SUB	ADJ
----------	------------------------------	---------------	-----

or they can feature *fenntartható fejlődés* in postpositional phrases like

<i>fenntartható fejlődés</i>	POST ( <i>nélkül</i> )
------------------------------	------------------------

These colligational patterns are reproduced together with their frequency in Graph 16.



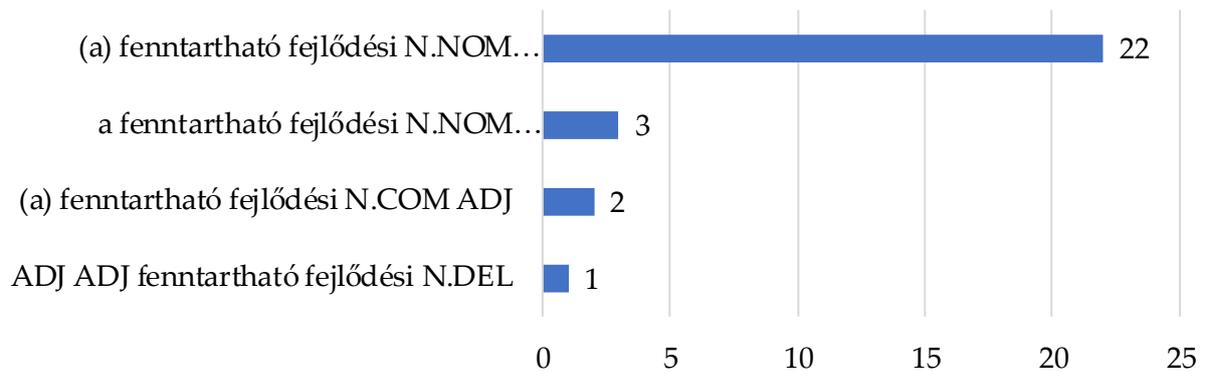
**Graph 16.** The colligational patterns of the nominative *fenntartható fejlődés* ‘sustainable development’ in the 2030 Agenda (Hungarian).

When *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ is employed in adjectival form (i.e. *fenntartható fejlődési* ‘(of) sustainable development’), it is usually found in the following colligational patterns:

- |     |                               |                               |                 |               |
|-----|-------------------------------|-------------------------------|-----------------|---------------|
| (a) | <i>fenntartható fejlődési</i> | N.NOM                         | (N.POSS3SG.ACC) | (ADJ)         |
| (a) | <i>fenntartható fejlődési</i> | N.COM                         | ADJ             |               |
| a   | <i>fenntartható fejlődési</i> | N.NOM                         | N.POSS3SG.NOM   | N.POSS3SG.INE |
| ADJ | ADJ                           | <i>fenntartható fejlődési</i> | N.DEL           |               |

In these colligational patterns, COM stands for “comitative” and DEL stands for “relative”.

The frequency of the colligational patterns involving the adjectival form *fenntartható fejlődési* can be read in Graph 17.



**Graph 17.** The colligational patterns of the adjectival form *fenntartható fejlődési* 'sustainable development' in the 2030 Agenda (Hungarian).

When *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' is used in the allative case (i.e. in the form *fenntartható fejlődéshez* 'for sustainable development') the two colligational patterns are:

*a fenntartható fejlődéshez* ADJ

ADJ *a fenntartható fejlődéshez*

The first colligational pattern occurs three times and the second colligational pattern appears a single time in the 2030 Agenda (Hungarian).

When *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' is employed in the sublative case (i.e. *fenntartható fejlődésre* 'to sustainable development'), the lexical item is included in the colligational patterns

*a fenntartható fejlődésre* ADJ

*a* N.NOM N.POSS3SG.SUB *és a fenntartható fejlődésre*

Both colligational patterns appear a single time.

When *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' is used in the accusative case (i.e. *fenntartható fejlődést* 'sustainable development') it is most significantly found in the colligational pattern

*a fenntartható fejlődést* ADJ

, which appears a single time.

Finally, when *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ is employed in the comitative case (i.e. *fenntartható fejlődéssel* ‘with sustainable development’), it is especially found in the colligational pattern

*a* *fenntartható fejlődéssel* ADJ

, which occurs two times.

Overall, as in the case of the English *sustainable development*, thanks to the semantic preference for positively charged lexemes and thanks to the broad positivity of the colligational patterns involving *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’, which usually treat sustainable development as a goal to be reached, the evaluative prosody of the Hungarian lexical item is positive. The evaluative prosody is positive in ecolinguistic terms: the semantics of *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ make it a condition to be wished for in order to grant a beneficial progress for the environment and its inhabitants.

#### 4.4.3. The Italian *SVILUPPO SOSTENIBILE*

The meaning by collocation of *sustainable development* in the 2030 Agenda Corpus is further expanded on with the examination of the collocational patterns of the Italian *SVILUPPO SOSTENIBILE* ‘sustainable development’.

In the 2030 Agenda (Italian), the lexical item *SVILUPPO SOSTENIBILE* ‘sustainable development’ occurs 106 times. The Italian version of the 2030 Agenda is the only version of the resolution where the lexical item occurs also in the plural form. In Italian this corresponds to *sviluppi sostenibili*, literally ‘sustainable developments’. The plural form of the lexical item *sustainable development* is a typicality of the Italian version of the Agenda, and it appears in the sentence *Ci riuniamo in un periodo di enormi sfide per gli sviluppi sostenibili* ‘We are meeting at a time of immense challenges to sustainable development’. *sviluppi sostenibili* seems to refer to the various developmental enterprises that the world’s countries and citizens need to embark on to sustainably reach the future.

The meaning by collocation of the Italian *SVILUPPO SOSTENIBILE* ‘sustainable development’ (from now on simply *sviluppo sostenibile* because of its invariable form) is gathered through #LancsBox’s collocation networks.

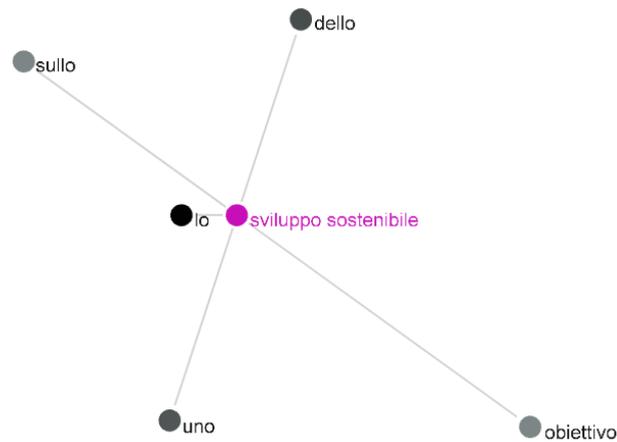
The collocates’ lists extracted for *sviluppo sostenibile* ‘sustainable development’ with a  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word span can be seen in Table 29. These collocates’ lists can be also read in Tables 61, 62 and 63 in the Appendix, together with their position in relation to the node, their Z-value, their frequency of co-occurrence with the node and their frequency of occurrence in the corpus independently of the node.

Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>lo</i> ‘the’	<i>lo</i> ‘the’	<i>lo</i> ‘the’
2	<i>dello</i> ‘of (the)’	<i>obiettivi</i> ‘goals’	<i>obiettivi</i> ‘goals’
3	<i>uno</i> ‘a’	<i>uno</i> ‘a’	<i>uno</i> ‘a’
4	<i>sullo</i> ‘on (the)’	<i>dello</i> ‘of (the)’	<i>raggiungimento</i> ‘achievement’
5	<i>obiettivo</i> ‘goal’	<i>per</i> ‘for’	<i>dello</i> ‘of (the)’
6		<i>degli</i> ‘of (the)’	<i>per</i> ‘for’
7		<i>sullo</i> ‘on (the)’	<i>degli</i> ‘of (the)’
8		<i>obiettivo</i> ‘goal’	<i>sullo</i> ‘on (the)’
9		<i>di</i> ‘of; to’	<i>pace</i> ‘peace’
10		<i>traguardi</i> ‘targets’	17 ‘17’
11			<i>di</i> ‘of; to’
12			<i>il</i> ‘the’
13			<i>e</i> ‘and’
14			<i>traguardi</i> ‘targets’
15			<i>gli</i> ‘the’
16			<i>che</i> ‘that’
17			<i>raggiungere</i> ‘achieve’
18			<i>obiettivo</i> ‘goal’
19			<i>riconosciamo</i> ‘(we) recognise, (we) acknowledge’
20			<i>un</i> ‘a’
21			<i>globale</i> ‘global’
22			<i>a</i> ‘to, in, on’

**Table 29.** List of collocates of *SVILUPPO SOSTENIBILE* ‘sustainable development’ in the 2030 Agenda Corpus (Italian).

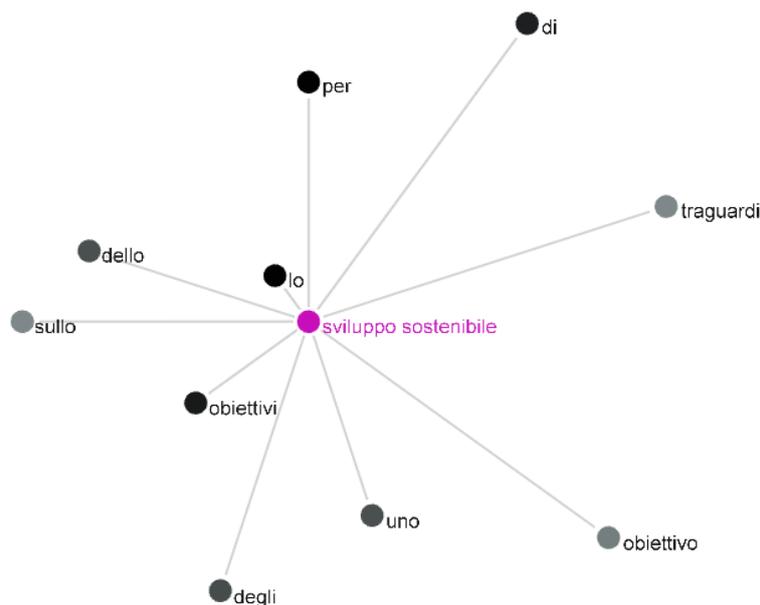
Within one word to the left and one word to the right, *sviluppo sostenibile* collocates with the content word *obiettivo* ‘goal’, and with the function words *lo* ‘the’, *dello* ‘of (the)’, *uno* ‘a’, *sullo* ‘on (the)’ (Figure 27). The collocates *lo* ‘the’, *dello* ‘of (the)’, *uno* ‘a’,

*sullo* ‘on (the)’ tend to stand on the left-hand side of the node, whereas the collocate *obiettivo* ‘goal’ tends to appear on its right-hand side.



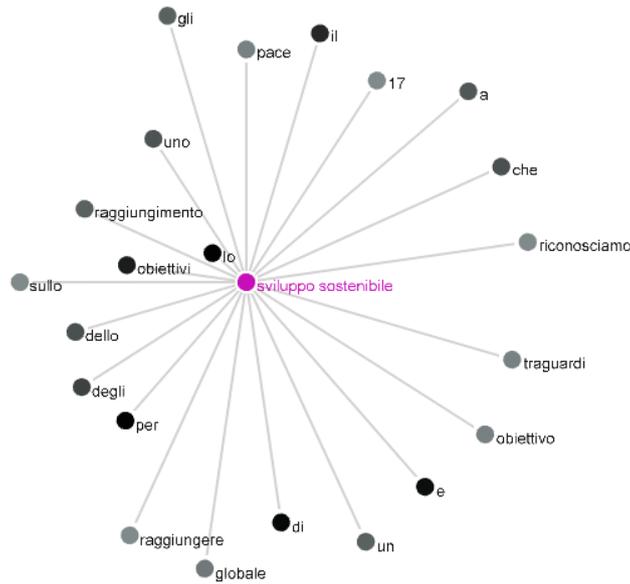
**Figure 27.** Collocation network of *SVILUPPO SOSTENIBILE* ‘sustainable development’ in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.

As it can be seen in Figure 28, within three words to the left and three words to the right, *sviluppo sostenibile* ‘sustainable development’ collocates with the content words *obiettivi* ‘goals’, *traguardi* ‘targets’ and with the function words *per* ‘for’, *degli* ‘of (the)’, *di* ‘of; to’ in addition to the ones already mentioned for the  $\pm 1$ -word span. The collocates *obiettivi* ‘goals’, *per* ‘for’, *degli* ‘of (the)’, *di* ‘of; to’ tend to occur on the left-hand side of the node, while the collocate *traguardi* ‘targets’ tends to occur on its right-hand side.



**Figure 28.** Collocation network of *SVILUPPO SOSTENIBILE* ‘sustainable development’ in the 2030 Agenda Corpus (Italian)– Z(10.0), 3L-3R, NC5-C5.

Within five words to the left and five words to the right, *sviluppo sostenibile* ‘sustainable development’ collocates with the content words *raggiungimento* ‘achievement’, *pace* ‘peace’, *raggiungere* ‘achieve’, *riconosciamo* ‘(we) recognise; (we) acknowledge’, *globale* ‘global’ and with the function words *17* ‘17’, *il* ‘the’, *e* ‘and’, *gli* ‘the’, *che* ‘that’, *un* ‘a’, *a* ‘to; in; on’ (Figure 29). These content and function words appear in addition to the ones already noted for the smaller collocation windows. The collocates *raggiungimento* ‘achievement’, *pace* ‘peace’, *gli* ‘the’, *raggiungere* ‘achieve’, *globale* ‘global’ tend to appear on the left-hand side of the node, the collocates *17* ‘17’, *e* ‘and’, *che* ‘that’, *riconosciamo* ‘(we) recognise; (we) acknowledge’, *un* ‘a’, *a* ‘to; in; on’ tend to stand on its right-hand side. *il* ‘the’ occurs both on the left- and on the right-side of the node.



**Figure 29.** Collocation network of *SVILUPPO SOSTENIBILE* ‘sustainable development’ in the 2030 Agenda Corpus (Italian)– Z(10.0), 5L-5R, NC5-C5.

Within one word to the left and one word to the right, the collocation networks of the Italian *sviluppo sostenibile* ‘sustainable development’ (see Table 29) and the collocation networks of the English *sustainable development* and of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ (see Tables 27 and 28 respectively) share the singular and plural forms of the lemma *GOAL*, represented in English by the word forms *goals* and *goal* and found in Hungarian with the plural noun *célok* ‘goals’ and in Italian with the singular form *obiettivo* ‘goal’. These nouns are used in comparable ways in the three languages.

Nevertheless, the Italian *obiettivo* ‘goal’ cannot be considered as a collocate proper within this collocation window because all its occurrences belong to the sentence that follows the one including *sviluppo sostenibile* ‘sustainable development’.

Apart from *obiettivo* ‘goal’, the collocates of *sviluppo sostenibile* ‘sustainable development’ are either determiners, namely *lo* ‘the’ and *uno* ‘a’, or prepositions, namely *dello* ‘of (the)’ and *sullo* ‘on (the)’.

With a  $\pm 1$ -word span, the definite article *lo* ‘the’ (singular masculine) modifies the noun phrase *sviluppo sostenibile* ‘sustainable development’ in 48 instances. The definiteness of the determiner *lo* ‘the’, however, is conveyed also by the complex prepositions *dello* ‘of (the)’ (singular masculine;  $f=14$ ), *sullo* ‘on (the)’ (singular masculine;  $f=5$ ), *allo* ‘to (the)’ (singular masculine;  $f=2$ ), which precede *sviluppo sostenibile* ‘sustainable development’ 21 times overall. Thus, *sviluppo sostenibile*

'sustainable development' is modified by determiners or prepositions that stress its definiteness in 65% of the cases. The remaining instances of *sviluppo sostenibile* 'sustainable development' are introduced by the indefinite article *uno* 'a' (masculine) in 12 cases and by no determiner or by the prepositions *di* 'of' ( $f=22$ ), *senza* 'without' ( $f=1$ ), and *tra* 'between' ( $f=1$ ) in 25 cases. The Hungarian and the Italian collocates' lists have in common the definite articles *a* 'the' (Hungarian) and *lo* 'the' (Italian). The definite article *the* appears among the collocates of *sustainable development* in the 2030 Agenda (English) with a  $\pm 2$ -word span, while the prepositions emerge starting with a  $\pm 2$ -word span. The difference in the width of the collocation window featuring determiners and prepositions in Italian, English and Hungarian depends on the morphological and syntactic peculiarities of the three languages: while Italian and Hungarian require a more extensive use of the definite article, in English only 23% of the occurrences of *sustainable development* are preceded by a definite article; while Italian and English make extensive use of prepositions, Hungarian encodes the same functions through affixes.

Within three words to the left and three words to the right, the Italian version of the 2030 Agenda shares with the English and Hungarian versions of the resolution the collocates *the-lo-a* and *for-per-érdekében* in addition to the aforementioned English lexemes *goals* and *goal*, that correspond to the sole *célok* 'goals' in Hungarian, and to *obiettivi* 'goals' and *obiettivo* 'goal' in Italian. These collocates are usually employed in comparable contexts in the three versions of the 2030 Agenda.

Among these shared collocates, the Italian preposition *per* 'for' appears for the first time within a collocation window of two words to the left and two words to the right of the node, namely a word farther than its English and Hungarian translational equivalents. This depends on the fact that, in most cases, prepositional phrases having the preposition *per* 'for' as their head and the noun phrase *sviluppo sostenibile* 'sustainable development' as their modifier need the modifier to be preceded by a definite or an indefinite article. The same applies to the Italian article *lo* 'the', which occurs in its Hungarian form (*a* 'the') within one word to the left and one word to the right of the node and in its English form (*the*) with a span of  $\pm 2$  words.

The Italian and the English collocates' lists share also the preposition *on-sullo*, and the noun *targets-traguardi*. Both word pairs are used in comparable ways in the two languages. The translational equivalent of *targets* is absent in the Hungarian list up to a span of  $\pm 6$  words, where it appears in the form *alcélok*. *traguardi* 'targets' refers

to the 169 Targets that the Agenda puts forward together with the 17 Sustainable Development Goals.

Comparing the collocates of *sviluppo sostenibile* ‘sustainable development’ within this collocation window with those already discussed for the  $\pm 1$ -word span, the novel and Italian-exclusive function words are the prepositions *degli* ‘of (the)’ (plural masculine) and *di* ‘of; to’. In English, *sustainable development* collocates with *of* and *to* starting from a  $\pm 4$ -word span. This depends on the different syntactic characteristics of Italian and English. For example, while in Italian noun phrases modify other noun phrases by belonging to prepositional phrases, in English they often feature straightforwardly as modifiers, as in

<i>Obiettivo-i</i>	<i>di</i>	<i>Svilupp-o</i>	<i>Sostenibil-e</i>
Goal-PL.M	of	Development-SG.M	Sustainable-SG.M
‘Sustainable Development Goals’			

Hungarian translational equivalents of *of* and *to* apparently do not occur in the lists of collocates of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’. This happens because the role played by the two function words in English and Italian is morphologically translated in Hungarian. Thus, modification or possession marked by *of* in English and by *di* and its related forms in Italian is encoded in a possessive suffix; the dative function of *to* is mainly rendered with the dative suffix *-nak/nek*; the infinitival function of *to* and *di* is conveyed through the verbal suffix *-ni*.

Within five words to the left and five words to the right, the comparison between the collocates’ list of the Italian *sviluppo sostenibile* ‘sustainable development’, the English *sustainable development*, and the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ adds other shared collocates to the ones already mentioned. Shared content words are: *peace-pace-béke*, *achieving-raggiungimento*, *raggiungere-elérését*, and *global-globale-globális*. Shared function words are: *the-lo,il,gli-a* and *and-e-és*. These content and function words participate in the same contexts in the three versions of the 2030 Agenda.

The Italian and the English collocation networks also share the singular noun *goal-obiettivo*, which appears in Hungarian with a  $\pm 7$ -word span in the form of *cél*, the verb *recognize-riconosciamo* (first person plural of the present indicative), which features in the Hungarian collocates’ list with a span of  $\pm 6$  words and in the form *elismerjük* (first person plural, present indicative, definite conjugation), and the

numeral 17. The nouns *goal-obiettivo-cél* and the verb *recognize-riconosciamo-elismerjük* are used almost exactly in the same contexts. For example, *goal, obiettivo, and cél* are usually employed at the beginning of sentences that follow the English node *sustainable development* and the Italian node *sviluppo sostenibile* ‘sustainable development’ to introduce the 17 goals put forward by the 2030 Agenda. *recognize, riconosciamo, and elismerjük* are usually found in the sentence that follows the one including *sustainable development* and its translational equivalents. The numeral 17 functions in English and Italian either to recall the 17 SDGs or to introduce the seventeenth goal of the Agenda.

The Italian and the English collocation networks also share the relative pronoun *that-che*, which appears among the collocates of the Hungarian *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ with a  $\pm 7$ -word span, the indefinite article *a,an-un*, and the prepositions *to,in,on-a*. Indefinite articles do not feature as collocates in the Hungarian collocation network because they are frequently translated with a zero article. Prepositions are absent because of morphological reasons: Hungarian tends to encode the functions of the Italian and English prepositions through suffixes and postpositions. These function words retain in Italian the same use that has already been described for English.

The only difference can be observed for the translation of the noun phrase *Sustainable Development Goals*, which takes the following forms in Italian: *gli obiettivi per lo sviluppo sostenibile* ( $f=3$ ), *gli obiettivi dello sviluppo sostenibile* ( $f=2$ ), and *gli Obiettivi di Sviluppo Sostenibile* ( $f=2$ ). These definite noun phrases show that the SDGs are a well-known concept for the readers of the 2030 Agenda.

However, while the English and Hungarian collocates’ lists also share the collocates *partnership-partnerség, poverty-szegénység* and *including-beleértve* within the  $\pm 5$ -word span, the corresponding Italian collocates are absent from the collocation networks up to this point. As far as the corresponding Italian collocates are concerned, *sviluppo sostenibile* ‘sustainable development’ collocates with *partnership* ‘partnership’ with a  $\pm 8$ -word span, with *povertà* ‘poverty’ with a  $\pm 7$ -word span, whereas the translational equivalents of *including* do not appear among the closest collocates of the node. The translational equivalents of *including* are adjectives like *COMPRESO* and *INCLUSO* but also conjunctions like *anche*. They are absent from the collocates’ lists of the Italian *sviluppo sostenibile* ‘sustainable development’ because of frequency reasons. In fact, the frequency of every single form of the lemmas *COMPRESO* and *INCLUSO*, which vary according to the number and gender of the noun they modify, and of the conjunction *anche* is lower compared to the frequency of the unchangeable *including*

and *beleértve*. This excludes the several forms of the Italian adjectives and of the conjunction from the collocates' lists of *sviluppo sostenibile* 'sustainable development'. The loanword *partnership* 'partnership' falls only in the  $\pm 8$ -word span because it is not the only translational equivalent of the English *partnership*: the noun *partnership* is translated with *partnership* in 9 cases and with *collaborazione* in the remaining 7 cases. The case of the Italian *povertà* 'poverty' is different: the Italian collocate appears farther from the node compared to English and to Hungarian because of syntactic reasons.

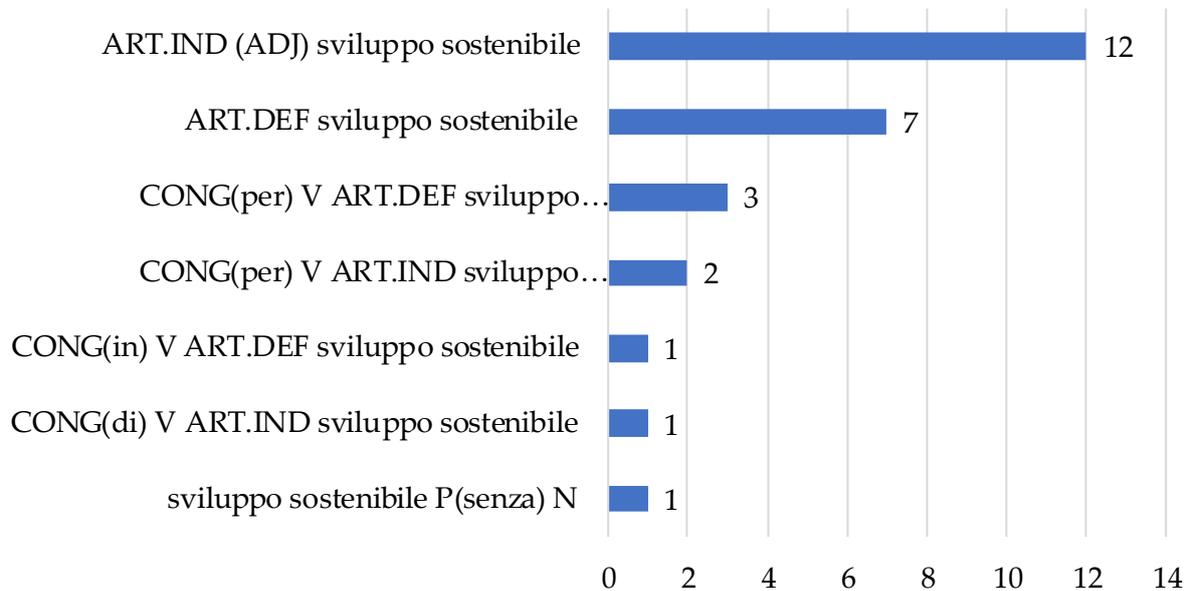
The meaning by collocation of the Italian *sviluppo sostenibile* 'sustainable development' is comparable with the semantics of the English and of the Hungarian lexical items. This comparable meaning is contributed to especially by a similar semantic preference. As it broadly occurs in the English and in the Hungarian cases, the Italian *sviluppo sostenibile* 'sustainable development' experiences a semantic preference for lexemes that concern the UN's 2030 Agenda (*obiettivi* 'goals', *traguardi* 'targets', *obiettivo* 'goal'), for material processes of achievement either represented by verbs or by nouns (*raggiungimento* 'achievement', *raggiungere* 'achieve'), for mental processes of understanding and admitting (*ricogliamo* '(we) recognise; (we) acknowledge') for lexemes that illustrate international political or social conditions (*pace* 'peace', *globale* 'global').

These lexemes complete the most significant colligational patterns of the Italian *sviluppo sostenibile* 'sustainable development', which are the only aspect of meaning by collocation that diverges in the Italian, English and Hungarian cases. The most significant colligational patterns of *sviluppo sostenibile* 'sustainable development' in noun phrases used either in isolation or as modifiers of verb phrases are:

		ART.IND (ADJ)	<i>sviluppo sostenibile</i>		
		ART.DEF	<i>sviluppo sostenibile</i>		
			<i>sviluppo sostenibile</i>	P ( <i>senza</i> )	N
CONG ( <i>per</i> )	V	ART.IND	<i>sviluppo sostenibile</i>		
CONG ( <i>per</i> )	V	ART.DEF	<i>sviluppo sostenibile</i>		
CONG ( <i>di</i> )	V	ART.IND	<i>sviluppo sostenibile</i>		
CONG ( <i>in</i> )	V	ART.DEF	<i>sviluppo sostenibile</i>		

In these colligational patterns, ART.IND stands for “indefinite article”, ART.DEF stands for “definite article” and CONG stands for “conjunction”.

These colligational patterns occur with the frequency shown in Graph 18.



**Graph 18.** The colligational patterns of noun phrases including *SVILUPPO SOSTENIBILE* ‘sustainable development’ in the 2030 Agenda (Italian).

The most frequent colligational pattern involving *sviluppo sostenibile* ‘sustainable development’ in a noun phrase is a noun phrase introduced by an indefinite article and eventually modified by an adjective.

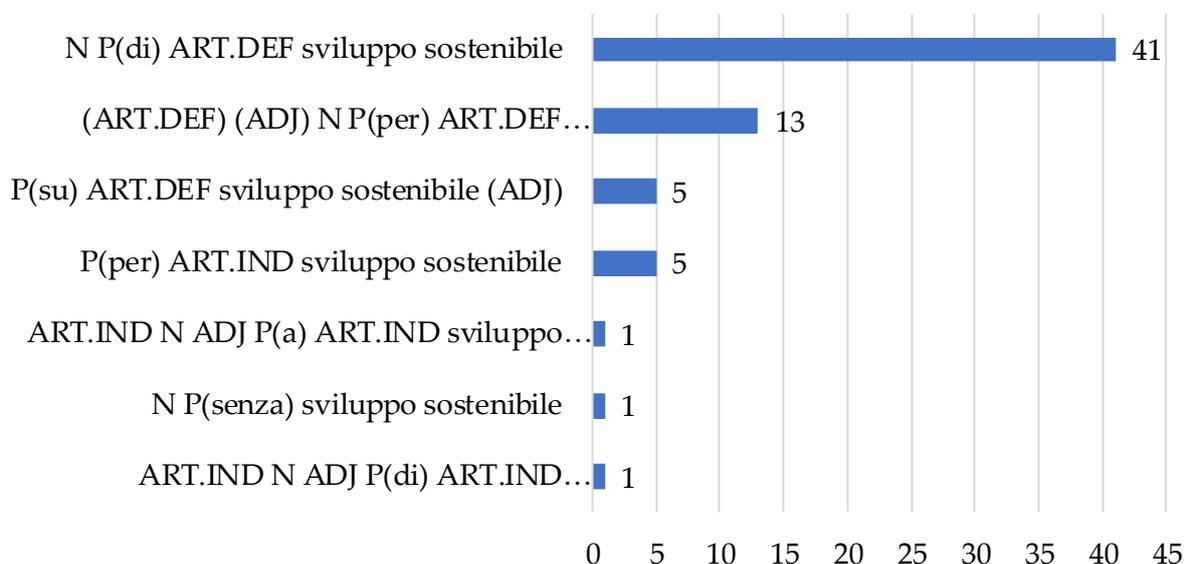
When noun phrases including *sviluppo sostenibile* ‘sustainable development’ modify prepositions in prepositional phrases, the most significant colligational patterns identified in the Italian version of the 2030 Agenda are the following:

(ART.DEF)	(ADJ)	N	P ( <i>per</i> )	ART.DEF	<i>sviluppo sostenibile</i>	
			P ( <i>per</i> )	ART.IND	<i>sviluppo sostenibile</i>	
		N	P ( <i>di</i> )	ART.DEF	<i>sviluppo sostenibile</i>	
ART.IND		N	P ( <i>di</i> )	ART.IND	<i>sviluppo sostenibile</i>	
			P ( <i>su</i> )	ART.DEF	<i>sviluppo sostenibile</i>	(ADJ)

N            P (*senza*)        *sviluppo sostenibile*

ART.IND        N            ADJ    P (*a*)        ART.IND        *sviluppo sostenibile*

The frequency of these colligational patterns can be seen in Graph 19.



**Graph 19.** The colligational patterns of prepositional phrases including *SVILUPPO SOSTENIBILE* ‘sustainable development’ in the 2030 Agenda (Italian).

The most frequent colligational pattern involving *sviluppo sostenibile* ‘sustainable development’ in a prepositional phrase is introduced by the preposition *di* ‘of’ and it is modified by a definite article.

As in the English and Hungarian cases, thanks to the semantic preference and thanks to the colligation displayed by the Italian *sviluppo sostenibile* ‘sustainable development’, the evaluative prosody of the lexical item might be positively evaluated. This means that the semantics of *sviluppo sostenibile* ‘sustainable development’ assigns the lexical item a beneficial role for the construction of equity and balance within the environment according to the ecological framework assumed for the current research.

## 4.5. Discussion

In conclusion, among the cultural keywords and the meanings by collocation of *sustainable, sustainability, sustainable development* identified in the English version of the 2030 Agenda and in its Hungarian and Italian counterparts, both similarities and differences have just been observed. These similarities and differences mainly depend on the morphological and syntactic characteristics of the three languages.

As far as cultural keywords are concerned, differences in word ranking of the most frequent and statistically significant lexemes of the English, Hungarian, and Italian versions of the 2030 Agenda Corpus could be explained by the morphological peculiarities of the three languages, which make them diversely rich in terms of inflectional variability. This inflectional variability tends to increase from English, an isolating language, to Hungarian, an agglutinative language, and Italian, a fusive language. For example, variation in the number of function words represented in the Italian frequency lists might depend on the high inflectional variation of the Italian function words; Italian variable function words, in fact, discriminate gender and number and this increases the number of word forms that can be found in the language (e.g. *la* 'the' and *il* 'the' are respectively the feminine and the masculine singular forms of the definite article). On the contrary, English and Hungarian function words do not distinguish gender.

As far as meanings by collocation are concerned, also the difference in the number of collocates computed for the English, Hungarian and Italian lexical items could depend on the morphological typology of the three languages. In fact, an increase in inflectional variability might result in the use of a higher number of word forms for equivalent lemmas and in the consequent scattering of the number of occurrences per word form.

This difference in the morphological typology of the three languages might heavily impact on the outcome of the extraction of collocational patterns when using a raw corpus, as it has been done in the present study. In fact, in the current research, the statistical measures employed to compute collocations pondered the frequency of occurrence of the collocate and/or the frequency of co-occurrence of the node and the collocate. When the statistical validity of a collocational pattern was calculated for a lemma or for a high-frequency word-form, it was likely for the lemma or for the high-frequency word-form to appear among the highest ranking collocates of the node. On the contrary, when the same statistical validity was calculated for a lower-frequency

word-form, then it was likely for it to appear in a lower position in the collocation list or not to appear at all. The differences in the number of collocates featuring in the 2030 Agenda Corpus for the English lexemes and their Hungarian and Italian translational equivalents often depended on the fact that some word-forms in the highly inflected languages were not frequent enough to be counted as collocates.

This limit in the collection of cultural keywords and collocation networks with raw corpora could be overcome with the use of morphologically annotated corpora, as it is suggested, among others, by Leech (1997) and Kübler and Zinsmeister (2015). Word level annotation, in fact, is helpful to meet the morphological characteristics of languages. As Leech (1997) observes, lemmatisation helps much indeed during the analysis of highly inflected languages. Even though all languages included in the 2030 Agenda Corpus are strongly suffixing, in fact, Hungarian and Italian stand out for their being highly inflectional (see, for instance, Kenesei et al. 1998 for Hungarian and Maiden and Robustelli 2000 for Italian; see also § 3.5 for more on this). This means that lemmas might unfold in discourse with a very high number of variants, challenging computational queries of Hungarian or Italian corpora. As a consequence, morphological annotation, and lemmatisation in particular, could help in grouping all word-forms belonging to a certain lemma and they would thus simplify corpus searches of words and the calculation of frequency lists.

Apart from corpus searches of words and calculation of frequency lists, also the computation of collocations could profit from morphologically annotated corpora, as Kübler and Zinsmeister (2015) notice (but see also Freddi 2013). This would be even truer if we looked for the collocates of a word, say *fenntartható* 'sustainable' in Hungarian and we found two forms of the same lemma as *fejlődés* 'development' or *fejlődéshez* 'for development', or also if we could not find such collocates only because their not being grouped in a single lemma slightly filtered them out of frequency thresholds.



## 5. Analysis of the *Sustainable development Corpus*

In the fifth chapter, I report the results of the analysis of the discursive construction of sustainable development in the Sustainable development Corpus. The discursive construction of sustainable development is studied in turn in the three sections of the corpus, namely the SusCorp (English), the SusCorp (Hungarian) and the SusCorp (Italian). The English section of the Sustainable development Corpus is explored in search for the discursive construction of sustainable development in the British press, the Hungarian section is studied in search for the discursive construction of sustainable development in the Hungarian press, the Italian section is expanded on in search for the discursive construction of sustainable development in the Italian press. The three subcorpora are accessed first in search for cultural keywords, which emerge among the most frequent and statistically most salient lexemes of the collections. The identification of cultural keywords is followed by the study of their meaning by collocation. This is bolstered with the analysis of the meaning by collocation of the lexeme *SUSTAINABLE* for English, and of its Hungarian and Italian equivalents (i.e. *FENNTARTHATÓ* for Hungarian and *SOSTENIBILE* for Italian).

### 5.1. Cultural keywords of the *SusCorp*

Cultural keywords are identified in the English, Hungarian and Italian sections of the Sustainable development Corpus in the form of lexemes which are esteemed as representative of the most meaningful political, social and cultural concerns reflected by the most influential quality papers of the United Kingdom, Hungary and Italy in relation to sustainable development after the publication of the 2030 Agenda. By reflecting these political, social and cultural concerns, cultural keywords contribute to the aboutness of the SusCorp.

#### 5.1.1. The most frequent lexemes

Cultural keywords are found first among the most frequent lexemes of the SusCorp (English), of the SusCorp (Hungarian) and of the SusCorp (Italian). The list of the most frequent lexemes of the SusCorp provides cultural keywords because it includes words that are extremely common in the linguistic habits of the corpus.

The most frequent lexemes of the three sections are collected with the aid of AntConc's Word List tool. Only the twenty most frequent content words of the subcorpora are carefully investigated with the reading of their concordance lines. The frequency of these content words is signalled with an *f* standing for "raw frequency". Cultural keywords are spotlighted among the most frequent content words of the subcorpora whenever these content words embody political, social or cultural matters that are meaningful for the contemporary British, Hungarian and Italian societies in relation to sustainable development.

Cultural keywords are investigated further by outlining their meaning by collocation. In this case, meaning by collocation is obtained with the analysis of collocational patterns and clusters. While collocational patterns highlight the lexemes that cultural keywords co-occur with most significantly, clusters emphasise the lexical items that they modify or that they are modified by. Collocational patterns are extracted with #LancsBox's GraphColl tool. Collocates are calculated with the Z association measure, a statistical threshold of 10.0, a frequency threshold of 20, and a span growing from  $\pm 1$  to  $\pm 5$  words. Clusters are gathered with AntConc's Cluster/N-gram tool. The tool is set to that clusters are at least two-words long and at the utmost three-words long, and so that they might be considered only when they appear at least five times in the subcorpus. This frequency threshold applies also to the reading of concordance lines: only concordances appearing at least five times in a corpus are usually considered for further study.

When analysing the SusCorp (Hungarian) and the SusCorp (Italian), results are reported in their original form together with their English translational equivalents. Translational equivalents are found in bilingual dictionaries, namely the Koltay-Kastner and Juhász (2000) dictionary for Hungarian and the Ragazzini (2019) dictionary for Italian. The identification process of English translational equivalents for the SusCorp differs from the corresponding identification process for the 2030 Agenda Corpus. For the analysis of the 2030 Agenda Corpus, in fact, English translational equivalents are found by matching Hungarian and Italian lexemes in the Hungarian and Italian versions of the 2030 Agenda with their English counterparts in the English version of the 2030 Agenda.

### 5.1.1.1. English

The identification of cultural keywords in the Sustainable development Corpus starts with the extraction of the most frequent lexemes of the SusCorp (English).

The most frequent lexemes of the SusCorp (English) are the function words *the* ( $f=25,783$ ), *to* ( $f=13,505$ ), *of* ( $f=12,455$ ), *and* ( $f=12,402$ ), and *in* ( $f=9,344$ ). The first content word that emerges in the corpus is *development*, which is ranked 27<sup>th</sup> with 1,625 occurrences. *development* is followed by the lexemes listed in Table 30.

Rank	Content word	<i>f</i>			
1	<i>development</i>	1,625	11	<i>government</i>	799
2	<i>said</i>	1,467	12	<i>climate</i>	748
3	<i>world</i>	1,370	13	<i>UK</i>	691
4	<i>people</i>	1,270	14	<i>change</i>	687
5	<i>countries</i>	1,039	15	<i>years</i>	631
6	<i>global</i>	1,000	16	<i>UN</i>	630
7	<i>sustainable</i>	931	17	<i>health</i>	585
8	<i>new</i>	916	18	<i>goals</i>	567
9	<i>says</i>	861	19	<i>international</i>	560
10	<i>year</i>	842	20	<i>now</i>	554

**Table 30.** The twenty most frequent content words of the SusCorp (English).

Table 30 includes only lexical verbs, nouns, adjectives and adverbs, but it does not consider possessives, and demonstrative and indefinite adjectives. In other terms, it comprises only those verbs that report a process, those nouns that present a clear-cut referent, and those adjectives and adverbs that qualify verbs, nouns, adjectives, or adverbs in an explicit way. These lexemes are regarded as the more appropriate for the identification of cultural keywords. As a consequence, also adverbs like *more* ( $f=1,491$ , ranking 29<sup>th</sup> in the complete frequency list) and *most* ( $f=601$ , ranking 77<sup>th</sup> in the complete frequency list) are excluded from the list.

The most frequent content words of the SusCorp (English) reflect the aboutness and the stylistic peculiarities of the articles included in the collection. As far as aboutness is concerned, the first topic dealt with in the corpus is sustainable development, which is encoded through the lexemes *development*, *sustainable* and *goals*. Sustainable development is written to be a matter that needs to be approached by the world's countries thanks to the lexemes *world*, *countries*, *global* and *international*. The *UK* stands out among the world's countries. The aboutness of the articles included in

the SusCorp (English) is contributed to also thanks to their frequently mentioning national and international politics (through *government* and *UN*) and thanks to their tackling the social and environmental dimensions of sustainable development. The social dimension of sustainable development is clearly represented by the lexemes *people* and *health*, while the environmental dimension is exemplified by the lexemes *climate* and *change*, which compose the lexical item *climate change*. As far as style is concerned, the English section of the SusCorp is characterized by an abundant use of reported speech, which is frequently introduced by the lexemes *said* and *says*, by a recurrent indication of time (as with the lexemes *year*, *years* and *now*), and by the repeated presentation of novelties (thanks to the adjective *new*).

Among these lexical items, cultural keywords can be spotted in the lexemes *development*, *global*, *sustainable*, *goals*, but also in the lexemes *climate* and *change*. These lexemes can be regarded as cultural keywords because of their referring to issues of political, social and cultural significance in relation to sustainable development for the United Kingdom in 2010s. The semantics of the cultural keyword *sustainable* is explored in § 5.2.1.

The noun *development* is a cultural keyword that refers to a condition of growth and evolution whose specificity is clarified by the nouns and adjectives that co-occur with *development*. *development* is modified by the nouns and adjectives that can be read in Table 31.

<b>Nouns and adjectives</b>	<b><i>f</i></b>		
<i>sustainable</i>	645	<i>human</i>	9
<i>international</i>	89	<i>national</i>	8
<i>global</i>	31	<i>proposed</i>	7
<i>millennium</i>	28	<i>restless</i>	6
<i>economic</i>	22	<i>rural</i>	5
<i>overseas</i>	18	<i>urban</i>	5

**Table 31.** Nouns and adjectives modifying *development* in the SusCorp (English).

These nouns and adjectives contribute to defining *development* as a condition that touches upon a national and a global dimension (*international*, *global*, *overseas*, *national*) and that regards the economic and the social spheres of life (*economic*, *human*). It is a condition that is encouraged (*proposed*) and that sometimes can be unsteady (*restless*). It occurs both in cities (*urban*) and in the surrounding countryside (*rural*) and it can be

governed either by the 2030 Agenda for Sustainable Development (*sustainable*) or by the 2015 Millennium Agenda (*millennium*).

At the same time *development* modifies the nouns in Table 32.

<b>Nouns</b>	<b>f</b>		
		<i>sector</i>	10
<i>goals</i>	309	<i>challenges</i>	8
<i>professionals</i>	61	<i>projects</i>	8
<i>goal</i>	35	<i>agencies</i>	7
<i>agenda</i>	25	<i>commission</i>	7
<i>assistance</i>	24	<i>cooperation</i>	7
<i>bank</i>	20	<i>issues</i>	7
<i>secretary</i>	19	<i>plans</i>	7
<i>capital</i>	17	<i>spending</i>	6
<i>programme</i>	16	<i>targets</i>	6
<i>committee</i>	11	<i>aid</i>	5
<i>institute</i>	11	<i>community</i>	5
<i>plan</i>	11	<i>initiatives</i>	5
<i>policy</i>	11	<i>model</i>	5
<i>banks</i>	10	<i>solutions</i>	5
<i>organisations</i>	10		

**Table 32.** Nouns modified by *development* in the SusCorp (English).

The nouns displayed in Table 32 show that development is tied to the SDGs (*goals, goal, agenda, targets*) and that it involves singular and plural actors (*professionals, bank, secretary, committee, institute, banks, organisations, sector, agencies, commission, community*), who might assist in the development enterprise (*assistance, capital, cooperation, spending, aid*) and plan it (*programme, plan, policy, projects, plans, initiatives, model*) in order to find solutions to problems (*challenges, issues, solutions*).

Moreover, *development* is associated to nouns and noun phrases through the conjunction *and*. When *and* precedes *development*, the lexeme is combined at least five times with the nouns *aid* ( $f=14$ ), *research* ( $f=10$ ), *peace* ( $f=5$ ) and with the adjective *humanitarian* ( $f=12$ ). When *and* follows *development*, the lexeme is combined at least five times with the adjective *humanitarian* ( $f=6$ ) and with the noun *prosperity* ( $f=5$ ).

The second cultural keyword identified for the SusCorp (English), namely the adjective *global*, refers to a property of internationality and universality, and it qualifies the nouns that can be read in Table 33.

<b>Nouns and noun phrases</b>	<b>f</b>		
<i>warming</i>	54	<i>leadership</i>	7
<i>goals</i>	53	<i>level</i>	7
<i>fund</i>	26	<i>challenges</i>	6
<i>poverty</i>	22	<i>effort</i>	6
<i>health</i>	21	<i>lead</i>	6
<i>economy</i>	20	<i>scale</i>	6
<i>community</i>	15	<i>temperature</i>	6
<i>education</i>	15	<i>witness</i>	6
<i>partnership</i>	15	<i>campaign</i>	5
<i>development professionals</i>	13	<i>consensus</i>	5
<i>population</i>	12	<i>economic growth</i>	5
<i>governance</i>	11	<i>environment facility</i>	5
<i>south</i>	10	<i>greenhouse gas emissions</i>	5
<i>compact</i>	9	<i>growth</i>	5
<i>issues</i>	9	<i>investors</i>	5
<i>leaders</i>	9	<i>learning crisis</i>	5
<i>movement</i>	9	<i>market</i>	5
<i>trade</i>	9	<i>pact</i>	5
<i>sustainable development</i>	8	<i>plan</i>	5
<i>impact</i>	7	<i>tax</i>	5

**Table 33.** Nouns and noun phrases modified by *global* in the SusCorp (English).

As it can be seen in Table 33, the adjective *global* modifies nouns and noun phrases that refer most importantly to politics. Politics is represented by lexical items like *goals*, *partnership*, *development professionals*, *governance*, *south*, *compact*, *issues*, *leaders*, *sustainable development*, *impact*, *leadership*, *level*, *challenges*, *effort*, *lead*, *scale*, *witness*, *campaign*, *consensus*, *growth*, *pact*, *plan*. Also the social sphere of life features among the nouns and noun phrases modified by *global* (*poverty*, *health*, *community*, *education*, *population*, *movement*, *learning crisis*), together with the environment (*warming*, *temperature*, *environment facility*, *greenhouse gas emissions*) and economy (*fund*, *economy*, *trade*, *economic growth*, *investors*, *market*, *tax*). The adjective *global* embodies a property that qualifies especially aspects of the political life of the world's communities.

The cultural keyword *goals* refers to the development targets set by politics for global growth and improvement. These goals are most importantly the *sustainable development goals* ( $f=271$ ), which are mentioned also in the form of *the 17 goals* ( $f=12$ ) or the *sustainability goals* ( $f=6$ ). Development targets can also be the *millennium development goals* ( $f=25$ ), but they can also be the *Paris goals* ( $f=6$ ) or *global goals* ( $f=53$ ) in general.

The cultural keyword *climate* appears in the corpus 748 times. It is used both as an autonomous lexeme (in 50 instances) and as a part of compounds (in 698 cases). When it is part of compounds, it can function as a pre-modifying first item (in 694 cases) or as a pre-modified second item (in 60 instances); some of the two compound forms overlap.

In most cases, *climate* appears in compounds as a pre-modifier. It modifies the noun *change* in more than half of its occurrences (429 out of 748, roughly 57%). With a span of 5 words to the left and 5 words to the right of *climate change*, the lexical item collocates with verbs that reproduce transformative material processes like *combat* ( $Z=56.47, f=10$ ), *address* ( $Z=51.02, f=17$ ), *tackle* ( $Z=46.75, f=15$ ), *tackling* ( $Z=44.92, f=11$ ); these verbs show that climate change is actively approached with the aim at reducing it. This active approach to climate change is stressed thanks to the war metaphor triggered by *combat*: climate change is an enemy that needs to be fought. *climate change* collocates also with nouns that highlight how the reduction of climate change needs striving, as in the case of *efforts* ( $Z=28.96, f=12$ ) and *action* ( $Z=28.32, f=15$ ). This commitment depends on the significant consequences of *climate change*, which is written to have *effects* ( $Z=43.50, f=10$ ) and a precise *impact* ( $Z=17.42, f=11$ ). The consequences of climate change are evident in society and they are showcased by collocates like *inequality* ( $Z=51.50, f=17$ ) and *poverty* ( $Z=23.69, f=14$ ). These consequences should be clearly shown and addressed in an international scale, as the collocates *global* ( $Z=22.93, f=24$ ) and *world* ( $Z=10.57, f=12$ ) suggest. The problems caused by climate change might be faced by the world via the information and recommendations that international summits on climate produce; international summits on climate are reminded of with collocates like *agreement* ( $Z=61.92, f=26$ ), *Paris* ( $Z=60.69, f=27$ ), and *UN* ( $Z=17.40, f=13$ ).

Apart from the extremely frequent *change*, the other lexemes that are pre-modified by *climate* refer to political actions (*action*,  $f=21$ ; *agreement*,  $f=16$ , and the relative plural form *agreements*,  $f=1$ ; *talks*,  $f=14$ ; *deal*,  $f=8$ ; *accord*,  $f=7$ ; *policy*,  $f=5$ , and *summit*,  $f=5$ ). They also present political protagonists that operate in relation to climate (*leader*,  $f=8$ , and the relative plural form *leaders*,  $f=2$ ). They can also regard the expenses required to tackle climate issues (*finance*,  $f=7$ , and *fund*,  $f=7$ ). *climate* is also pre-modifier of nouns that refer to disciplines involved with its study or to people that deal with it (*science*,  $f=7$ ; *scientists*,  $f=5$ ). Climate is described as causing and undergoing severe difficulties (*crisis*,  $f=6$ ; *risks*,  $f=5$ ). Climate is also associated with social problems, which

result from challenging environmental conditions (as with *justice*,  $f=5$ ). It is vexed by harmful conditions and elements like *pollutants* ( $f=7$ ).

Compounds constituted by *climate* and another lexeme are modified most frequently by adjectives like *global* ( $f=8$ ), *green* ( $f=5$ ) and *short-lived* ( $f=5$ ). These adjectives qualify *climate* and its compounds as belonging both to an international dimension (as in *global*), and they make them swing between a positive (*green*) and a negative condition (*short-lived*).

*climate* compounds are most frequently the arguments of verbs like *ADDRESS* ( $f=20$ ) and *TACKLE* ( $f=20$ ), *COMBAT* ( $f=15$ ), *FIGHT* ( $f=7$ ), *AFFECT* ( $f=6$ ). These verbs are examples of transformational material processes and they demonstrate that *climate* compounds are linguistically treated as goals to halt or achieve. The halting or achieving of actions, states, or events related to climate takes the form of a war as the verbs *COMBAT* and *FIGHT* imply.

*climate* compounds are additionally found in prepositional phrases whose head are prepositions like: *on* ( $f=69$ ), *about* ( $f=7$ ), *over* ( $f=5$ ), when *climate* compounds are the topic of discussion; *of* ( $f=76$ ) and *in* ( $f=7$ ), when *climate* compounds specify the previous noun or verb; *to* ( $f=36$ ), which usually completes the previous noun phrase or verb phrase; *for* ( $f=15$ ), which assigns compounds of *climate* the role of goals; *by* ( $f=14$ ), when *climate* compounds function as the actors of passive verbs; *from* ( $f=13$ ), which makes compounds of *climate* the origin of events or states; *against* ( $f=7$ ), which makes climate change the goal of a battle encoded with the phrases *the battle against climate change* ( $f=2$ ), *the fight against climate change* ( $f=2$ ), and *the global struggle against climate change* ( $f=1$ ).

*climate* compounds are also found in noun group complexes where they are matched with the nouns and noun phrases *poverty* ( $f=9$ ), *energy* ( $f=8$ ), *inequality* ( $f=8$ ), *sustainable development* ( $f=6$ ). This matching occurs either with the conjunction *and* or by juxtaposition. The nouns and noun phrases that match with *climate* compounds depict an image of climate-related concerns that interest first a social dimension (*poverty*, *inequality*), second an economic dimension (*energy*). Both dimensions are explicitly connected to *sustainable development*.

### 5.1.1.2. Hungarian

After the study of the most frequent lexemes of the SusCorp (English) and the identification of cultural keywords in the English subcorpus, the SusCorp (Hungarian)

is explored to find the most frequent lexemes appearing in the Hungarian subcorpus and to observe whether cultural keywords can be spotted among them.

The most frequent lexemes of the Hungarian section of the Sustainable development Corpus are function words: the definite articles *a* ‘the’ ( $f=25,193$ ) and *az* ‘the’ ( $f=10,312$ ), the conjunctions *és* ‘and’ ( $f=5,937$ ), *hogy* ‘that’ ( $f=4,134$ ), and *is* ‘also’ ( $f=3,216$ ). The first content word that appears in the frequency list of the SusCorp (Hungarian) is *fenntartható* ‘sustainable’, which ranks 11<sup>th</sup> with 831 occurrences; the other most frequent content words of the collection can be read in Table 34.

Rank	Content word	<i>f</i>			
1	<i>fenntartható</i> ‘sustainable’	831	11	<i>ENSZ</i> ‘United Nations’	386
2	<i>magyar</i> ‘Hungarian’	766	12	<i>lehet</i> ‘(he/she/it) can be’	382
3	<i>van</i> ‘(he/she/it) is’	691	13	<i>fejlődés</i> ‘development’	363
4	<i>európai</i> ‘European’	625	14	<i>kormány</i> ‘government’	343
5	<i>volt</i> ‘(he/she/it) was’	581	15	<i>nagy</i> ‘big; Nagy (proper noun)’	320
6	<i>új</i> ‘new’	477	16	<i>Unió</i> ‘(of the) Union’	317
7	<i>mondta</i> ‘(he/she/it) said’	474	17	<i>ország</i> ‘country’	313
8	<i>gazdasági</i> ‘economic’	460	18	<i>lesz</i> ‘(he/she/it) will be’	295
9	<i>Magyarország</i> ‘Hungary’	451	19	<i>világ</i> ‘world’	293
10	<i>nemzetközi</i> ‘international’	414	20	<i>nemzeti</i> ‘national’	281

**Table 34.** The twenty most frequent content words of the SusCorp (Hungarian).

Table 34 comprises only lexical verbs, nouns, adjectives and adverbs. Within these word classes, demonstrative and indefinite adjectives and adjectives like *több* ‘more’ are excluded because of their not reproducing events, entities, or properties that qualify these events and entities in the world; the same applies to adverbs like *már* ‘already’, *csak* ‘only’, *még* ‘still, yet’ which signal the time frame that surrounds the articles included in the SusCorp (Hungarian).

The content words listed in Table 34 reflect the aboutness and style of the collection. As far as aboutness is concerned, the SusCorp (Hungarian) revolves around the core topic of sustainability (thanks to the adjective *fenntartható* ‘sustainable’) and it handles it in terms of development (with the noun *fejlődés* ‘development’). Sustainable development is debated in relation to the opposition between a national and an international dimension, with the lexemes *magyar* ‘Hungarian’, *Magyarország* ‘Hungary’, and *nemzeti* ‘national’ on the national side and *európai* ‘European’, *nemzetközi* ‘international’, *Unió* ‘(of the) Union’ (i.e. the European Union), and *világ*

'world' on the international side; the lexeme *ország* 'country' stands in between. The national and the international engagement for sustainable development require the actions of politics, represented by governments and organizations (*ENSZ* 'United Nations' and *kormány* 'government') and they are mainly regarded in economic terms (*gazdasági* 'economic'). In terms of style, the articles included in the SusCorp (Hungarian) are characterized by a very frequent use of the existential verb, set in the present, in the past, and in the future (*van* '(he/she/it) is', *volt* '(he/she/it) was', *lesz* '(he/she/it) will be'), but also in its potential form (*lehet* '(he/she/it) can be'). The SusCorp (Hungarian) also features verbal processes of saying (thanks to *mondta* '(he/she/it) said', definite conjugation) and it relies on qualities like novelty (*új* 'new') and bigness (*nagy* 'big', which can be used also as a proper name).

Among these content words, the lexemes *fenntartható* 'sustainable' and *fejlesztés* 'development' can be regarded as cultural keywords because of their pointing at some of the most meaningful political, social and cultural issues raised by Hungarian news discourse on sustainable development between 2016 and 2018. The adjective *fenntartható* 'sustainable' is pinpointed as a cultural keyword because it addresses one of the aspirations that are longed for worldwide and that leads the wellbeing of the world's population and environment. The semantics of the adjective is outlined in detail in § 5.2.2. The noun *fejlesztés* 'development' is chosen as a cultural keyword because it mirrors a condition that has been pursued by the world's population for centuries and that posits challenges to the management of the planet. In the SusCorp (Hungarian), *fejlesztés* 'development' is mainly described as *fenntartható* 'sustainable' ( $f=302$ ), but it can also be *gazdasági* 'economic' ( $f=16$ ). Decisions on *fenntartható fejlesztés* 'sustainable development' are taken by the Hungarian parliament (*országgyűlés* 'parliament',  $Z=133.01$ ,  $f=59$ ), which is governed by a president (*elnöke* '(its) president',  $Z=35.85$ ,  $f=22$ ) which governs on commissions (*bizottságának* 'to (its) commission',  $Z=126.37$ ,  $f=38$ ; *bizottsága* 'commission',  $Z=124.70$ ,  $f=30$ ).

### 5.1.1.3. Italian

The exploration of the cultural keywords of the Sustainable development Corpus continues with the retrieval of the most frequent lexemes of the SusCorp (Italian) and with the identification of cultural keywords among them.

The most frequent lexemes of the SusCorp (Italian) are the function words *di* 'of' ( $f=12,750$ ), *e* 'and' ( $f=9,810$ ), *il* 'the' (singular masculine,  $f=6,623$ ), *la* 'the' (singular feminine,  $f=5,735$ ), and *che* 'that' ( $f=5,215$ ).

The highest-ranking content word of the wordlist (i.e. *sviluppo* 'development') appears in the 29<sup>th</sup> position with 1,340 occurrences. It is followed by the content words that can be read in Table 35.

Rank	Content word	f			
1	<i>sviluppo</i> 'development'	1,340	11	<i>parte</i> 'part'	366
2	<i>sostenibile</i> 'sustainable'	1,042	12	<i>presidente</i> 'president'	357
3	<i>Italia</i> 'Italy'	751	13	<i>stato</i> 'state; been'	348
4	<i>anni</i> 'years'	638	14	<i>paese</i> 'country'	333
5	<i>paesi</i> 'countries'	487	15	<i>ambiente</i> 'environment'	330
6	<i>mondo</i> 'world'	455	16	<i>Europa</i> 'Europe'	327
7	<i>anno</i> 'year'	424	17	<i>città</i> 'city; cities'	321
8	<i>lavoro</i> 'job; work'	411	18	<i>crescita</i> 'growth'	309
9	<i>oggi</i> 'today'	391	19	<i>persone</i> 'people'	304
10	<i>obiettivi</i> 'goals'	370	20	<i>sostenibilità</i> 'sustainability'	303

**Table 35.** The twenty most frequent content words of the SusCorp (Italian).

Table 35 encompasses lexical verbs, nouns, adjectives and adverbs. For the purposes of this study, verbs that play both a lexical and an auxiliary function like *è* '(he/she/it) is' ( $f=3,594$ ), *ha* '(he/she/it) has' ( $f=1,525$ ), *sono* '(I) am; (they) are' ( $f=1,419$ ), *hanno* '(they) have' ( $f=527$ ), *essere* 'to be' ( $f=425$ ), *sia* '(I) were; (he/she/it) was' (subjunctive,  $f=374$ ) are excluded from the list of the most frequent content words analysed to trace aboutness and style; the same applies to adverbs like *più* 'more' ( $f=1,594$ ), *solo* 'only' ( $f=502$ ), *oltre* 'beyond' ( $f=364$ ), *ancora* 'still' ( $f=358$ ), to demonstrative adjectives like *questo* 'this' ( $f=620$ ), to indefinite adjectives or pronouns like *tutti* 'all' (plural masculine,  $f=418$ ), or to numerals like *milioni* 'millions' ( $f=460$ ), *uno* 'one' ( $f=389$ ), *due* 'two' ( $f=336$ ), *prima* 'first; before' ( $f=313$ ), *2030* '2030' ( $f=303$ ). These lexical items contribute to the style that can be envisaged in the SusCorp (Italian).

As Table 35 shows, in terms of aboutness, the SusCorp (Italian) ties sustainable development (expressed through *sostenibile* 'sustainable', *sviluppo* 'development', *obiettivi* 'goals', *sostenibilità* 'sustainability') to a national and to an international panorama (with *Italia* 'Italy', *paesi* 'countries', *mondo* 'world', *stato* 'state; been', *paese* 'country', *Europa* 'Europe'). The most frequent lexemes of the subcorpus hint at the

environmental and social dimensions of sustainability (through the lexemes *ambiente* 'environment' and *persone* 'people') and they associate sustainable development to growth (*crescita* 'growth') and employment (*lavoro* 'job; work'). Then, at the interplay between the interests of society, environment and economy, they identify cities (*città* 'city; cities'). They also recognise that sustainable development is intertwined with politics (*presidente* 'president'). In terms of style, the articles included in the Italian section of the Sustainable development Corpus frequently refer to time through the nouns *anni* 'years' and *anno* 'year', but also through the adverb *oggi* 'today'.

Among the content words displayed in Table 35, some can be esteemed cultural keywords for the SusCorp (Italian) because they mirror and shape the political, social and cultural context that saw the rise of quality news discourse on sustainable development in Italy between 2016 and 2018. These cultural keywords are *sostenibile* 'sustainable', *sviluppo* 'development', *obiettivi* 'goals', *sostenibilità* 'sustainability'. They can be regarded key because they refer to one of the most debated issues of the 2010s in Italy, namely sustainable development.

In the SusCorp (Italian), the noun *sviluppo* 'development' is qualified as *sostenibile* 'sustainable' in 59% of its occurrences ( $f=787$ ; the semantics of *sostenibile* 'sustainable' is explored in § 5.2.3). In the remaining cases, *sviluppo* 'development' is classified as *economico* 'economic' ( $f=32$ ), *agricolo* 'agricultural' ( $f=7$ ), *umano* 'human' ( $f=7$ ), *urbano* 'urban' ( $f=5$ ), and it can be described as *locale* 'local' ( $f=7$ ), *globale* 'global' ( $f=6$ ), *equo* 'equitable' ( $f=5$ ). *sviluppo* 'development' in general and *sviluppo sostenibile* 'sustainable development' in particular are the matter or the goal of institutions (e.g. *fondazione* 'foundation',  $Z=63.99$ ,  $f=62$ ) and they are planned through models and strategies like *modello* 'model' ( $Z=55.59$ ,  $f=43$ ), *strategia* 'strategy' ( $Z=29.91$ ,  $f=21$ ), *politiche* 'policies' ( $Z=24.88$ ,  $f=27$ ). *sviluppo* 'development' is the goal of the material process *raggiungere* 'achieving' ( $Z=27.64$ ,  $f=20$ ), and it is also the target of evolutionary processes that involve countries in the lexical item *in via di sviluppo* 'developing' ( $f=52$ ). These development processes are associated to *obiettivi* 'goals' ( $Z=119.91$ ,  $f=155$ ), namely the seventeen SDGs.

Also the noun *obiettivi* 'goals' functions as a cultural keyword in the SusCorp (Italian). Since the release of the United Nations' 2030 *Agenda for Sustainable Development*, in fact, the achievement of sustainable development has been bound to the realisation of the seventeen sustainable development goals. Because of this, the noun *obiettivi* 'goals' is typically found in lexical items like *obiettivi di sviluppo sostenibile* 'sustainable development goals' ( $f=116$ ) or *obiettivi per lo sviluppo sostenibile* 'sustainable

development goals' ( $f=11$ ), or simply *obiettivi di sviluppo* 'development goals' ( $f=7$ ), but it can also be encompassed in *obiettivi dell'Agenda 2030* '2030 Agenda goals' ( $f=12$ ).

The latest cultural keyword identified among the most frequent content words of the SusCorp (Italian) is *sostenibilità* 'sustainability'. *sostenibilità* 'sustainability' is defined as a *concetto* 'concept' ( $Z=32.01, f=7$ ) or as a *tema* 'topic' ( $Z=19.69, f=8$ ) associated with other concepts like *innovazione* 'innovation' ( $Z=26.46, f=8$ ) and pursued through actions and plans like *bilancio* 'balance' ( $Z=23.67, f=7$ ), *progetto* 'project' ( $Z=13.68, f=7$ ), *politiche* 'policies' ( $Z=13.65, f=7$ ), *piano* 'plan' ( $Z=12.67, f=6$ ), *temi* 'topics' ( $Z=12.63, f=5$ ), *progetti* 'projects' ( $Z=12.61, f=6$ ). *sostenibilità* 'sustainability' is described as *ambientale* 'environmental' ( $f=26$ ), *economica* 'economic' ( $f=8$ ) and *sociale* 'social' ( $f=5$ ), and consequently it covers all the dimensions of sustainable development introduced by the United Nations (i.e. the economic, social and environmental dimension).

### 5.1.2. The statistically significant lexemes

The identification of cultural keywords in the Sustainable development Corpus continues with the retrieval of the statistical keywords rising from the SusCorp (English), the SusCorp (Hungarian) and the SusCorp (Italian). Statistical keywords spotlight lexemes that are statistically more salient in the English, Hungarian and Italian quality news discourse on sustainable development published online between 2016 and 2018, compared to a general reference of online use of English, Hungarian and Italian. These statistically salient lexemes are employed as a productive ground to find cultural keywords, namely those lexemes that help reflect and shape the culture that originated the English, Hungarian and Italian broadsheet articles on sustainability included in the Sustainable development Corpus. The list of the most salient statistical keywords of the SusCorp provides cultural keywords because it contains words that are peculiar to the discourse represented in the corpus compared to the discourse represented by another corpus which is used as a reference.

The statistical keywords of the SusCorp (English) are retrieved on the Sketch Engine platform by using the SusCorp (English) as the study corpus and the enTenTen 2012 and the English section of Timestamped JSI webcorpus 2014-2020 as reference corpora. The statistical keywords of the SusCorp (Hungarian) are retrieved on the Sketch Engine platform by using the SusCorp (Hungarian) as the study corpus and the Hungarian web 2012 (HuTenTen12) corpus and the Hungarian section of the Timestamped JSI webcorpus 2014-2020 as the reference corpora. For Italian, statistical

keywords are extracted by comparing the SusCorp (Italian), used as the study corpus, with the itTenTen 2016 and the Italian section of Timestamped JSI webcorpus 2014-2020, employed as reference corpora. The association measure adopted for the extraction of the statistical keywords is simple maths, set on a 10 score for rare words. The minimum frequency of occurrence of a statistical keyword is fixed on 20. Statistical keywords are searched for among lemmas when comparing the SusCorp (English) and the enTenTen 2012, among word types when comparing the SusCorp (English) and the English section of Timestamped JSI webcorpus 2014-2020. Statistical keywords are searched for among lemmas when comparing the SusCorp (Hungarian) and the huTenTen 2012 and for word types when comparing the SusCorp (Hungarian) and the Hungarian section of Timestamped JSI webcorpus 2014-2020. Statistical keywords are retrieved among lemmas when comparing the SusCorp (Italian) and the itTenTen 2016 and among word types when comparing the SusCorp (Italian) and the Italian section of Timestamped JSI webcorpus 2014-2020.

As in the case of the most frequent content words of the SusCorp, the lexicogrammatical patterns of the twenty highest-ranking statistical keywords are expanded on by means of collocational patterns, clusters and concordances.

### 5.1.2.1. English

The analysis of the cultural keywords characterising the Sustainable development Corpus proceeds with the study of the statistical keywords of the SusCorp (English).

The statistical keywords of the SusCorp (English) add details on the aboutness of the corpus. The twenty most significant statistical keywords extracted by comparing the SusCorp (English) and the enTenTen 2012 can be seen in Table 37.

<b>Rank</b>	<b>Statistical keyword</b>		
1	<i>SUSTAINABLE</i>	11	<i>FAIRTRADE</i>
2	<i>UN</i>	12	<i>NGO</i>
3	<i>SDGs</i>	13	<i>GUARDIAN</i>
4	<i>SDG</i>	14	<i>SUSTAINABILITY</i>
5	<i>HUMANITARIAN</i>	15	<i>INEQUALITY</i>
6	<i>TRUMP</i>	16	<i>INDIGENOUS</i>
7	<i>G20</i>	17	<i>GLOBAL</i>
8	<i>CLIMATE</i>	18	<i>TB</i>
9	<i>POVERTY</i>	19	<i>REFUGEE</i>
10	<i>BREXIT</i>	20	<i>SANITATION</i>

**Table 36.** The twenty most significant statistical keywords of the SusCorp (English) compared to the enTenTen 2012.

The statistical keywords of the SusCorp (English) highlight the specificity of the corpus by stressing first that its texts are shaped around the topic of sustainable development. Sustainable development is mirrored through the lexemes *sustainable*, *SDGs*, *SDG* and *sustainability*, and it is dealt with in its social and environmental dimensions in more specific terms compared to the general matters summoned among the most frequent content words of the corpus. The social dimension of sustainable development is described with the lexemes *poverty*, *inequality*, *indigenous*, *TB* (i.e. Tuberculosis), *refugee*, and *sanitation*. The SusCorp (English) portrays the engagement with the social dimension of sustainable development as concerned especially with the health of people. Health is of interest *per se*, as an independent topic of discussion, but also in relation to the economic conditions of human beings: economic conditions impact on human beings' wellbeing and consequently on the chances they have to contract diseases; this holds true both for individuals and for entire populations. Moreover, the social dimension of sustainable development cares for the movement of peoples around the world and for the reasons why they migrate from one place to another (as with the lexeme *refugees*). The environmental dimension of sustainable development is touched upon by the lexeme *climate*. The social and the environmental dimensions of sustainable development are tended to by international politics (clear in the lexeme *global*), which is represented by lexemes referring to international organizations like *UN*, *G20*, and *Ngos* but also by politicians like *Trump* and by political decisions like *Brexit*. International organizations contribute to sustainability with an aid that can be described as *humanitarian*. This humanitarian engagement impacts also on the economic dimension of sustainable development: it witnesses the growth of *Fairtrade* as opposed to trade which does not respect the rights of the protagonist involved in the exchange. Not only do the statistical keywords of the SusCorp (English) point at international matters, but they also refer to national issues and thus they bolster the cultural specificity of the collection. The most typical in this is the lexeme *Guardian*, which explicitly mentions the most represented broadsheet of the corpus.

When this first keyword list is enhanced with the statistical keywords extracted by comparing the SusCorp (English) to the English section of Timestamped JSI webcorpus 2014-2020, the aboutness of the Sustainable development Corpus is expanded in various aspects. In relation to the environmental dimension of sustainable development, the environment is recalled through the statistical keywords

*environmental* and *biodiversity*. *environmental* offers a very general picture of what the British press means when it reports about sustainable development; the lexeme *biodiversity* raises the issue of the variety of animate and inanimate beings inhabiting the world and needing to be protected when countries develop in a sustainable way. In addition, other culture-specific statistical keywords emerge. *DfID* stands for *Department for International Development*, and it refers to a department of the British government which operates nationally and internationally for the welfare of its citizens. National specificity in relation to sustainable development is also highlighted with the lexeme *Macquarie*, namely an Australia-based investment bank. The British press assigns a paramount importance to Macquarie because of the bank's concern for environmental issues and because of its investments in the infrastructures of the United Kingdom.

Comparing the list of the most frequent content words of the SusCorp (English), which can be seen in Table 30, with the list of the highest-ranking statistical keywords of the collection, which is displayed in Table 36, the same words emerge as cultural keywords of this snapshot of the British news discourse released on sustainable development after the publication of the 2030 Agenda: *development*, *global*, *sustainable*, *goals* and *climate*. The high and statistically significant frequency of *development* and *sustainable* is doubtlessly due to their belonging to the query item employed for the collection of SusCorp's (English) newspaper articles. The same can be stated also for *global* and *goals*. The lexeme *climate*, on the contrary, bears much of the culture that shaped the texts incorporated in the corpus. It can be regarded as a cultural keyword especially for the most represented broadsheet of the collection, namely *The Guardian*, which includes 78% of the occurrences of the word.

### **5.1.2.2. Hungarian**

The identification of the cultural keywords characterising the Sustainable development Corpus is carried on with the extraction of the statistical keywords of the SusCorp (Hungarian).

The statistical keywords of the SusCorp (Hungarian) enhance the features of the aboutness and style that have been noted thanks to the most frequent lexemes of the subcorpus. The twenty most salient statistical keywords collected when comparing the SusCorp (Hungarian) and the Sketch Engine's huTenTen12 corpus can be found in Table 37.

Rank	Statistical keyword		
1	<i>FENNTARTHATÓ</i> 'sustainable'	11	<i>TRUMP</i> 'Trump'
2	<i>MIGRÁCIÓ</i> 'migration'	12	<i>MIGRÁNS</i> 'migrant'
3	<i>ENSZ</i> 'United Nations'	13	<i>VILÁGSZERVEZET</i> 'world organization'
4	<i>ÁDER</i> 'Áder'	14	<i>GLOBÁLIS</i> 'global'
5	<i>FENNTARTHATÓSÁGI</i> '(of) sustainability'	15	<i>SZIJJÁRTÓ</i> 'Szijjártó'
6	<i>FEJLŐDÉSI</i> '(of) development'	16	<i>FEJLŐDÉS</i> 'development'
7	<i>KLÍMAVÁLTOZÁS</i> 'climate change'	17	<i>VÍZÜGYI</i> '(of) water issue'
8	<i>MIGRÁCIÓS</i> '(of) migration'	18	<i>ÉGHAJLATVÁLTOZÁS</i> 'climate change'
9	<i>FENNTARTHATÓSÁG</i> 'sustainability'	19	<i>MAJTÉNYI</i> 'Majtényi'
10	<i>SALLAI</i> 'Sallai'	20	<i>FÖLDMŰVELÉSÜGYI</i> '(of) agriculture issue'

**Table 37.** The twenty most significant statistical keywords of the SusCorp (Hungarian) compared to huTenTen 2012.

These statistical keywords confirm that the core topic of the SusCorp (Hungarian) is sustainable development, represented by the lemmas *FENNTARTHATÓ* 'sustainable', *FENNTARTHATÓSÁGI* '(of) sustainability', *FEJLŐDÉSI* '(of) development', *FENNTARTHATÓSÁG* 'sustainability', and *FEJLŐDÉS* 'development'. Sustainable development happens to be one of the most debated matters of the collection because of corpus design: the Sustainable development Corpus is planned in such a way as to include broadsheet articles explicitly referring to sustainable development. It should be noted that *FENNTARTHATÓSÁGI* '(of) sustainability' and *FENNTARTHATÓSÁG* 'sustainability' emerge among the lexemes that refer to sustainable development in the SusCorp (Hungarian), while in the Hungarian version of the 2030 Agenda they are employed exclusively in relation to debt sustainability (see § 4.3.2 on this point).

Another aspect of the aboutness of the SusCorp (Hungarian) identified among the most frequent content words of the collection and confirmed by statistical keywords is the political nature of sustainable development. Politics is referred to with *ENSZ* 'United Nations', *ÁDER* 'Áder', *SALLAI* 'Sallai', *TRUMP* 'Trump', *VILÁGSZERVEZET* 'world organization', *SZIJJÁRTÓ* 'Szijjártó' and *MAJTÉNYI* 'Majtényi'. These keywords distinguish between a national and an international level of politics. National politics is represented by *ÁDER* 'Áder', *SALLAI* 'Sallai', *SZIJJÁRTÓ* 'Szijjártó' and *MAJTÉNYI* 'Majtényi'. *ÁDER* 'Áder' refers to János Áder, Hungary's president: János Áder is mentioned in the SusCorp (Hungarian) for his engagement with the sustainable management of water; water is included in the list of the most meaningful statistical keywords of the collection through the lemma *VÍZÜGYI* '(of) water issue'. *SALLAI* 'Sallai'

refers to Benedek R. Sallai, a Hungarian MP formerly belonging to the Hungarian green party *LMP – Magyarország Zöld Pártja (Zöldek)*. *SZIJJÁRTÓ* ‘Szijjártó’ refers to Péter Szijjártó, who has been Hungary’s Minister of foreign affairs since 2014. *MAJTÉNYI* ‘Majtényi’ refers to László Majtényi, a scholar of the Hungarian Academy of Sciences. International politics is represented by the United Nations (*ENSZ*), by the American president Donald Trump (*TRUMP*), and by the lemma *VILÁGSZERVEZET* ‘world organization’, which refers to world organizations in general. The global dimension of politics in relation to sustainable development is reproduced also through the lexeme *GLOBÁLIS* ‘global’.

In addition, as far as the aboutness of the SusCorp (Hungarian) is concerned, sustainable development meaningfully correlates with migration via the lemmas *MIGRÁCIÓ* ‘migration’, *MIGRÁCIÓS* ‘(of) migration’, and *MIGRÁNS* ‘migrant’. It is interesting to notice that in the SusCorp (Hungarian) migration is reported with lexemes sharing the root *migr-* (i.e. *migr-áció* ‘migration’, *migr-ációs* ‘(of) migration’, *migr-áns* ‘migrant’), even though a common way of referring to migration in Hungarian is through lexemes characterized by the root *bevándorol-* (i.e. *bevándorl-ás* ‘migration’, *bevándorl-ási* ‘(of) migration’, *bevándorló* ‘migrant’).

Another significant topic in the SusCorp (Hungarian) is climate change. Climate change is rendered in two forms in the subcorpus: *KLÍMAVÁLTOZÁS* ‘climate change’ and *ÉGHAJLATVÁLTOZÁS* ‘climate change’. A similar environmental commitment also stimulates the presence of the lemma *FÖLDMŰVELÉSÜGYI* ‘(of) agriculture issue’ among the most meaningful statistical keywords of the subcorpus.

When enriching the list of these statistical keywords with those extracted by comparing the SusCorp (Hungarian) and the Hungarian section of Timestamped JSI webcorpus 2014-2020, sustainable development is referred to with the lexeme *célok* ‘goals’. The environmental dimension of sustainable development is filled with the lexemes *környezeti* ‘environmental’ and *környezetvédelmi* ‘environmentalist’, whereas the social dimension of sustainable development is further expanded on with the lexeme *szegénység* ‘poverty’.

The statistically salient lexemes of the SusCorp (Hungarian) host cultural keywords that reflect the most meaningful political, social and cultural issues of the Hungarian quality news discourse on sustainable development between 2016 and 2018. These cultural keywords are: *FENNTARTHATÓ* ‘sustainable’, *FENNTARTHATÓSÁGI* ‘(of) sustainability’, *FEJLŐDÉSI* ‘(of) development’, *FENNTARTHATÓSÁG* ‘sustainability’, *FEJLŐDÉS* ‘development’ in relation to sustainable development; *MIGRÁCIÓ* ‘migration’, *MIGRÁCIÓS* ‘(of) migration’, and *MIGRÁNS* ‘migrant’ in relation to migration;

KLÍMAVÁLTOZÁS 'climate change' and ÉGHAJLATVÁLTOZÁS 'climate change' in relation to climate change. The lexical items having to do with sustainable development are key because of their constituting the core of the collection: sustainable development is the cultural ground on which the articles included in the SusCorp (Hungarian) are based. The politics, society and culture mirrored by the SusCorp (Hungarian) are more peculiarly reflected through the lexemes encoding migration and climate change issues.

As far as migration is concerned, the collocational patterns of *MIGRÁNS* 'migrant' portray the migrants dealt with in the SusCorp (Hungarian) as people who move towards Europe (*Európába* 'to Europe',  $Z=60.80, f=5$ , with *Európa* meaning 'Europe' and the *-ba* suffix being the back-vowel form of the illative case) for several reasons, some of them being linked to climate change. Migrants are associated with refugees (*menekültek* 'refugees',  $Z=46.84, f=5$ , with *menekültek* including *menekült* 'refugee' and *-ek*, the front-vowel form of the plural suffix) in a statistically significant way, exemplified by the collocational patterns *menekültek és a migránsok* ( $f=2$ ) and *menekültek és migránsok* ( $f=3$ ), both meaning 'refugees and migrants'. In addition, the lemma *MIGRÁNS* 'migrant' is usually introduced by the definite article but it is also frequently modified by numerals (e.g. *3800 migráns* '3800 migrants'), which signal a quantification of the individuals migrating. Additional modifiers portray the distinction between legal and illegal migrants, and they also differentiate migrants according to the place they come from and according to the reasons why they moved to Europe. By means of collocation, migration (encoded through the lemmas *MIGRÁCIÓ* 'migration' and *MIGRÁCIÓS* '(of) migration') is significantly described as globally distributed (*globális* 'global',  $Z=42.61, f=20$ , and *nemzetközi* 'international',  $Z=24.49, f=15$ ). Migration is frequently deemed illegal (*illegális* 'illegal',  $Z=91.17, f=17$ ) and it raises issues for the countries that have to cope with it: it is regarded as a crisis (*válság* 'crisis',  $Z=24.07, f=5$ ) that needs to be solved (*megoldás* 'solution',  $Z=27.76, f=5$ ). According to the articles included in the SusCorp (Hungarian), the solution to the issues raised by migration is not to be found merely in the reception of the migrants (*befogadás* 'reception',  $Z=56.55, f=5$ ) but also in international agreements on the matter, represented by collocates like *csomag* 'pack' ( $Z=38.86, f=5$ ), *csomagját* 'its pack' ( $Z=76.99, f=6$ ; with *csomag* meaning 'pack', the suffix *-ja* signalling the third person singular possessive suffix, and *-t* standing for the accusative case), *ENSZ* 'United Nations' ( $Z=30.03, f=15$ ). Also national politics plays a role in solving the issues posited by migration, as it is clarified by lexemes like *Orbán* 'Orbán', referring to Hungary's Prime Minister ( $Z=17.84, f=6$ ), or the general *politikai* 'political' ( $Z=13.82, f=6$ ). Thus, in the Hungarian section of the

Sustainable development Corpus, migration is narrated as a phenomenon that involves people moving towards Europe sometimes illegally and other times with the right to be welcomed as refugees. The number of people migrating to Europe is seldom so high that migration is deemed critical and in need to be solved with an international enterprise.

As far as climate change is concerned, the cultural keyword *KLÍMAVÁLTOZÁS* 'climate change' is used 137 times and *ÉGHAJLATVÁLTOZÁS* 'climate change' appears 60 times. Both cultural keywords are meaningfully used with the postposition *elleni* 'against', which forms an attributive adjective phrase when it is preceded by *klímaváltozás* (i.e. the nominative form of *KLÍMAVÁLTOZÁS* meaning 'climate change') or *éghajlatváltozás* (i.e. the nominative form of *ÉGHAJLATVÁLTOZÁS* meaning 'climate change' as well); the adjective phrases *klímaváltozás elleni* ( $f=31$ ) and *éghajlatváltozás elleni* ( $f=6$ ), both meaning 'against climate change', typically modify nouns and noun phrases that have to do with war, namely *harc* 'battle' ( $f=15$ ) and *küzdelem* 'battle' ( $f=14$ ). In the SusCorp (Hungarian), climate change is often addressed with a war metaphor. Both Hungarian forms standing for 'climate change' are written to be a global issue (through the collocate *globális* 'global',  $Z=24.37$ ,  $f=9$ ). The collocational patterns that *KLÍMAVÁLTOZÁS* 'climate change' belongs to clarify that the global enterprise of climate change is carried out by international organizations like *FAO* 'FAO' ( $Z=40.14$ ,  $f=5$ ) and *ENSZ* 'United Nations' ( $Z=15.19$ ,  $f=6$ ) and that it needs to be tackled in order to prevent negative consequences caused by the changing climate. These consequences are reported of with the lexemes *hatásainak* ( $Z=107.26$ ,  $f=7$ ) and *hatásai* ( $Z=72.21$ ,  $f=5$ ). *hatásainak* literally means 'to its effects' and it consists of the nominal stem *hatás* 'effect' followed by the plural suffix *-i*, which is used with plural possessive forms in combination with a first or second person possessive suffix or without any additional suffixes when the possessor is a third person possessor (cf. Manzelli 1990), and by the suffix *-nak*, signalling the dative case; *hatásai* 'its effects' is the nominative form of the same lexeme added only with the plural suffix. In the SusCorp (Hungarian), climate change is depicted as a phenomenon causing negative effects that should be tackled in a warlike way in order to reduce it.

### 5.1.2.3. Italian

The identification of the cultural keywords interspersed in the Sustainable development Corpus continues with the gathering of the statistical keywords of the

SusCorp (Italian), which highlight the peculiarities of the Italian news discourse on sustainable development.

The statistical keywords resulting from the comparison of the SusCorp (Italian) and of the itTenTen 2016 corpus can be seen in Table 38.

<b>Rank</b>	<b>Statistical keyword</b>		
1	<i>SOSTENIBILE</i> 'sustainable'	11	<i>TRUMP</i> 'Trump'
2	<i>AGENDA</i> 'Agenda'	12	<i>PIL</i> 'GDP'
3	<i>ASVIS</i> 'ASviS'	13	<i>GREEN</i> 'green'
4	<i>SOSTENIBILITÀ</i> 'sustainability'	14	<i>POVERTÀ</i> 'poverty'
5	<i>ONU</i> 'UN'	15	<i>UNITE</i> 'United'
6	<i>DISUGUAGLIANZA</i> 'disequality'	16	<i>RICICLARE</i> 'to recycle'
7	<i>NAZIONI</i> 'Nations'	17	<i>GLOBALE</i> 'global'
8	<i>G7</i> 'G7'	18	<i>SVILUPPO</i> 'development'
9	<i>ECONOMY</i> 'economy'	19	<i>PAESI</i> 'countries'
10	<i>CLIMATICO</i> 'climate'	20	<i>G20</i> 'G20'

**Table 38.** The twenty most significant statistical keywords of the SusCorp (Italian) compared to itTenTen 2016.

The statistical keywords of the SusCorp (Italian) contribute to the aboutness of the corpus by highlighting the importance of sustainable development. Sustainable development is mainly evoked with the lemmas *SOSTENIBILE* 'sustainable', *SOSTENIBILITÀ* 'sustainability', *AGENDA* 'Agenda' and *SVILUPPO* 'development', but it is recalled also with the lemma *ASVIS* 'ASviS'. *ASVIS* 'ASviS' stands for "Alleanza Italiana per lo Sviluppo Sostenibile", namely 'Italian Alliance for Sustainable Development'. ASviS is a network that was created in 2016 and that "brings together 270 member organizations among the civil society" in order "to raise the awareness of the Italian society, economic stakeholders and institutions about the importance of the 2030 Agenda for Sustainable Development, and to mobilize them in order to pursue the Sustainable Development Goals (SDGs)", as it is claimed in the website of the organisation.

Thanks to statistical keywords, sustainable development is discursively constructed also in relation to the political actors that have been involved in promoting sustainability within the realm of the 2030 Agenda. Some of these political actors are introduced by means of the statistical keywords *ONU* 'UN', *NAZIONI* 'Nations', *UNITE* 'United' and *PAESI* 'countries'. Other political actors indirectly contributing to the sustainable development venture are *TRUMP* 'Trump', *G7* 'G7' and *G20* 'G20'.

Among the statistical keywords of the SusCorp (Italian), the environmental dimension of sustainable development is the most significantly represented. This dimension is mirrored by the lemmas *GREEN* 'green', which modifies several nouns and noun phrases having to do with society and economy, *RICICLARE* 'to recycle', which introduces recycling as one of the positive actions that can be done for the sake of sustainable development, and *CLIMATICO* 'climate', which is almost always used together with *cambiamento* 'change' in the noun phrase *cambiamento climatico* 'climate change'.

It is interesting to notice that environmental concerns are meaningfully addressed with the English adjective *GREEN* 'green' instead of its Italian equivalent *VERDE* 'green'. *VERDE* 'green' occurs in the SusCorp (Italian) almost as frequently as *GREEN* 'green': the former appears 148 times, the latter 159 times. Nevertheless, despite this similar raw frequency, the English adjective *GREEN* 'green' stands out among the lemmas of the subcorpus because of its appearing in a statistically significant way in the SusCorp (Italian) compared to the itTenTen 2016. This is due to the fact that the adjective *VERDE* 'green' is definitely more common in Italian compared to its English equivalent *GREEN* 'green'. In the SusCorp (Italian), the English adjective *GREEN* 'green' modifies lexemes that refer to economic and social issues; some of these are in Italian, while others are in English. The Italian nouns and noun phrases modified by *GREEN* 'green' are: *investimenti* 'investments' (f=3), *economia* 'economy' (f=2), *edilizia* 'construction' (f=2), *finanza* 'finance' (f=2), *imprese* 'enterprises' (f=2), *obbligazioni* 'bonds' (f=2), *svolta* 'turning point' (f=2), *Accordo* 'deal' (f=1), *aziende* 'companies' (f=1), *benessere* 'wellbeing' (f=1), *brevetti* 'patents' (f=1), *chiave* 'key' (f=1), *comportamenti* 'behaviours' (f=1), *emissione* 'emission' (f=1), *fiera* 'fair' (f=1), *flotta* 'fleet' (f=1), *impegno* 'commitment' (f=1), *interesse* 'interest' (f=1), *istanza* 'application' (f=1), *Italia* 'Italy' (f=1), *lavoro* 'work' (f=1), *misure* 'measures' (f=1), *mobilità* 'mobility' (f=1), *orientamento* 'positioning' (f=1), *parola* 'word' (f=1), *reputazione* 'reputation' (f=1), *riconversione* 'conversion' (f=1), *rivoluzione* 'revolution' (f=1), *scelte* 'choices' (f=1), *scommessa* 'bet' (f=1), *sfida* 'challenge' (f=1), *trasformazione* 'transformation' (f=1), *trasporto* 'transport' (f=1), *turismo* 'tourism' (f=1), *veicoli* 'vehicles' (f=1). The English nouns and noun phrases modified by *GREEN* 'green' are: *economy* (f=58), *BOND* (f=12), *city* (f=6), *finance* (f=3), *Evaluation* (f=2), *jobs* (f=2), *Leaf* (f=2), *Public Procurement* (f=2), *Team* (f=2), *Way* (f=2), *boom* (f=1), *business* (f=1), *Capital* (f=1), *Energy Storage* (f=1), *Festival* (f=1), *Financial Center* (f=1), *Lab* (f=1), *lobby* 'lobby' (f=1), *Power* (f=1), *social consumption tax* (f=1), *washing* (f=1).

As it happens to the English adjective *GREEN* 'green', also the English noun *ECONOMY* 'economy' is recognized as a statistical keyword for the SusCorp (Italian)

while the Italian equivalent *ECONOMIA* 'economy' is not. This happens in spite of the English noun appearing only 93 times in the subcorpus, while the Italian noun occurs 326 times. The former stands out in a statistical way because it is rarer to find *ECONOMY* 'economy' in Italian compared to *ECONOMIA* 'economy'.

The economic dimension of sustainable development is reflected also by the lemma *PIL* 'GDP'. *PIL* 'GDP' refers to Gross Domestic Product, being it Italian or of other countries. Even the social dimension of sustainable development is represented among the keywords through the lemmas *DISEGUAGLIANZA* 'inequality' and *POVERTÀ* 'poverty'.

The aboutness of the Italian section of the Sustainable development Corpus is further described by the statistical keywords that are calculated by comparing the SusCorp (Italian) and the Italian section of the Timestamped JSI webcorpus 2014-2020. The only keyword improving the aboutness of the SusCorp (Italian) is the adjective *sociale* 'social'.

Among the statistically significant lexemes of the SusCorp (Italian) some cultural keywords can be identified. In addition to the lexemes referring to sustainable development that have been emphasised when studying the most frequent lexemes of the SusCorp (Italian), namely *sostenibile* 'sustainable', *sviluppo* 'development', *obiettivi* 'goals', *sostenibilità* 'sustainability', *GREEN* 'green' can be deemed a cultural keyword for the Italian quality press because of its standing out among the abundant statistical keywords referring to environmental issues. *GREEN* 'green' is regarded as a cultural keyword because it mirrors and shapes an aspect of environmental concerns, which are one of the most urgent political, social and cultural issues of the 2010s in Italy. At the same time, also the fact that the lexeme appears in its English form is peculiar for the Italian culture (Antelmi 2018).

## 5.2. Meaning by collocation of *sustainable* in the *SusCorp*

As it has already been observed, the English adjective *sustainable* and its Hungarian and Italian translational equivalents (i.e. *FENNTARTHATÓ* for Hungarian and *SOSTENIBILE* for Italian) are esteemed as cultural keywords for the Sustainable development Corpus. This surely follows from corpus design, as the SusCorp is planned so that every article counts at least one occurrence of the lexical item *sustainable development*, but it also depends on the importance of sustainability in the British, Hungarian and Italian news discourse of the second half of 2010s.

The semantics of the English adjective *sustainable* and of its Hungarian and Italian translational equivalents in the SusCorp are traced by means of collocation in the following part of the study: the statistically significant co-occurrence of lexemes with *sustainable* and its Hungarian and Italian equivalents hints at the lexico-grammatical company that the adjective keeps; this lexico-grammatical company is further emphasised by means of colligation, namely the tendency of a node to occur in specific grammatical patterns, and semantic preference, namely the tendency of a node to occur in specific semantic environments; collocation, colligation and semantic preference contribute to the establishment of evaluative prosody, which assesses whether the adjective acquires a positive, neutral or negative connotation as a consequence of its collocational patterns.

The collocational patterns of the English, Hungarian and Italian adjectives are sketched with the aid of the Sketch Engine and of the #LancsBox's GraphColl tool. First, the collocational behaviour of the nodes is outlined by means of word sketches. Word sketches summarise the most important collocational patterns of a node and they organise them according to underlying grammatical relationships. Word sketches are extracted on the Sketch Engine platform with the LogDice association measure, by setting the statistical threshold to 6.0, the minimum frequency threshold to 20, and by indicating that the searched lexeme should be treated as a lemma and that it belongs to the word class of the adjectives. The word sketch of the English *sustainable* is searched retrieved by typing the string `sustainable`, the word sketch of the Hungarian *FENNTARTHATÓ* is found with the string `fenntartható`, the word sketch of the Italian *SOSTENIBILE* is extracted with the string `sostenibile`.

Second, collocation networks are found with #LancsBox's GraphColl. Collocation networks allow to visualise the collocational patterns of a node, and according to Brezina (2018b: 273) they can be "useful visual summaries of key meaning relationships in the data pointing to crucial aspects of the semantic structure of a text or corpus". Collocation networks add details to the meaning by collocation outlined with the word sketches. Collocation networks are retrieved with a collocation window ranging from one to five words to the left and to the right of the node; they are calculated with the Z statistical measure, set on a statistical threshold of 10.0, and by allowing only collocates that appear at least 20 times both independently in the corpus and in association with the node. Collocational patterns are extracted from the SusCorp (English) by querying the subcorpus in GraphColl with the string `/sustainable/` because of the adjective's being invariable. Collocational patterns are built from the SusCorp (Hungarian) by querying the subcorpus in GraphColl through

the string /fenntartható\*/ with the use of the wildcard \*, which allows to retrieve all forms of a lemma. Collocational patterns are found in GraphColl for the SusCorp (Italian) with the string /sostenibile|sostenibili/, which encompasses both the singular and the plural forms of the adjective.

Collocational patterns are labelled as it has been done when analysing the 2030 Agenda Corpus (see § 4.2 on this point). Labels mention first the association measure utilised for the calculation of collocation, then the statistical threshold set to identify the strongest collocates, followed by the span and by the minimum frequency of co-occurrence of node and collocate and by the minimum frequency of occurrence of the sheer collocate in the corpus.

The collocational patterns found with word sketches and with collocation networks are investigated further by reading their concordance lines with AntConc's Concordance tool.

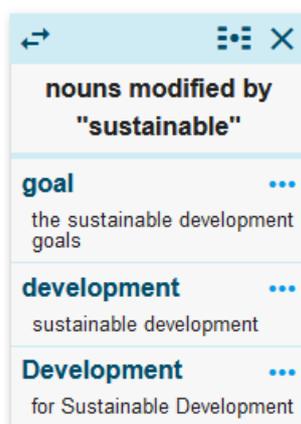
### **5.2.1. The English *SUSTAINABLE***

The study of the meaning by collocation of *sustainable* in the Sustainable development Corpus starts with the analysis of the collocational patterns of the English adjective *sustainable* in the SusCorp (English).

The adjective *sustainable* appears 931 times in the English section of the Sustainable development Corpus: in other terms, it occurs on average 1.86 times per article.

#### **5.2.1.1. Word sketch**

The meaning by collocation of the English adjective *sustainable* is traced first thanks to the word sketch that can be extracted from Sketch Engine. The word sketch retrieved for the adjective is reproduced in Figure 30, and it showcases the most peculiar collocational patterns of *sustainable*.



**Figure 30.** Word sketch of *SUSTAINABLE* in the SusCorp (English) – LogDice(6.0), NC20-C20.

As it can be seen in Figure 30, the nouns that are modified by *sustainable* the most are *goal* and *development*. *sustainable* appears 647 times together with the noun *development*, and thus the pattern *sustainable development* makes up for 70% of the occurrences of the adjective. This is taken into consideration when analysing the collocational patterns of *sustainable*. The noun *goals* is almost always modified by *sustainable* in the noun phrase *sustainable development goals* ( $f=273$ ) and it is found in the noun phrase *sustainable goals* only 4 times. Thus, *sustainable development* is associated with the SDGs through the noun phrase *sustainable development goals* in roughly 42% of its occurrences.

Apart from *goals* and *development*, the adjective *sustainable* left-modifies several other nouns and noun phrases. These nouns and noun phrases refer first to the world's future and to the path that the world should take aiming at sustainability, as with the lexemes *FUTURE* ( $f=12$ ), *growth* ( $f=9$ ), *way* ( $f=5$ ). Other nouns or noun phrases modified by *sustainable* refer to economic and environmental matters, namely to the economic and environmental dimensions of sustainable development. When they refer to economic issues, they can either be about economy in general (*ECONOMY*,  $f=6$ ) or they can hint at specific fields of economy. These specific fields of economy are finance (*investing*,  $f=16$ ; *INVESTMENT*,  $f=8$  occurrences), and management and business (*BUSINESS*,  $f=9$ ; *COMPANY*,  $f=6$ ).

At the crossroads between economy and environment, *sustainable* is associated with nouns and noun phrases related to agriculture and farming via the lexemes *palm oil* (or *palm-oil*;  $f=8$ ) and *agriculture* ( $f=6$ ). Linked to some extent to agriculture and farming, natural elements are written to undergo *sustainable use* ( $f=7$ ) and *sustainable management* ( $f=5$ ). *sustainable use* is modified by prepositional phrases like *of terrestrial ecosystems* ( $f=2$ ), *of land and other natural resources* ( $f=1$ ), *of the sea* ( $f=1$ ), *of the world's oceans* ( $f=1$ ), and *of water* ( $f=1$ ). Terrestrial and marine environments and elements are equally

said to be employed in a sustainable way. The same happens for the use of *sustainable management*. *sustainable management* is modified by the prepositional phrases of *natural resources* (f=2) and *of water and sanitation for all* (f=2) and by the prepositional phrase of *forests* (f=1). Despite the alleged sustainable management and sustainable use of natural resources, natural resources suffer *consumption and production* (f=4 occurrences) with the alternative form *production and consumption* (f=1 occurrence), but also *catering* (f=1). At the interplay between environment, economy and society, the adjective *sustainable* modifies also a lexeme associated with energy issues, namely *energy* (f=5).

The nouns and noun phrases that *sustainable* modifies suggest that the meaning by collocation of the adjective is tied especially to the economic and environmental dimensions of sustainable development, although the social dimension of sustainability is not totally neglected. This is further stressed by the collocation networks that surround the adjective from 1 to 5 words to the left and to the right of the adjective.

### 5.2.1.2. Collocation networks

The meaning by collocation outlined through word sketches for the English adjective *sustainable* in the SusCorp (English) is further expanded on thanks to the retrieval of the adjective's collocation networks.

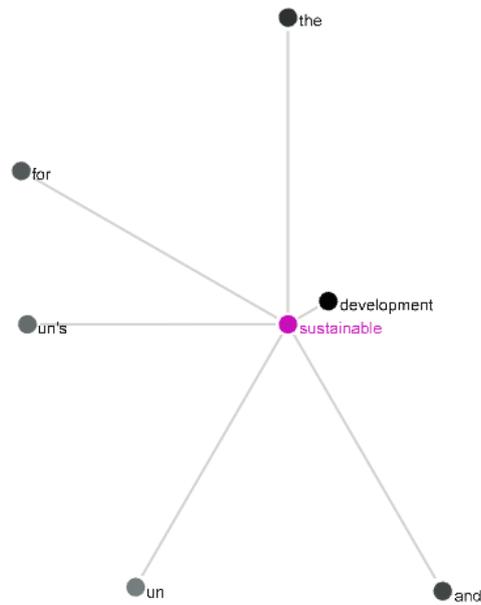
The list of the collocates retrieved for *sustainable* with  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word collocation windows can be seen in Table 39. The collocates' lists endowed with the collocations' statistical values, with the position of the collocate in relation to the node, with the frequency of co-occurrence of node and collocate and with the frequency of occurrence of the sole collocate in the corpus can be read in Tables 64, 65, 66 in the Appendix.

Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>development</i>	<i>development</i>	<i>development</i>
2	<i>un's</i>	<i>goals</i>	<i>goals</i>
3	<i>un</i>	<i>un's</i>	<i>un's</i>
4	<i>the</i>	<i>sdgs</i>	<i>the</i>
5	<i>for</i>	<i>the</i>	<i>sdgs</i>
6	<i>and</i>	<i>goal</i>	<i>and</i>
7		17	<i>goal</i>
8		<i>agenda</i>	<i>for</i>
9		<i>and</i>	<i>adopted</i>
10		<i>investing</i>	17

11	<i>for</i>	2030
12	<i>capital</i>	<i>to</i>
13	2030	<i>agenda</i>
14	<i>of</i>	<i>of</i>
15	<i>un</i>	<i>achieve</i>
16	<i>to</i>	<i>investing</i>
17	<i>on</i>	<i>a</i>
18	<i>a</i>	<i>capital</i>
19	<i>business</i>	<i>un</i>
20	<i>in</i>	<i>on</i>
21	<i>at</i>	<i>in</i>
22	<i>new</i>	<i>global</i>
23	<i>more</i>	<i>is</i>
24	<i>global</i>	<i>growth</i>
25		<i>new</i>
26		<i>business</i>
27		<i>by</i>
28		<i>water</i>
29		<i>at</i>
30		<i>will</i>
31		<i>world</i>
32		<i>which</i>
33		<i>investment</i>
34		<i>be</i>
35		<i>more</i>
36		<i>that</i>
37		<i>as</i>
38		<i>said</i>
39		<i>are</i>
40		<i>we</i>
41		<i>with</i>
42		<i>it</i>
43		<i>says</i>
44		<i>not</i>
45		<i>its</i>
46		<i>this</i>
47		<i>has</i>

**Table 39.** List of collocates of *SUSTAINABLE* in the SusCorp (English).

Figure 31 depicts the collocation network extracted for the adjective *sustainable* with #LancsBox and employing a collocation window of  $\pm 1$  words. As it can be seen in Figure 31 and in Table 39, with a span of  $\pm 1$  words, *sustainable* collocates with the content words *development*, *UN's* and *UN*, and with the function words *the*, *for* and *and*. All collocates but *development* are usually found on the left-hand side of the node.



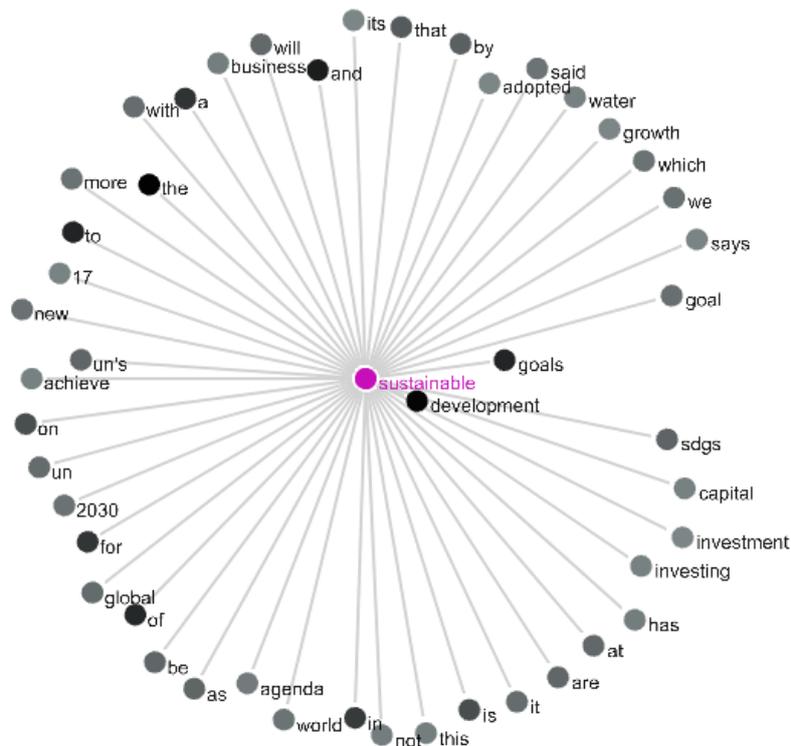
**Figure 31.** Collocation network of *SUSTAINABLE* in the SusCorp (English) – Z(10.0), 1L-1R, NC20-C20.

Between three words to the left and three words to the right, *sustainable* collocates with the content words *goals, SDGs, goal, agenda, investing, capital, business, new, more* and *global*, and with the function words *17, 2030, of, to, on, a, in* and *at*. These complement the content and function words that have already emerged with a  $\pm 1$ -word span, namely *development, UN's, UN, the, for* and *and*. As it can be noticed in Figure 32, the content words *agenda, business, new, more, global* mainly stand on the left-hand side of *sustainable*, and the content words *goals, SDGs, goal, investing* and *capital* usually stand on the right-hand side of the node. The function words *17, 2030, of, to, on, a* and *in* usually appear on the left-hand side of *sustainable*, the function word *at* normally occurs on the right-hand side of the node.



**Figure 32.** Collocation network of *SUSTAINABLE* in the SusCorp (English) – Z(10.0), 3L-3R, NC20-C20.

Between five words to the left and five words to the right, *sustainable* collocates with the same content and function words that have already been spotlighted with the  $\pm 1$ - and  $\pm 3$ - word spans. The adjective also collocates with the content words *adopted*, *achieve*, *growth*, *water*, *world*, *investment*, *said* and *says* and with the function words *by*, *will*, *which*, *that*, *as*, *we*, *with*, *it*, *not*, *its* and *this*. Halfway between content words and function words are the collocates *is*, *be*, *are*, and *has* which can be used both as lexical verbs and as auxiliary verbs. As it can be observed in Figure 33, the content words *achieve* and *world* tend to appear on the left-hand side of *sustainable*, whereas the content words *adopted*, *growth*, *water*, *investment*, *said*, *says* tend to stand on the right-hand side of the node. The function words *will*, *be*, *as*, *with* tend to occur on the left-hand side of *sustainable*, the function words *is*, *by*, *which*, *that*, *are*, *we*, *it*, *not*, *this*, *has* tend to appear on the right-hand side of the node, while *its* stands equally on the left- and on the right-hand side of the adjective.



**Figure 33.** Collocation network of *SUSTAINABLE* in the SusCorp (English) – Z(10.0), 5L-5R, NC20-C20.

**Nouns.** The nouns collocating with the adjective *sustainable* within a  $\pm 1$ -word collocation window introduce the most salient semantic features of the adjective in the Sustainable development Corpus. As it has been noticed also when analysing the word sketch in Figure 30, the strongest collocate of the adjective *sustainable* is the noun *development*. Apart from *development*, the other nouns that the adjective *sustainable* collocates with within a  $\pm 1$ -word span are *UN's* and *UN*. *UN's* contributes to the rise of collocational patterns like *UN's sustainable development goals* ( $f=30$ ) with its singular form *UN's sustainable development goal* ( $f=4$ ), but also less frequent patterns like *UN's sustainable development plans* ( $f=1$ ), *UN's sustainable development solutions network* ( $f=1$ ), *UN's sustainable development target* ( $f=1$ ), and *UN's sustainable energy for all* ( $f=1$ ). *UN* contributes to the collocational patterns *UN sustainable development goals* ( $f=13$ ) with its singular form *UN sustainable development goal* ( $f=3$ ), *UN sustainable development summit* ( $f=3$ ), *UN sustainable development solutions network* ( $f=2$ ), *UN sustainable development targets* ( $f=2$ ), and *UN sustainable development agenda* ( $f=1$ ). All these patterns show that in the SusCorp (English) the strongest collocational patterns of *sustainable* involve sustainable development and that sustainable development is referred to especially in relation to the United Nations and to the 2030 Agenda's SDGs.

This observation is strengthened when analysing the collocates emerging for *sustainable* within a  $\pm 3$ -word span. Between three words to the left and three words to

the right, *sustainable* collocates with other nouns that explicitly refer to the 2030 Agenda for Sustainable Development and to its 17 Sustainable Development Goals. For instance, the adjective *sustainable* collocates with the noun *agenda* and with the numeral 2030 in the patterns *2030 Agenda for Sustainable Development* ( $f=11$ ), *2030 Agenda on Sustainable Development* ( $f=1$ ), *2030 Sustainable Development Agenda* ( $f=1$ ), *2030 UN Sustainable Development Agenda* ( $f=1$ ), and *Sustainable Development Agenda 2030* ( $f=1$ ). Moreover, *sustainable* collocates also with the nouns *goals*, *sdgs*, *goal*, and together with the collocate 17 it belongs to the patterns *sustainable development goals* ( $f=273$ ; 49 of these occurrences are completed in the form *sustainable development goals (SDGs)*), *sustainable development goal* ( $f=31$ ), *17 sustainable development goals* ( $f=16$ ), *sustainable goals* ( $f=2$ ), *17 global goals for sustainable development* ( $f=1$ ), *17 UN sustainable development goals* ( $f=1$ ), *global sustainable energy goals* ( $f=1$ ), and *new Global Goals for Sustainable Development* ( $f=1$ ).

Within a  $\pm 3$ -word span, the adjective *sustainable* collocates also with nouns that have to do with economy. These are *investing*, *capital* and *business*. *sustainable* meaningfully co-occurs with *investing* in the patterns *sustainable investing* ( $f=13$ ) and *sustainable and impact investing* ( $f=4$ ). It mainly collocates with *capital* in the pattern *Sustainable Development Capital* ( $f=17$ ). It co-occurs with *business* in the patterns *sustainable business* ( $f=7$ ), *World Business Council for Sustainable Development* ( $f=5$ ), *Business and Sustainable Development Commission* ( $f=4$ ), *business and the sustainable development goals* ( $f=2$ ), *business diplomacy for sustainable development* ( $f=1$ ). Of these economic terms, *capital* and *business* are strongly intertwined with sustainable development.

Between five words to the left and five words to the right, among the nouns that *sustainable* collocates with in the SusCorp (English), *growth* mirrors one of the core tenets of sustainability, namely sustainable growth. The noun phrase *sustainable growth* makes up for 45% of the co-occurrences of *sustainable* and *growth* (9 out of 20). *sustainable growth* is often found in postmodifier or adjunct prepositional phrases, or it can function as a goal in material processes. In the remaining cases, noun phrases including *sustainable* are associated with lexical items like *economic growth* ( $f=5$ ) or *green growth* ( $f=1$ ), but they can also combine with the growth of specific places (as in *Bogota's growth*,  $f=1$ ).

Within the  $\pm 5$ -word collocation window, *sustainable* collocates with *water* in a wide variety of patterns. The most frequent patterns have to do with the access, management and use of water in the light of sustainable development: *availability and sustainable management of water and sanitation for all* ( $f=2$ ), *sustainable access to water and sanitation* ( $f=2$ ), and *sustainable development in water and sanitation* ( $f=2$ ); *sustainable access*

to safe drinking water and basic sanitation (f=1), sustainable approach to water management (f=1), sustainable development of water and sanitation access (f=1), sustainable development of water resources manager (f=1), sustainable use of water (f=1), sustainable water management (f=1), and sustainable water program manager (f=1). As in the case of *growth*, also the collocation *sustainable-water* is usually included in clauses characterized by material processes like *ensure* (f=4) and *achieve* (f=3). These patterns show that water contributes to the issue of sustainability by requiring to be handled in a sustainable way and to be dealt with in association with sanitation for the wellbeing of the world's peoples.

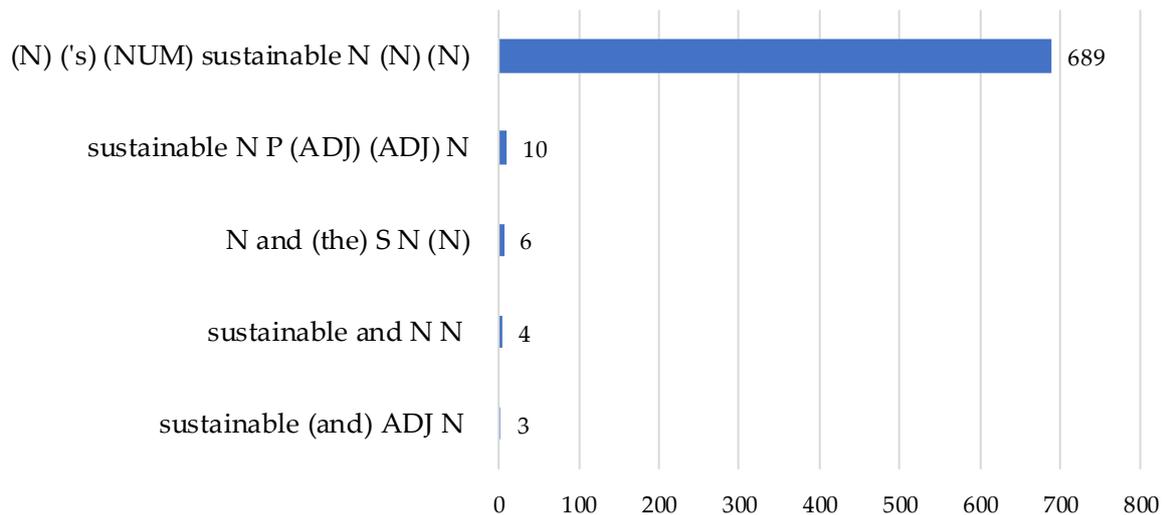
Also *sustainable* and *world* collocate in a wide variety of patterns. The most frequent collocational patterns of the two lexemes are *World Business Council for Sustainable Development* (f=5) and *sustainable world* (f=4). The other collocations do not find correspondences one with the other, but they bestow an aura of internationality to the occurrences of *sustainable* in the SusCorp (English).

*sustainable* and *investment* collocate in patterns that enhance the semantic preference for economic issues that has already been described for the ±3-word span. The most frequent collocational pattern binding *sustainable* and *investment* is *sustainable investment* (10 occurrences out of 20). The other patterns where *sustainable* and *investment* meaningfully co-occur are: *new investment for sustainable growth* (f=1), *the "green" investment fund Sustainable Development Capital* (f=1), and *the Sustainable Development Investment Partnership* (f=1). These patterns highlight the need for investing towards sustainability.

When the adjective *sustainable* collocates with nouns, it modifies noun phrases in the following colligational patterns:

(N)	('s)	(NUM)	<i>sustainable</i>	N	(N)	(N)		
			<i>sustainable</i>	<i>and</i>	N	N		
			<i>sustainable</i>	( <i>and</i> )	ADJ	N		
N	<i>and</i>	( <i>the</i> )	<i>sustainable</i>	N	(N)			
			<i>sustainable</i>	N	P	(ADJ)	(ADJ)	N

These colligational patterns are displayed also in Graph 20 together with their frequency in the SusCorp (English).



**Graph 20.** The colligational patterns of *SUSTAINABLE* and nouns in the SusCorp (English).

As it can be seen in Graph 20, the most common colligational pattern binding *sustainable* and nouns involves the adjective in a noun phrase that can be left-modified by a noun phrase eventually used in the possessive form (e.g. *UN's sustainable development*) and that can itself left-modify another noun phrase (e.g. *sustainable development goals*).

The available slots of these colligational patterns are filled with nouns that tend to cluster around a robust semantic preference for aspects concerning the 2030 Agenda (*development, goals, UN's, SDGs, goal, agenda, UN*). In addition, *sustainable* features semantic preferences for economic issues (*investing, capital, business and investment*) in an international dimension (*world*) and a sparse lexical preference for a generic improvement (*growth*) and for one of the environment elements that debates on sustainability focus on, namely *water*.

**Adjectives.** Within a  $\pm 3$ -word span, the adjective *sustainable* collocates with the adjective *new* in patterns like *new sustainable development goals* ( $f=9$ ), *new standards for sustainable development* ( $f=2$ ) and *new sustainable development agenda* ( $f=2$ ), and *new climate and sustainable development* ( $f=1$ ), *new investment for sustainable development* ( $f=1$ ), *new policy and sustainable development scenarios* ( $f=1$ ), *new sustainable development consensus* ( $f=1$ ), *new sustainable development goal* ( $f=1$ ), *new UN sustainable development goals* ( $f=1$ ), and *new UN sustainable development targets* ( $f=1$ ). In some of these patterns the adjective *new* modifies a noun phrase including the adjective *sustainable*, whereas in others *new* modifies a noun phrase which is associated with another noun phrase

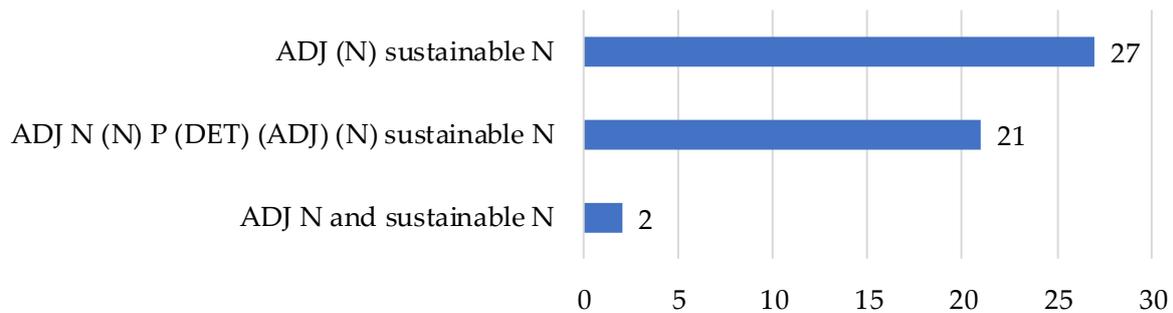
including *sustainable*. These associations are created with conjunctions or with prepositions. The new sustainable development goals and targets that the SusCorp (English) refers to are the goals and targets that resulted from the release of the 2030 Agenda for Sustainable Development and its 17 Sustainable Development Goals. The 2030 Agenda and the 17 SDGs substituted the Millennium Development Goals posited in the previous United Nations Millennium Declaration. The collocate *new*, in fact, appears in articles published short after the issue of the 2030 Agenda.

Between three words to the left and three words to the right, *sustainable* collocates with the adjective *global* in the patterns *global sustainable development goals* (f=4), *global sustainable development* (f=2), *global sustainable development director* (f=1), *Global Sustainable Development Lecturer* (f=1), *global sustainable energy goals* (f=1), *global sustainable finance agenda* (f=1), *Global Sustainable Investment Review* (f=1), and *Global Sustainable Seafood Initiative* (f=1). In these cases, sustainable development, energy, finance, investments and the handling of seafood are written to regard the whole planet thanks to their being qualified by the attribute *global*. *sustainable* collocates with *global* also in the following patterns: *global goals for sustainable development* (f=2), *global partnership for sustainable development* (f=2), *global campaign supporting Sustainable Development Goal #5* (f=1), *global criteria for sustainable homes* (f=1), *global head of sustainable finance* (f=1), and *global movement of sustainable companies* (f=1). In these instances, *global* does not modify something which is deemed sustainable: what is written to be sustainable qualifies a global enterprise.

When *sustainable* collocates with an adjective, it participates in the following colligational patterns:

ADJ	N	(N)	P	(DET)	(ADJ)	(N)	<i>sustainable</i>	N
					ADJ	(N)	<i>sustainable</i>	N
			ADJ	N	<i>and</i>	<i>sustainable</i>	N	N

The frequency of these colligational patterns in the SusCorp (English) can be found in Graph 21.



**Graph 21.** The colligational patterns of *SUSTAINABLE* and adjectives in the SusCorp (English).

As it can be noted in Graph 21, the most frequent colligational pattern linking *sustainable* with adjectives witnesses *sustainable* belonging to a noun phrase modified by the pinpointed adjectives.

The adjectives that *sustainable* collocates with shape a semantic preference for internationality (*global*) and novelty (*new*).

**Adverbs.** With a  $\pm 3$ -word span, the adjective *sustainable* collocates with the adverb *more*. The collocate *more* usually precedes *sustainable* in the adjective phrase *more sustainable* (16 occurrences out of 28). *more sustainable* modifies nouns and noun phrases like *future* ( $f=2$ ), but also *builds* ( $f=1$ ), *economic*, *social and environmental path* ( $f=1$ ), *economy* ( $f=1$ ), *employment* ( $f=1$ ), *food production* ( $f=1$ ), *homes* ( $f=1$ ), *investments* ( $f=1$ ), *jobs* ( $f=1$ ), *planet* ( $f=1$ ), *practices* ( $f=1$ ), and *world* ( $f=1$ ). The use of the comparative form *more sustainable* as an attribute implies that the entities represented by the nouns or noun phrases modified by the adjective phrase initially lie in a non-sustainable or non-fully sustainable condition and that this condition needs to be improved towards sound sustainability. The entities that need to undergo this process towards a more robust sustainability range from economic matters to infrastructural issues, from social matters like food production and employment to a general condition of the planet. *more sustainable* is paired in two instances with the adjective phrase *better-paid* via coordination (in the adjective phrase complex *more sustainable and better-paid*,  $f=1$ , with the alternative form *more sustainable and better paid*,  $f=1$ ) and to other adjectives by juxtaposition. In this case, it is found in the following adjective phrases: *more effective and sustainable* ( $f=1$ ), *more equal and sustainable* ( $f=1$ ), *more equitable, sustainable* ( $f=1$ ), *more equitable and sustainable* ( $f=1$ ), and *more sustainable, prosperous and equitable* ( $f=1$ ). In all these cases, sustainability is mainly associated with social equity.

**Verbs.** Within a  $\pm 5$ -word collocation window, the adjective *sustainable* collocates with the verbs *adopted*, *achieve*, *said* and *says*. When *sustainable* collocates with the verb *adopted*, the adjective usually belongs to a noun phrase that participates in an event as a goal of the material process of adopting. The actor of the material processes of adopting is usually the United Nations, but it can also be *world leaders* ( $f=1$ ), or it can be unexpressed in the sentence and retrievable only in a larger context. The process is employed both in its active and in its passive forms.

*sustainable* usually collocates with the verb *achieve* in the pattern *achieve the sustainable development goals* (10 occurrences out of 23). The pattern is frequently employed in infinitive clauses with an adverbial function or after auxiliary verbs.

The collocation binding the adjective *sustainable* and the verb forms *said* and *says* reflects the specificity of the discourse type represented by the SusCorp (English), namely news discourse: the co-occurrences of *said* and *says* are most frequently used to report speech that includes the adjective *sustainable*. Speech is reported directly 23 times and indirectly only 5 times. The collocations *sustainable-said* and *sustainable-says* point out that not only is sustainability something to act for, as it emerges from the statistically significant presence of the material processes *adopted* and *achieve* in the list of *sustainable's* collocates, but also something to report and discuss. This reporting happens more frequently in the past with the verb form *said* ( $f=33$ ) and less frequently in the present with the verb form *says* ( $f=20$ ). In addition, *sustainable* collocates with *said* and *says* also in clauses where the adjective qualifies the noun phrase that functions as subject of the reporting verb. This happens 18 times. In this case, the adjective can either describe the title of the expert who is interviewed in the article, or it can function as an attribute.

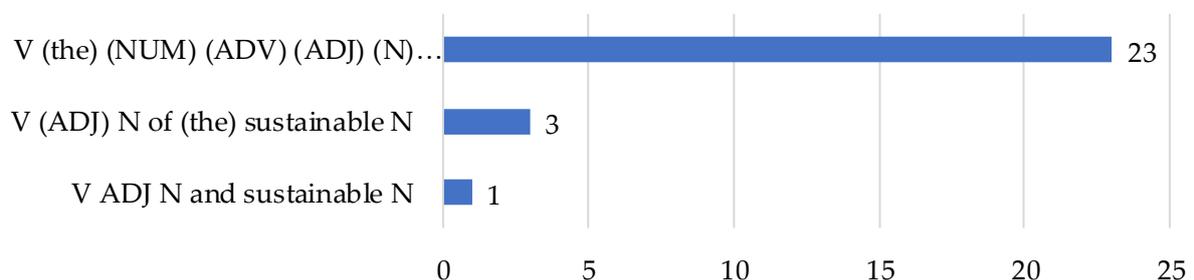
With a  $\pm 5$ -word span, *sustainable* collocates with a single auxiliary verb (i.e. *will*). This indicates that the most frequent and statistically salient modality of the SusCorp's (English) articles is one of reality set across past, present and future tenses. The past tense is represented by the verb forms *adopted*, *said* and *has*; the present tense marks the verb forms *achieve* (whenever it is used in a finite mode), *says*, *is* and *are*; the future is encoded with *will*. The present is the most represented tense. When *sustainable* collocates with *will*, the auxiliary verb precedes lexical verbs that can be mostly classified as material processes: *achieve* ( $f=2$ ), *promote* ( $f=2$ ), *require* ( $f=2$ ), *set* ( $f=2$ ), *bring* ( $f=1$ ), *command* ( $f=1$ ), *deliver* ( $f=1$ ), *do* ( $f=1$ ), *form* ( $f=1$ ), *fuel* ( $f=1$ ), *help* ( $f=1$ ), *maintain* ( $f=1$ ), *provide* ( $f=1$ ), *put* ( $f=1$ ) and *undermine* ( $f=1$ ). Relational processes are represented by *be* ( $f=6$ ) and there is a single instance of mental process (*need*,  $f=1$ ).

When *sustainable* collocates with the verb *is*, the verb is employed both as an auxiliary and as a lexical verb. When it is used as a lexical verb, *is* qualifies noun phrases or adjective phrases that include *sustainable* as: *a disguised attempt to impose global communism (f=1)*, *a key objective (f=1)*, *a “utopian, socialist nightmare” (f=1)*, *clear (f=1)*, *devoted to partnerships that build capacity and speed implementation (f=1)*, *fast becoming mainstream (f=1)*, *“mission impossible” (f=1)*, *not just another box to tick (f=1)*, *not just one that does good for others (f=1)*, *particularly important (f=1)*, and *to provide everybody on the plant with a legal identity by 2030 (f=1)*. At the same time, *is* qualifies as *sustainable* the following noun phrases: *DfID and USAid (f=1)*, *growth (f=1)*, *relying on donors for money (f=1)*, *the objective (f=1)*, and *the outcome (f=1)*. The collocates *be* and *are* behave similarly. *has* tends to function as an auxiliary verb and it is used as a lexical verb only in few occurrences.

When *sustainable* collocates with purely lexical verbs, it is found in the following colligational patterns:

V	(the)	(NUM)	(ADV)	(ADJ)	(N)	<i>sustainable</i>	N
V	(ADJ)	N	<i>of</i>	(the)	<i>sustainable</i>	N	N
	V	ADJ	N	<i>and</i>	<i>sustainable</i>	N	N

The frequency of these colligational patterns can be seen in Graph 22.



**Graph 22.** The colligational patterns of *SUSTAINABLE* and lexical verbs in the SusCorp (English).

As it can be observed in Graph 22, the most common colligational pattern connecting *sustainable* and lexical verbs witnesses *sustainable* modifying a noun phrase that functions as goal of a transitive verb.

The verb slots available in these colligational patterns are characterised by a semantic preference for material processes of achieving (*adopted* and *achieve*).

**Prepositions.** Within a collocation window of  $\pm 1$  word, *sustainable* collocates with the preposition *for*. *sustainable* is preceded by the preposition *for* in prepositional phrases that usually function as post-modifiers of nouns or noun phrases. These prepositional phrases are endowed with a semantics of goal. The noun phrase that modifies the prepositional head *for* and that includes *sustainable* is *sustainable development* in most of the cases (56 out of 66). The other occurrences feature noun phrases like *sustainable growth* ( $f=2$ ), *Sustainable Futures* ( $f=1$ ), *sustainable homes* ( $f=1$ ), *sustainable investing* ( $f=1$ ), *sustainable investments* ( $f=1$ ), *sustainable lifestyles* ( $f=1$ ), *sustainable solutions* ( $f=1$ ), *Sustainable Transport and Communications* ( $f=1$ ), and *sustainable use of the world's oceans* ( $f=1$ ).

Between three words to the left and three words to the right, *sustainable* collocates with the prepositions *of*, *to*, *on*, *in* and *at*. In most cases, *sustainable* belongs to a noun phrase that modifies the prepositions in prepositional phrases. In other cases, *sustainable* belongs to a noun phrase that is post-modified by a prepositional phrase having the aforementioned prepositions as head. When the adjective is included in the prepositional phrase, these prepositional phrases can function as postmodifiers or they can be adjuncts.

With a  $\pm 5$ -word span, among the function words that *sustainable* collocates with, *by* and *with* are the heads of prepositional phrases that include the adjective, or that modify complex noun phrases including the adjective. In addition, they frequently form prepositional phrases that operate as adjuncts for the clause that *sustainable* belongs to.

**Determiners.** Between one word to the left and one word to the right, *sustainable* collocates with the definite article *the*. The definite article *the* precedes *sustainable* in 200 cases and it follows it in 2 cases. When the definite article precedes the adjective, it modifies the noun phrase including *sustainable* by assigning it a certain degree of definiteness, as in the most frequent noun phrase *the sustainable development* ( $f=186$ ).

With a  $\pm 3$ -word span, *sustainable* collocates with the indefinite article *a*. The indefinite article *a* is used to assign an aura of indefiniteness or novelty to the noun phrases that *sustainable* belongs to, as in *a sustainable future* ( $f=7$ ), *a sustainable company* ( $f=4$ ), *a sustainable approach* ( $f=2$ ), *a sustainable way* ( $f=2$ ), *a sustainable and continuous service* ( $f=1$ ), *a more sustainable economy* ( $f=1$ ), and *a healthy and sustainable environment* ( $f=1$ ).

**Pronouns.** Between five words to the left and five words to the right, when co-occurring with *sustainable*, the collocates *which* and *that* usually introduce relative clauses that modify a phrase that includes the adjective, or they introduce relative clauses that include the adjective. In addition, *that* is used for object clauses that encompass *sustainable*.

When *sustainable* collocates with the personal and demonstrative pronouns and adjectives *it*, *its* and *this*, these lexical items usually function as cohesive devices. Behaving as cohesive devices, *it*, *its* and *this* point at textual referents that appear before or after them and they contribute to shaping the texts so that the adjective *sustainable* is perfectly integrated in the flow of discourse.

*sustainable* collocates with the personal pronoun *we* to signal some of the participants in the endeavour towards sustainable development. The personal pronoun replaces the following groups: a generic *we* that frequently stands for the impersonal *there* ( $f=12$ ); the United Kingdom's businesses, entrepreneurs, and scientists ( $f=4$ ); newspaper journalists ( $f=3$ ); United Kingdom's citizens ( $f=2$ ); Unilever (i.e. a multinational consumer goods company that tries to meet daily needs for personal care, hygiene and nutrition;  $f=2$ ); the United Nations ( $f=1$ ); the United Kingdom's politicians ( $f=1$ ); a global panel on the sustainable development goals ( $f=1$ ); members of the University of Brighton ( $f=1$ ), members of the University of Plymouth ( $f=1$ ), and members of Nottingham Trent University ( $f=1$ ); UBS Group (i.e. an investment bank and financial services company;  $f=1$ ); Kiron Open Higher Education (i.e. an education model that addresses refugees so that they can access higher education and successful learning with digital methods;  $f=1$ ); *The Lancet journal* and Australia's Lowitja Institute (i.e. an organisation cooperating for the health and wellbeing of Australia's indigenous people;  $f=1$ ); and researchers of Stockholm Resilience Centre (i.e. an international research centre the studies social-ecological systems;  $f=1$ ). These are some of the participants to discussions on sustainable development.

**Other function words.** Between five words to the left and five words to the right, the function word *as* co-occurs with *sustainable* in several patterns, almost all sharing a common function of comparison or identification: one entity is compared or identifies with something that is qualified as *sustainable*. *as* is used as a preposition in 27 instances. In these cases, *as* is employed to describe lexical items in two possible patterns: in the first pattern, an item is endowed with the quality of sustainability by being followed by a prepositional pattern including the adjective *sustainable*; in the

second pattern, a lexical item modified by *sustainable* is qualified in relation to something else, which appears in a prepositional phrase having *as* as head. The first pattern appears 19 times, while the second occurs 8 times. Both patterns describe or identify an item by relating it to another; one of the items is endowed with the quality of sustainability and it is usually associated with lexical items encoding improvement of some sort. Sustainability is also adopted as a means of comparison when *as* is used as an adverb (in 7 cases); in these instances, *as* compares two entities through a common quality. Finally, *as* is used as a conjunction in a single sentence.

Within a  $\pm 5$ -word span, the adjective *sustainable* and the negator *not* collocate in clauses that deprive an action or an entity of the quality of sustainability. *not* negates the verbs *BE* ( $f=7$ ), *ACHIEVE* ( $f=4$ ), *MEAN* ( $f=1$ ), *PLAY* ( $f=1$ ) and *PROTECT* ( $f=1$ ). Three of these verbs encode material processes (i.e. *ACHIEVE*, *PLAY* and *PROTECT*), one encodes a relational process (i.e. *BE*), and one a mental process (i.e. *MEAN*). When the relational process *BE* is negated with *not*, the following negative associations arise: *this* and *sustainable* ( $f=1$ ); *The dependence on big institutional donors such as DfID and USAid* and *sustainable* ( $f=1$ ); *relying on donors for money* and *sustainable* ( $f=1$ ); *they* and *part of a sustainable locally-led sanitation system* ( $f=1$ ); *sustainable portfolio* and *just one that does good for others* ( $f=1$ ); *ecologically sustainable development* and *just another box to tick* ( $f=1$ ).

Summing up, the meaning by collocation of *sustainable* is described in the following way: the quality of being *sustainable* correlates with the qualities of being new and global; the nature of sustainability, however, seems to depend first and foremost on economic matters; these economic matters are treated in an international dimension under the light of the 2030 Agenda for Sustainable Development and the related 17 Sustainable Development Goals.

The collocates of the adjective display a semantic preference for sustainable development as it is depicted by the United Nations' 2030 Agenda (*development, goals, SDGs, goal, 17, 2030, agenda*), but also for international organisations and for an international dimension (*UN's, UN, global, world*). *sustainable* also has a semantic preference for economic lexemes (*investing, capital, business, investment*) and for development in general (*growth*), which can also touch upon environmental concerns like *water*. In addition, they report two material processes that feature sustainability as goal (*adopted, achieve*) and two verbal processes that treat sustainability as a matter worth discussing (*said, says*).

The lexemes representing this semantic preference participate in colligational patterns where the adjective *sustainable* modifies a noun phrase that can be used in

isolation as the argument of a verb or that can be employed in a prepositional phrase that operates as a modifier or as a circumstantial.

The connotation of the collocates, of the semantic preference, and of the colligational patterns of *sustainable* is overall neutral, slightly leaning towards positivity. According to the ecological framework adopted in the present study, the evaluative prosody of the adjective proves neutral, as the adjective *sustainable* qualifies neither detrimental nor beneficial stories of sustainability.

### 5.2.2. The Hungarian *FENNTARTHATÓ*

The semantics of the adjective *sustainable* in the Sustainable development Corpus is enhanced with the study of the collocational behaviour of the Hungarian adjective *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian).

The adjective *FENNTARTHATÓ* ‘sustainable’ occurs 851 times in the Hungarian section of the Sustainable development Corpus: it appears on average 1,70 times per article.

#### 5.2.2.1. Word sketch

The meaning by collocation of *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian) is first outlined by means of the adjective’s word sketch, which highlights the most significant grammatical relations that connect the adjective with the surrounding lexemes.

The word sketch of *FENNTARTHATÓ* ‘sustainable’ is shown in Figure 34. Figure 34 showcases the nouns that are most meaningfully modified by *FENNTARTHATÓ* ‘sustainable’ and the most common usage patterns of the Hungarian adjective.



**Figure 34.** Word sketch of *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian) – LogDice(6.0), NC20-C20.

According to the word sketch in Figure 34, the adjective *FENNTARTHATÓ* ‘sustainable’ directly or indirectly modifies the lemmas *FEJLŐDÉS* ‘development’, *CÉL* ‘goal’, *BIZOTTSÁG* ‘commission’, *ORSZÁGGYŰLÉS* ‘parliament’, *ENSZ* ‘United Nations’, *TANÁCS* ‘council’, *SZEMPONT* ‘point of view’, *ELNÖK* ‘president’, and *ÜLÉS* ‘sitting’ and it is usually found in the nominative case.

The strongest collocate of *FENNTARTHATÓ* ‘sustainable’ is *FEJLŐDÉS* ‘development’, as the two lexemes co-occur in 639 cases, covering roughly 75% of the total occurrences of the adjective. The high frequency of the lexical item *FENNTARTHATÓ FEJLŐDÉS* ‘sustainable development’ impacts on the collocational patterns that are retrieved for the adjective *FENNTARTHATÓ* ‘sustainable’: most collocational patterns included in the word sketch regard co-occurrences of *FENNTARTHATÓ FEJLŐDÉS*

‘sustainable development’ and another lexical item. In the SusCorp (Hungarian), development is referred to both with the noun *FEJLŐDÉS* ‘development’ and with the noun *FEJLESZTÉS* ‘development’. *FEJLESZTÉS* ‘development’ shares the lexical stem *fejl-* and the nominal derivational morpheme *-és* with *FEJLŐDÉS* ‘development’: they can be morphologically chunked as *FEJL-ESZT-ÉS* and *FEJL-ŐD-ÉS* respectively and they differ because of the derivational morpheme standing in the middle of the word. The derivational morpheme chosen for the slot standing between the lexical stem and the nominal morpheme builds two different verbs and consequently two different nouns: *FEJLŐDÉS* ‘development’ results from the deriving of a verb with the derivational morpheme *-ődik*, whereas *FEJLESZTÉS* ‘development’ is generated with the derivational morpheme *-eszt*.

Sustainable development is also referred to through the collocates *CÉL* ‘goal’ and *ENSZ* ‘United Nations’. The other lemmas collocating with *FENNTARTHATÓ* ‘sustainable’ in the word sketch refer to political institutions (i.e. *BIZOTTSÁG* ‘commission’, *ORSZÁGGYŰLÉS* ‘parliament’, and *TANÁCS* ‘council’), practices (i.e. *ÜLÉS* ‘sitting’), and actors (i.e. *ELNÖK* ‘president’): they aid in constructing for *FENNTARTHATÓ* ‘sustainable’ a semantics associated with politics.

### 5.2.2.2. Collocation networks

The study of the semantics of *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian), which is outlined through the reading of the adjective’s word sketch, is enhanced by means of the collocation network that surrounds the adjective.

The most significant collocational patterns of *FENNTARTHATÓ* ‘sustainable’ can be glimpsed from Table 40, which collects the collocates that are extracted with  $\pm 1$ -,  $\pm 3$ - and  $\pm 5$ -word spans. The collocates’ lists can be found also in Tables 67, 68 and 69 in the Appendix, endowed with the collocates’ position in relation to the node and with statistical and frequency values.

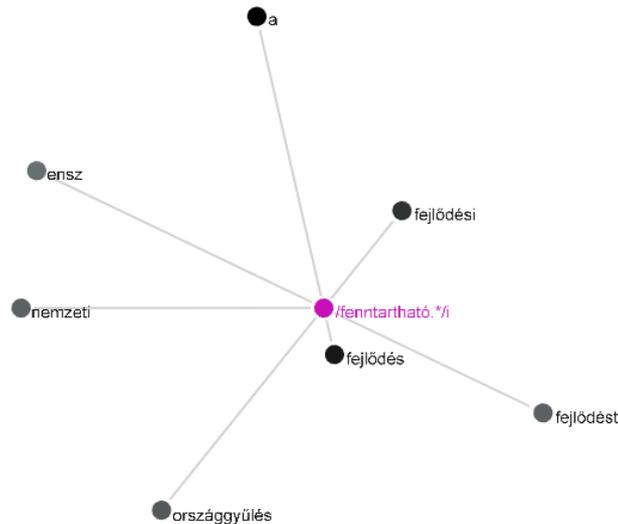
Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>fejlődés</i> ‘development’	<i>fejlődés</i> ‘development’	<i>fejlődés</i> ‘development’
2	<i>fejlődési</i> ‘development’	<i>fejlődési</i> ‘development’	<i>fejlődési</i> ‘development’
3	<i>fejlődést</i> ‘development’	<i>fejlődést</i> ‘development’	<i>a</i> ‘the’
4	<i>országgyűlés</i> ‘parliament’	<i>országgyűlés</i> ‘parliament’	<i>fejlődést</i> ‘development’
5	<i>a</i> ‘the’	<i>bizottsága</i> ‘its commission’	<i>célok</i> ‘goals’
6	<i>nemzeti</i> ‘national’	<i>célok</i> ‘goals’	<i>országgyűlés</i> ‘parliament’

7	ENSZ 'UN'	<i>a</i> 'the'	<i>bizottsága</i> 'its commission'
8		<i>bizottságának</i> 'to its commission'	<i>bizottságának</i> '(to) its commission'
9		<i>tanács</i> 'council'	<i>és</i> 'and'
10		ENSZ 'UN'	ENSZ 'UN'
11		<i>és</i> 'and'	<i>tanács</i> 'council'
12		<i>nemzeti</i> 'national'	<i>az</i> 'the'
13		<i>szempontjából</i> 'from its point of view'	<i>nemzeti</i> 'national'
14		<i>környezeti</i> 'environmental'	<i>ülésén</i> 'on its sitting'
15		<i>fejlesztési</i> 'developmental'	<i>szempontjából</i> 'from its point of view'
16		<i>az</i> 'the'	<i>hosszú</i> 'long'
17		<i>gazdasági</i> 'economic'	<i>környezeti</i> 'environmental'
18		<i>érdekében</i> 'for'	<i>gazdasági</i> 'economic'
19		<i>van</i> '(he/she/it) is'	<i>érdekében</i> 'for'
20		<i>hogy</i> 'that'	<i>szóló</i> 'about'
21			<i>fejlesztési</i> 'developmental'
22			<i>elnöke</i> 'its president'
23			<i>által</i> 'through'
24			<i>is</i> 'also'
25			<i>hogy</i> 'that'
26			<i>fontos</i> 'important'
27			<i>Magyarország</i> 'Hungary'
28			<i>világ</i> 'world'
29			<i>amelyek</i> 'that'
30			<i>kell</i> 'have to'
31			<i>valamint</i> 'as well as'
32			<i>van</i> '(he/she/it) is'
33			<i>el</i> '-'
34			<i>nem</i> 'not'
35			<i>olyan</i> 'such'
36			<i>szerint</i> 'according to'
37			<i>európai</i> 'European'
38			<i>fenntartható</i> 'sustainable'
39			<i>ki</i> '-'
40			<i>egy</i> 'a'

**Table 40.** List of collocates of *FENNTARHATÓ* 'sustainable' in the SusCorp (Hungarian).

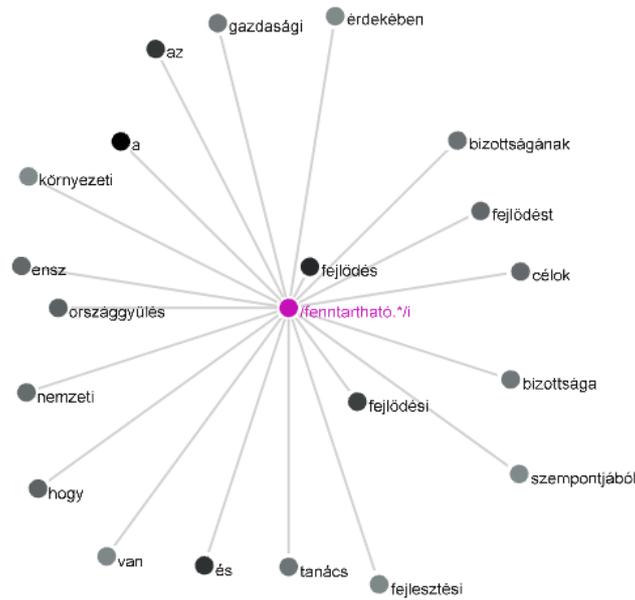
Between one word to the left and one word to the right, *FENNTARTHATÓ* 'sustainable' collocates with the content words *fejlődés* 'development', *fejlődési* 'development', *fejlődést* 'development', *országgyűlés* 'parliament', *nemzeti* 'national', *ENSZ* 'UN', and with the function word *a* 'the'. As it can be seen in Figure 35, the content words *fejlődés* 'development', *fejlődési* 'development', *fejlődést* 'development' tend to occur on the left-

hand side of *FENNTARTHATÓ* ‘sustainable’, whereas the content words *országgyűlés* ‘parliament’, *nemzeti* ‘national’, *ENSZ* ‘UN’ tend to appear on its right-hand side. The function word *a* ‘the’ tends to stand on the left-hand side of the node.



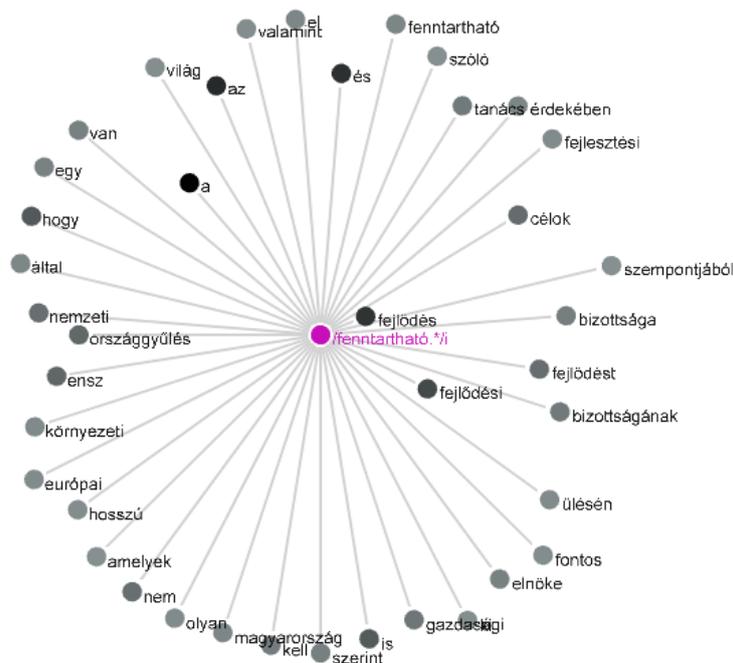
**Figure 35.** Collocation network of *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 1L-1R, NC20-C20.

Within a  $\pm 3$ -word collocation window, the Hungarian *FENNTARTHATÓ* ‘sustainable’ collocates with the content words *bizottsága* ‘its commission’, *célok* ‘goals’, *bizottságának* ‘to its commission’, *tanács* ‘council’, *szempontjából* ‘from its point of view’, *környezeti* ‘environmental’, *fejlesztési* ‘developmental’, *gazdasági* ‘economic’, *érdekében* ‘for’ and with the function words *és* ‘and’, *az* ‘the’, *van* ‘(he/she/it) is’, *hogy* ‘that’. These content and function words are added to the ones that have already been observed with a  $\pm 1$ -word span. As Figure 36 shows, the content word *környezeti* ‘environmental’ usually stands on the left-hand side of *FENNTARTHATÓ* ‘sustainable’, while the content words *bizottsága* ‘its commission’, *célok* ‘goals’, *bizottságának* ‘to its commission’, *tanács* ‘council’, *szempontjából* ‘from its point of view’, *fejlesztési* ‘developmental’, *gazdasági* ‘economic’, *érdekében* ‘for’ usually stand on its right-hand side. The function words *és* ‘and’, *az* ‘the’, *van* ‘(he/she/it) is’, *hogy* ‘that’ usually appear on the left-hand side of *FENNTARTHATÓ* ‘sustainable’.



**Figure 36.** Collocation network of *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 3L-3R, NC20-C20.

With a  $\pm 5$ -word span, the adjective *FENNTARTHATÓ* ‘sustainable’ collocates with the content words *ülésén* ‘on its sitting’, *hosszú* ‘long’, *szóló* ‘about’, *elnöke* ‘its president’, *fontos* ‘important’, *Magyarország* ‘Hungary’, *világ* ‘world’, *olyan* ‘such’, *európai* ‘European’, *fenntartható* ‘sustainable’ and with the function words *által* ‘through’, *is* ‘also’, *amelyek* ‘that’, *kell* ‘have to’, *valamint* ‘as well as’, *el* ‘-’, *nem* ‘not’, *szerint* ‘according to’, *ki* ‘-’, *egy* ‘a’. These content and function words contribute to the semantics of *FENNTARTHATÓ* ‘sustainable’ together with the content and function words that have been noted with the  $\pm 3$ - and  $\pm 5$ -word spans. As Figure 37 shows, the content words *hosszú* ‘long’, *Magyarország* ‘Hungary’, *világ* ‘world’ usually occur on the left-hand side of *FENNTARTHATÓ* ‘sustainable’, whereas the content words *ülésén* ‘on its sitting’, *szóló* ‘about’, *elnöke* ‘its president’, *fontos* ‘important’, tend to stand on its right-hand side. The function words *által* ‘through’, *amelyek* ‘that’, *kell* ‘have to’, *valamint* ‘as well as’, *el* ‘-’, *nem* ‘not’, *szerint* ‘according to’, *egy* ‘a’ tend to appear on the left-hand side of *FENNTARTHATÓ* ‘sustainable’, while the function words *is* ‘also’, *ki* ‘-’ usually stand on its right-hand side.



**Figure 37.** Collocation network of *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian) – Z(10.0), 5L-5R, NC20-C20.

As far as content words are concerned, the collocational patterns of *FENNTARTHATÓ* ‘sustainable’ contribute to the adjective’s meaning mainly by shaping a semantic preference for nouns that refer to development, namely *fejlődés* ‘development’ (singular, nominative case), *fejlődési* ‘development’ (adjective derived from *fejlődés* by means of the derivational suffix *-i*), *fejlődést* ‘development’ (singular, accusative case), *célok* ‘goals’ (plural, obtained from the noun *cél* ‘goal’ and the back-vowel grammatical suffix *-ok*), *fejlesztési* ‘developmental’ (adjective derived from *fejlesztés* ‘development’ with the derivational morpheme *-i*).

The adjective *FENNTARTHATÓ* ‘sustainable’ displays also a semantic preference for lexemes that hint at politics, like *országgyűlés* ‘parliament’ (singular, nominative), *bizottsága* ‘its commission’ (singular possessive form of *bizottság* ‘commission’ with the third person singular possessive suffix *-a*), *bizottságának* ‘(to) its commission’ (singular possessive form of *bizottság* ‘commission’ with the third person singular possessive suffix *-a* and the back-vowel dative case suffix *-nak*), *ENSZ* ‘United Nations’, *tanács* ‘council’ (singular, nominative), *ülésén* ‘on its sitting’ (with the singular noun *ülés* meaning ‘sitting’, *-e* standing for the third person singular possessive, and *-n* for the superessive case suffix), *elnöke* ‘its president’ (singular nominative form of *elnök* ‘president’ added with the third person singular possessive suffix *-e*).

The collocates of *FENNTARTHATÓ* 'sustainable' also semantically cluster in relation to geo-political dynamics by distinguishing a national and an international dimension: *nemzeti* 'national' (adjective derived from the noun *nemzet* 'nation' with the derivational suffix *-i*), *Magyarország* 'Hungary' (nominative), *világ* 'world' (singular, nominative), *európai* 'European' (adjective derived from the noun *Európa* 'Europe' and the derivational suffix *-i*).

The meaning of the adjective *FENNTARTHATÓ* 'sustainable' is further enriched by its collocating with lexemes that detail a semantic preference for the economic and environmental dimensions of sustainable development: *gazdasági* 'economic' mirrors the economic dimension of sustainability; *környezeti* 'environmental' reflects the environmental dimension.

The collocate *szempontjából* 'from its point of view' (consisting of the singular noun *szempont* 'point of view', of the third person singular possessive suffix *-ja*, and of the back-vowel relative case suffix *-ból*) highlights that some statements of the SusCorp (Hungarian) are written under the light of the sustainable development issue. The collocate *szerint* 'according to' plays a comparable role: it ascribes statements on sustainable development to politicians and experts. These tendencies constitute the semantic preference of *FENNTARTHATÓ* 'sustainable'.

Two verbs emerge among the collocates of *FENNTARTHATÓ* 'sustainable': the auxiliary verb *kell* 'have to' and the existential verb *van* '(he/she/it) is'. *kell* 'have to' assigns a semantic shade of obligation to processes that host the adjective *FENNTARTHATÓ* 'sustainable' in their arguments or circumstantials. These processes are mainly material processes like *valósítani* 'to achieve' ( $f=3$ ), *biztosítani* 'to ensure' ( $f=2$ ), *építenie* 'to build' (used when the subject of the auxiliary is a third person singular subject;  $f=2$ ), *átalakítani* 'to transform' ( $f=1$ ), *erősíteni* 'to strengthen' (used when the subject of the auxiliary is a third person singular subject;  $f=1$ ), *felkészíteni* 'to prepare' ( $f=1$ ), *fokozni* 'to increase' ( $f=1$ ), *fordítani* 'to turn' ( $f=1$ ), *helyezni* 'to place' ( $f=1$ ), *lépni* 'to move' ( $f=1$ ), *létrehozni* 'to establish' ( $f=1$ ), *megtenniük* 'to accomplish' (used when the subject of the auxiliary is a third person plural subject;  $f=1$ ), *mutatniuk* 'to show' (use when the subject of the auxiliary is a third person plural subject;  $f=1$ ), *összehangolni* 'to unite' ( $f=1$ ), *változtatni* 'to change' ( $f=1$ ). Processes modified by *kell* 'have to' can also be relational, as in the case of *lennie* 'to be' (used when the subject of the auxiliary is a third person singular subject;  $f=3$ ), mental, as with *tudnunk* 'to know' (used when the subject of the auxiliary is a first person plural subject;  $f=1$ ), and verbal, as in the case of *szólnia* 'to say' (used when the subject of the auxiliary is a third person singular subject;  $f=2$ ).

As far as function words are concerned, the adjective *FENNTARTHATÓ* ‘sustainable’ is very frequently preceded by the definite article *a* ‘the’, which signals the word’s definiteness in discourse. In noun phrases consisting of iterative sequences of nouns, the nouns following the first are usually found in possessive chains where the second noun and the following nouns are generally marked by a possessive suffix, as in

<i>Országgyűlés-Ø</i>	<i>fenntartható-Ø</i>	<i>fejlődés-Ø</i>
parliament-SG-NOM	sustainable-SG-NOM	development-SG-NOM

<i>bizottság-Ø-a-Ø</i>	<i>ülés-Ø-é-n</i>
commission-SG-POSS.3SG-NOM	sitting-SG-POSS.3SG-SUP

‘the sitting of the parliament commission on sustainable development’.

In these possessive chains, the second to last noun is frequently marked with the dative case suffix, as in

<i>meghatározott-Ø</i>	<i>fenntartható-Ø</i>	<i>fejlődés-Ø</i>
defined-SG-NOM	sustainable-SG-NOM	development-SG-NOM

<i>cél-ja-i-nak</i>	<i>hatékony-Ø</i>	<i>végrehajtás-Ø-á-ban</i>
goal-POSS.3SG-PL-DAT	effective-SG-NOM	realization-SG-POSS.3SG-INES

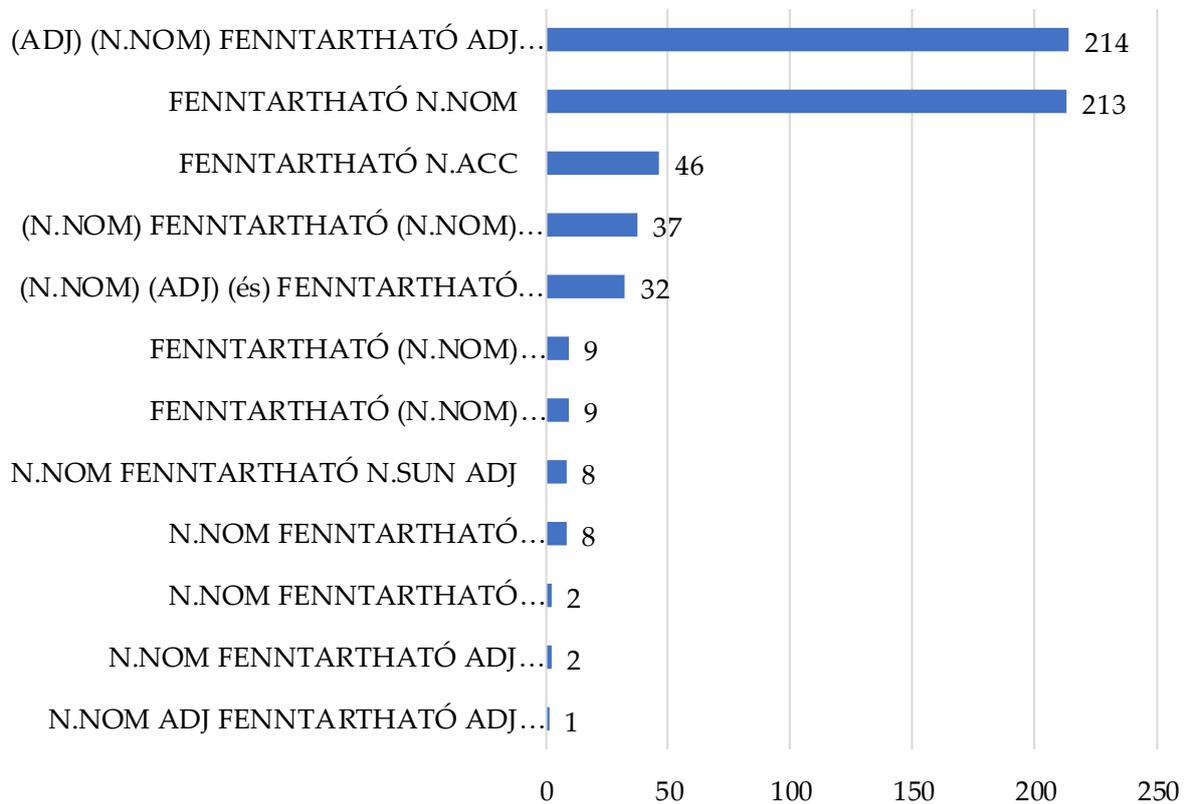
‘in the effective realization of the defined sustainable development goals’

The lexemes co-occurring with *FENNTARTHATÓ* ‘sustainable’ mainly gather in the following colligational patterns:

(ADJ)	(N.NOM)	<i>FENNTARTHATÓ</i>	ADJ	(N.NOM)	
N.NOM	ADJ	<i>FENNTARTHATÓ</i>	ADJ	N.POSS3SG.NOM	
	N.NOM	<i>FENNTARTHATÓ</i>	ADJ	N.POSS3SG.DAT	
		<i>FENNTARTHATÓ</i>	N.NOM		
(N.NOM)	(ADJ)	(és)	<i>FENNTARTHATÓ</i>	(N.NOM)	N.POSS3SG.NOM
	(N.NOM)	<i>FENNTARTHATÓ</i>	(N.NOM)		N.POSS3SG.DAT

	<i>FENNTARTHATÓ</i>	(N.NOM)	N.POSS3SG.ELA
	<i>FENNTARTHATÓ</i>	(N.NOM)	N.POSS3SG.INE
N.NOM	<i>FENNTARTHATÓ</i>	N.POSS3SG.NOM	N.POSS3SG.ELA
N.NOM	<i>FENNTARTHATÓ</i>	N.POSS3SG.NOM	N.POSS3SG.INE
N.NOM	<i>FENNTARTHATÓ</i>	N.SUB	ADJ
	<i>FENNTARTHATÓ</i>	N.ACC	

The frequency of these colligational patterns in the SusCorp (Hungarian) can be seen in Graph 23.



**Graph 23.** The colligational patterns of *FENNTARTHATÓ* ‘sustainable’ in the SusCorp (Hungarian).

Graph 23 shows that the most frequent colligational patterns of *FENNTARTHATÓ* ‘sustainable’ involve the adjective in noun phrases where the adjective modifies a noun marked by the nominative case (i.e. by zero marking), eventually accompanied by another adjective. This noun phrase can be pre-modified by another noun phrase.

As far as evaluative prosody is concerned, most collocational, colligational and semantic patterns of *FENNTARTHATÓ* 'sustainable' display a neutral connotation in relation to the ecological framework adopted in the present study. The collocates assigning an explicit polarity to clauses including *FENNTARTHATÓ* 'sustainable' are the negator *nem* 'not', which apparently adds an aura of negativity to the sentences that include the adjective, and the adjective *fontos* 'important', which on the contrary fills the evaluative prosody of *FENNTARTHATÓ* 'sustainable' with positivity. The negator *nem* 'not', however, very often negates verbs or verb phrases characterized by a potential modality, either because they are held by a potential auxiliary like *lehet* 'can' or because of their being morphologically marked by the potential derivational suffix *-hat/-het*. This potential modality is usually associated with processes that clarify the requirements that need to be met and those actions that need not to be done in order for sustainable development to be achieved. As a consequence, the negator *nem* 'not' contributes to the positive semantic prosody of *FENNTARTHATÓ* 'sustainable'.

### 5.2.3. The Italian *SOSTENIBILE*

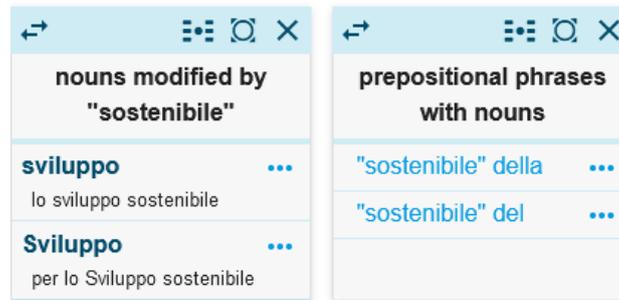
The description of the meaning by collocation of *sustainable* in the Sustainable development Corpus is enriched thanks to the analysis of the collocational tendencies of the Italian adjective *SOSTENIBILE* 'sustainable' in the SusCorp (Italian).

The adjective *SOSTENIBILE* 'sustainable' appears 1,136 times in the Italian section of the Sustainable development Corpus, namely 2.27 times per article on average.

#### 5.2.3.1. Word sketch

The meaning by collocation of the Italian *SOSTENIBILE* 'sustainable' is first outlined with a word sketch, which showcases the most significant grammatical patterns that involve the adjective.

The word sketch of *SOSTENIBILE* 'sustainable' (Figure 38) shows a snapshot of the semantics of the adjective by highlighting the nouns it modifies most meaningfully and the prepositional phrases it is modified by most commonly.



**Figure 38.** Word sketch of *SOSTENIBILE* ‘sustainable’ in the SusCorp (Italian) – LogDice(6.0), NC20-C20.

The strongest noun collocate of *SOSTENIBILE* ‘sustainable’ is *sviluppo* ‘development’, which is considered by the word sketch twice because of its statistically significantly starting both in lower case and in upper case. In the SusCorp (Italian), the adjective *SOSTENIBILE* ‘sustainable’ modifies the noun *sviluppo* ‘development’ 787 times and the noun phrase *sviluppo sostenibile* ‘sustainable development’ makes up for approximately 69% of the occurrences of *SOSTENIBILE* ‘sustainable’. The collocational patterns produced by the noun *sviluppo* ‘development’ can be grammatically analysed through the colligation

N *sostenibile*

, which is enriched when the prepositional phrases represented on the right side of Figure 38 come into play. In these cases, *SOSTENIBILE* ‘sustainable’ is spotted in the boosted colligational patterns

N                    *sostenibile*                    P                    (DET)                    (ADJ)                    N.

### 5.2.3.2. Collocation networks

The acquaintance with the semantics of the Italian adjective *SOSTENIBILE* ‘sustainable’ is enhanced with the analysis of the collocation networks of the lexeme in the SusCorp (Italian). These collocation networks summarise a broader collocational behaviour for the adjective compared to the word sketch that has just been presented.

Table 41 displays the collocates computed for the adjective with a collocation window ranging from one to five words to the left and to the right of the node. The lists of collocates added with statistical and frequency thresholds can be found in the Appendix (Tables 70, 71, and 72).

Rank	Collocate (1L-1R)	Collocate (3L-3R)	Collocate (5L-5R)
1	<i>sviluppo</i> 'development'	<i>sviluppo</i> 'development'	<i>sviluppo</i> 'development'
2	<i>e</i> 'and'	<i>lo</i> 'the'	<i>lo</i> 'the'
3	<i>più</i> 'more'	<i>dello</i> 'of (the)'	<i>dello</i> 'of (the)'
4		<i>obiettivi</i> 'goals'	<i>obiettivi</i> 'goals'
5		<i>uno</i> 'a'	<i>per</i> 'for'
6		<i>per</i> 'for'	<i>uno</i> 'a'
7		<i>equo</i> 'equitable'	<i>e</i> 'and'
8		<i>sullo</i> 'on (the)'	<i>di</i> 'of'
9		<i>e</i> 'and'	17 '17'
10		<i>allo</i> 'to (the)'	<i>sullo</i> 'on (the)'
11		<i>mobilità</i> 'mobility'	<i>equo</i> 'equitable'
12		<i>di</i> 'of'	2030 '2030'
13		2030 '2030'	<i>allo</i> 'to (the)'
14		<i>il</i> 'the'	<i>mobilità</i> 'mobility'
15		<i>in</i> 'in'	<i>il</i> 'the'
16		<i>che</i> 'that'	<i>fondazione</i> 'foundation'
17		<i>più</i> 'more'	<i>un</i> 'a'
18		<i>un</i> 'a'	<i>italiana</i> 'Italian'
19		<i>la</i> 'the'	<i>in</i> 'in'
20		<i>è</i> '(he/she/it) is'	<i>che</i> 'that'
21		<i>con</i> 'with'	<i>modello</i> 'model'
22		<i>a</i> 'to'	<i>Nazioni</i> 'Nations'
23		<i>non</i> 'not'	<i>gli</i> 'the'
24			<i>Unite</i> 'United'
25			<i>la</i> 'the'
26			<i>a</i> 'to'
27			<i>territorio</i> 'territory'
28			<i>della</i> 'of (the)'
29			<i>futuro</i> 'future'
30			<i>sociale</i> 'social'
31			<i>è</i> '(he/she/it) is'
32			<i>ha</i> '(he/she/it) has'
33			<i>una</i> 'a'
34			<i>più</i> 'more'
35			<i>le</i> 'the'
36			<i>del</i> 'of (the)'
37			<i>si</i> '-'
38			<i>i</i> 'the'
39			<i>città</i> 'city, cities'
40			<i>delle</i> 'of (the)'
41			<i>nel</i> 'in (the)'
42			<i>con</i> 'with'
43			<i>sia</i> '(he/she/it) is; both'
44			<i>anche</i> 'also'

45	<i>dalla</i> 'from (the)'
46	<i>su</i> 'on'
47	<i>non</i> 'not'
48	<i>sul</i> 'on (the)'
49	<i>questo</i> 'this'
50	<i>alla</i> 'to (the)'
51	<i>al</i> 'to (the)'
52	<i>dei</i> 'of (the)'
53	<i>tra</i> 'between'
54	<i>da</i> 'from'
55	<i>ai</i> 'to (the)'
56	<i>ma</i> 'but'
57	<i>sono</i> '(they) are'
58	<i>degli</i> 'of (the)'
59	<i>come</i> 'like; how'

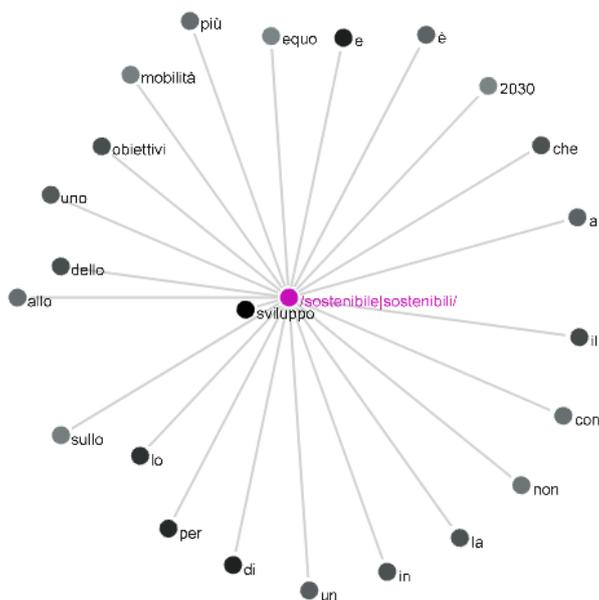
**Table 41.** List of collocates of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian) – Z(10.0), 1L-1R/3L-3R/5L-5R, NC20-C20.

Between one word to the left and one word to the right, *SOSTENIBILE* 'sustainable' collocates with the content word *sviluppo* 'development', and with the function words *e* 'and' and *più* 'more'. As it can be noticed in Figure 39, the content word *sviluppo* 'development' and the function word *più* 'more' tend to occur on the left-hand side of the node, whereas the function word *e* 'and' usually appears on the right-hand side of the node.



**Figure 39.** Collocation network of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian) – Z(10.0), 1L-1R, NC5-C5.

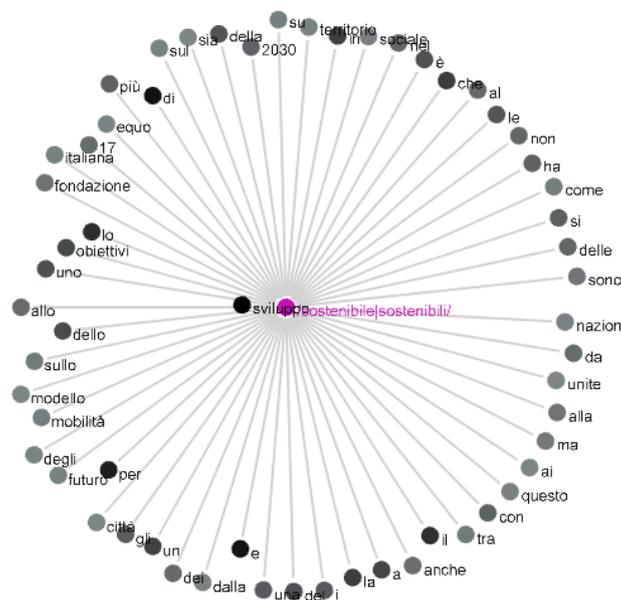
Within a  $\pm 3$ -word collocation window, *SOSTENIBILE* 'sustainable' collocates with the same content and function words that have already been observed for the  $\pm 1$ -word span. In addition, the Italian adjective co-occurs also with the content words *obiettivi* 'goals', *equo* 'equitable', *mobilità* 'mobility' and with the function words *lo* 'the', *dello* 'of (the)', *uno* 'a', *per* 'for', *sullo* 'on (the)', *allo* 'to (the)', *di* 'of', *2030* '2030', *il* 'the', *in* 'in', *che* 'that', *un* 'a', *la* 'the', *è* '(he/she/it) is', *con* 'with', *a* 'to', *non* 'not'. Figure 40 shows that the content words *obiettivi* 'goals', *equo* 'equitable', *mobilità* 'mobility' usually appear on the left-hand side of *SOSTENIBILE* 'sustainable'. The function words *lo* 'the', *dello* 'of (the)', *uno* 'a', *per* 'for', *sullo* 'on (the)', *allo* 'to (the)', *di* 'of', *2030* '2030', *un* 'a' tend to occur on the left-hand side of the node, whereas the function words *il* 'the', *in* 'in', *che* 'that', *la* 'the', *è* '(he/she/it) is', *con* 'with', *a* 'to', *non* 'not' usually appear on its right-hand side.



**Figure 40.** Collocation network of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian) – Z(10.0), 3L-3R, NC5-C5.

With a  $\pm 5$ -word span, *SOSTENIBILE* 'sustainable' collocates with the content words *fondazione* 'foundation', *italiana* 'Italian', *modello* 'model', *Nazioni* 'Nations', *Unite* 'United', *territorio* 'territory', *futuro* 'future', *sociale* 'social', *città* 'city, cities' and with the function words *17* '17', *gli* 'the', *della* 'of (the)', *ha* '(he/she/it) has', *una* 'a', *le* 'the', *del* 'of (the)', *si* '-', *i* 'the', *delle* 'of (the)', *nel* 'in (the)', *sia* '(he/she/it) is; both', *anche* 'also', *dalla* 'from (the)', *su* 'on', *sul* 'on (the)', *questo* 'this', *alla* 'to (the)', *al* 'to (the)', *dei* 'of (the)', *tra* 'between', *da* 'from', *ai* 'to', *ma* 'but', *sono* '(they) are', *degli* 'of (the)', *come* 'like;

how'. These content and function words are summed to the ones that have already been mentioned for the  $\pm 1$ - and  $\pm 3$ -word spans. As it can be seen in Figure 41, the content words *fondazione* 'foundation', *italiana* 'Italian', *modello* 'model', *futuro* 'future', *città* 'city, cities' tend to occur on the left-hand side of *SOSTENIBILE* 'sustainable', while the content words *Nazioni* 'Nations', *Unite* 'United', *territorio* 'territory', *sociale* 'social' usually appear on the right-hand side of the node. The function words *17* '17', *gli* 'the', *sul* 'on (the)', *dei* 'of (the)', *degli* 'of (the)' usually stand on the left-hand side of *SOSTENIBILE* 'sustainable', whereas the function words *della* 'of (the)', *ha* '(he/she/it) has', *una* 'a', *le* 'the', *del* 'of (the)', *si* '-', *i* 'the', *delle* 'of (the)', *nel* 'in (the)', *anche* 'also', *su* 'on', *questo* 'this', *alla* 'to (the)', *al* 'to (the)', *tra* 'between', *da* 'from', *ai* 'to', *ma* 'but', *sono* '(they) are', *come* 'like; how' tend to occur on the right-hand side of the node. The function words *sia* '(he/she/it) is; both' and *dalla* 'from (the)' stand equally on the left- and on the right-side of *SOSTENIBILE* 'sustainable'.



**Figure 41.** Collocation network of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian) – Z(10.0), 5L-5R, NC5-C5.

As far as semantic preference is concerned, the collocates of *SOSTENIBILE* 'sustainable' cluster together in few but very clear-cut semantic fields. The first semantic preference of the adjective is for lexemes that refer to sustainable development as it is envisaged by the United Nations' 2030 Agenda. This semantic preference is represented by the lexemes *sviluppo* 'development', *obiettivi* 'goals', *2030* '2030', *17* '17'. The noun *sviluppo* 'development' is the head of the noun phrase *sviluppo sostenibile* 'sustainable development', which occurs 787 times in the SusCorp (Italian). As it can be glimpsed



9	<i>Agenda 2030 di Sviluppo Sostenibile</i> '2030 Agenda for Sustainable Development'	1
10	<i>Agenda del 2030 per lo Sviluppo Sostenibile</i> '2030 Agenda for Sustainable Development'	1
11	<i>Agenda di Sviluppo Sostenibile 2030</i> '2030 Agenda for Sustainable Development'	1
12	<i>Agenda Sostenibile dell'ONU per il 2030</i> 'UN's 2030 Agenda for Sustainable Development'	1

**Table 42.** The ways of referring to the UN's 2030 Agenda in the SusCorp (Italian).

As it can be seen in Table 42, the 2030 Agenda is mentioned in several ways when the collocates 2030 '2030' is involved. The various lexical items share references to the *agenda* 'agenda' and to 2030 '2030', but they variably include also references to *sviluppo sostenibile* 'sustainable development' and to the *ONU* 'UN'. The main lexical and grammatical features that diversify these lexical items regard the morphosyntactic aspects of the use of prepositional phrases, of the choice of the preposition that functions as head of these prepositional phrases, and of the position of the various elements in the lexical item.

Also the collocates *obiettivi* 'goals' and 17 '17' serve a semantic preference for sustainable development as it is meant by the UN's 2030 Agenda. In particular, *obiettivi* 'goals' and 17 '17' acknowledge the Agenda's sustainable development goals. *obiettivi* 'goals' is usually found in the patterns displayed in Table 43:

Rank	Names for the SDGs	<i>f</i>
1	<i>obiettivi di sviluppo sostenibile</i> 'sustainable development goals'	116
2	<i>obiettivi per lo sviluppo sostenibile</i> 'sustainable development goals'	11
3	<i>obiettivi dello sviluppo sostenibile</i> 'sustainable development goals'	3
4	<i>obiettivi per uno sviluppo sostenibile</i> 'sustainable development goals'	2
5	<i>obiettivi mondiali di sviluppo sostenibile</i> 'world sustainable development goals'	1
6	<i>obiettivi sostenibili</i> 'sustainable goals'	1

**Table 43.** The ways of referring to the UN's Sustainable Development Goals in the SusCorp (Italian).

As in the case of the lexical items identifying the 2030 Agenda (Table 42), also the lexical items mentioning the 17 Sustainable Development Goals are several (Table 43).

They diverge because of the prepositional phrases that bind *obiettivi* 'goals' and *sviluppo sostenibile* 'sustainable development'. These prepositional phrases change in terms of the selected preposition and of the way it is modified by the noun phrases including *sviluppo sostenibile* 'sustainable development'. Some of these lexical items are modified by the numeral 17 '17', giving rise to the patterns 17 *obiettivi di sviluppo sostenibile* '17 sustainable development goals' (f=34) and 17 *obiettivi per lo sviluppo sostenibile* '17 sustainable development goals' (f=4).

In addition, *SOSTENIBILE* 'sustainable' collocates with lexemes semantically connected because of their encoding political actors. Through these lexemes, sustainable development is narrated to be approached in a political way. The lexemes showing a semantic preference for politics are *fondazione* 'foundation', *italiana* 'Italian', *Nazioni* 'Nations', and *Unite* 'United'. *Nazioni* 'Nations' and *Unite* 'United' co-occur in the lexical item *Nazioni Unite* 'United Nations' (f=20). *Nazioni Unite* 'United Nations' are usually mentioned in relation to the SDGs. Through the lexeme *fondazione* 'foundation', the articles of the SusCorp (Italian) name organisations that engage with sustainability issues: *Fondazione per lo sviluppo sostenibile* 'Sustainable development foundation' (f=46), *Fondazione sviluppo sostenibile* 'Sustainable development foundation' (f=12), but also *Fondazione Barilla Center* 'Barilla Center Foundation' (f=1), *Fondazione Pirelli* 'Pirelli Foundation' (f=1), *Fondazione Prada* 'Prada Foundation' (f=1), *Fondazione Pubblicità Progresso* 'Pubblicità Progresso Foundation' (f=1). Also the lexeme *italiana* 'Italian' (feminine singular) names organisations that engage with sustainability issues by appearing almost always in *Alleanza Italiana per lo Sviluppo Sostenibile* 'Italian Alliance for Sustainable Development' (f=26).

The other content words that *SOSTENIBILE* 'sustainable' collocates with do not revolve around any specific semantic preference. The adjective *equo* 'equitable' summons general ethics, and it is used in association with the adjective *SOSTENIBILE* 'sustainable' in the adjective phrases *equo e sostenibile* 'equitable and sustainable' (f=16), *sostenibile, equo ed inclusivo* 'sustainable, equitable and inclusive' (f=3), and *sostenibile e equo* 'sustainable and equitable' (f=2). These adjective phrases join sustainability and equity by means of the conjunction *e* 'and', and they modify the nouns *sviluppo* 'development' (f=10), *benessere* 'wellbeing' (f=9), *commercio* 'trade' (f=1), *futuro* 'future' (f=1).

The collocates *città* 'city; cities' and *territorio* 'territory' on the one hand, and *mobilità* 'mobility' on the other respectively recall places and the transport system that allows to move within and between places. In the case of the co-occurrence of *SOSTENIBILE* 'sustainable' and *città* 'city; cities', the collocation narrates a story of *CITTÀ*

*SOSTENIBILE* ‘sustainable city; sustainable cities’ (f=5), of *città e comunità sostenibili* ‘sustainable cities and communities’ (f=3), of *città e territorio sostenibili* ‘sustainable cities and territories’ (f=1), of *città più sicure, sostenibili e meglio connesse* ‘safer, more sustainable and more connected cities’ (f=1). This story suggests that cities are sustainable or that they should be so. When *SOSTENIBILE* ‘sustainable’ collocates with *territorio* ‘territory’, they co-occur in patterns like *sviluppo sostenibile del territorio* ‘sustainable territory development’ (f=4), *gestione sostenibile del territorio* ‘sustainable territory management’ (f=3), *gestione sostenibile integrale del territorio* ‘integral, sustainable territory management’ (f=1), *progetti sostenibili di gestione del territorio* ‘sustainable projects of territory management’ (f=1), *territorio incentrato sullo sviluppo sostenibile* ‘territory focused on sustainable development’ (f=1). In these patterns, the development and the management of the territory are qualified as sustainable. A similar story can be read also for *mobilità* ‘mobility’. *mobilità* ‘mobility’ is described as *SOSTENIBILE* ‘sustainable’ in patterns like *mobilità sostenibile* ‘sustainable mobility’ (f=18), *mobilità più sostenibile* ‘more sustainable mobility’ (f=1), *mobilità urbana sostenibile* ‘sustainable urban mobility’ (f=1), *mobilità urbana sempre più sostenibile* ‘more and more sustainable urban mobility’ (f=1).

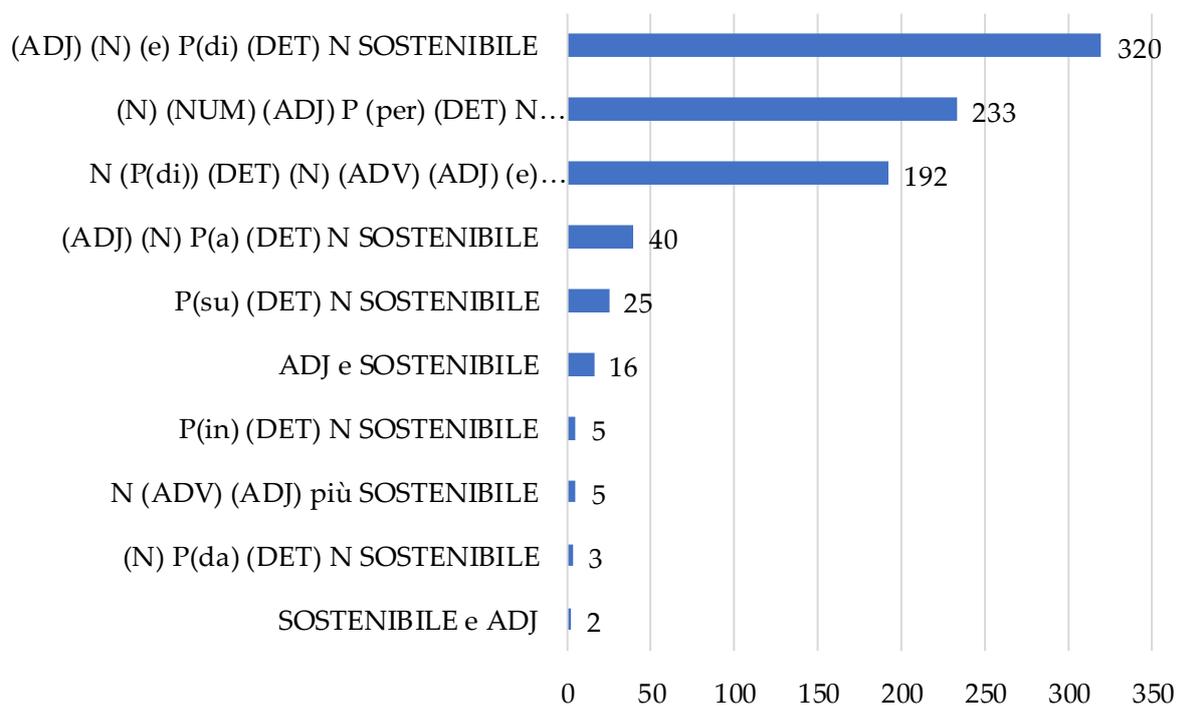
The adjective *SOSTENIBILE* ‘sustainable’ experiences also a semantic preference for a lexeme that reproduces an abstract concept of forward-directed aspiration (i.e. *futuro* ‘future’). *SOSTENIBILE* ‘sustainable’ collocates with *futuro* ‘future’ in the patterns *futuro sostenibile* ‘sustainable future’ (f=9), *futuro dell’Italia equo e sostenibile* ‘equitable and sustainable future for Italy’ (f=1), *futuro dell’umanità condiviso e realmente sostenibile* ‘shared and truly sustainable future for humanity’ (f=1), *futuro digitale sostenibile* ‘sustainable digital future’ (f=1), *futuro economicamente e socialmente sostenibile* ‘economically and socially sustainable future’ (f=1), *futuro più sostenibile* ‘more sustainable future’ (f=1).

The last semantic preference of *SOSTENIBILE* ‘sustainable’ in the SusCorp (Italian) involves the noun *modello* ‘model’, namely a lexeme that embodies a plan, a strategy. *SOSTENIBILE* ‘sustainable’ and *modello* ‘model’ co-occur in the patterns *modello di sviluppo sostenibile* ‘sustainable development model’ (f=10), *modello sostenibile* (f=2), *modello dell’Eni per lo sviluppo sostenibile* ‘Eni’s model for sustainable development’ (f=1), *modello di sviluppo equo e sostenibile* ‘equitable and sustainable development model’ (f=1), *modello di sviluppo più sostenibile* ‘more sustainable development model’ (f=1), *modello Eni per lo sviluppo sostenibile* ‘Eni model for sustainable development’ (f=1), *modello per uno sviluppo sostenibile* ‘model for sustainable development’ (f=1).

The lexemes that embody these semantic preferences are organised in the following, main colligational patterns:

N	(P( <i>di</i> ))	(DET)	(N)	(ADV)	(ADJ)	( <i>e</i> )	(ADV)	SOSTENIBILE
						ADJ	<i>e</i>	SOSTENIBILE
								SOSTENIBILE <i>e</i> ADJ
			N	(ADV)	(ADJ)	<i>più</i>		SOSTENIBILE
		(ADJ)	(N)	P( <i>a</i> )	(DET)	N		SOSTENIBILE
			(N)	P( <i>da</i> )	(DET)	N		SOSTENIBILE
	(ADJ)	(N)	( <i>e</i> )	P( <i>di</i> )	(DET)	N		SOSTENIBILE
	(N)	(NUM)	(ADJ)	P( <i>per</i> )	(DET)	N		SOSTENIBILE
				P( <i>in</i> )	(DET)	N		SOSTENIBILE
				P( <i>su</i> )	(DET)	N		SOSTENIBILE

The frequency of these colligational patterns can be seen in Graph 24.



**Graph 24.** The colligational patterns of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian).

As it can be seen in Graph 24, the most frequent colligational patterns of *SOSTENIBILE* 'sustainable' involve the adjective in a noun phrase that modifies prepositional phrases introduced by the prepositions *di* 'of' and *per* 'for'. The preposition *di* 'of' is the head of prepositional phrases that function as modifiers or as possessors of noun or adjective phrases, whereas *per* 'for' is the head of prepositional phrases that function as modifiers or as goals of the processes embodied by the noun phrases that the prepositional phrase with *per* 'for' is attached to.

Thanks to the semantic preference of *SOSTENIBILE* 'sustainable' and thanks to the connotation of the colligational patterns it is included in, the evaluative prosody of the adjective can be regarded neutral overall. The adjective is usually surrounded by neutral lexemes and patterns, and the only lexeme that seems to bear a negative connotation is the negation *non* 'not'. The negation *non* 'not', however, negates practices that should be avoided in order to make sustainable development possible. Thus, it adds a positive hue to the meaning by collocation of the adjective *SOSTENIBILE* 'sustainable'.

### 5.3. Discussion

Summing up, both similarities and differences have been noticed when identifying cultural keywords and when retrieving the meaning by collocation of the English *sustainable*, of the Hungarian *FENNTARTHATÓ* 'sustainable' and of the Italian *SOSTENIBILE* 'sustainable' in the SusCorp (English), in the SusCorp (Hungarian) and in the SusCorp (Italian) respectively. These similarities and differences depend first on morphological and syntactic differences across languages, but also on the peculiarity of the discourse type represented by the SusCorp.

As far as the identification of cultural keywords is concerned, it can be noticed that the three subcorpora share a core of key lexical items referring to sustainable development: sustainability starts with the lexemes *development*, *sustainable*, *goals* in the SusCorp (English), with *FEJLŐDÉS* 'development', *FEJLŐDÉSI* 'development', *FENNTARTHATÓ* 'sustainable', *FENNTARTHATÓSÁG* 'sustainability' in the SusCorp (Hungarian), and with *sviluppo* 'development', *sostenibile* 'sustainable', *sostenibilità* 'sustainability' and *obiettivi* 'goals' in the SusCorp (Italian). This commonality of cultural keywords can be easily explained as a consequence of the design of the Sustainable development Corpus. The corpus, in fact, is planned so that all texts included in it mention sustainable development at least once. Despite this common

core, cultural keywords slightly diverge in the three subcorpora: in the SusCorp (English), the environmental dimension predominates though the lexemes *climate* and *change*; in the SusCorp (Hungarian), the social and the environmental dimensions coexist thanks to the reference to migration (*MIGRÁCIÓ* 'migration', *MIGRÁCIÓS* '(of) migration', *MIGRÁNS* 'migrant') and climate change (*KLÍMAVÁLTOZÁS* 'climate change', *ÉGHAJLATVÁLTOZÁS* 'climate change'); in the SusCorp (Italian), the environmental dimension prevails (*GREEN* 'green'). These cultural keywords emphasise the political, social and cultural specificity of the British, Hungarian and Italian quality press between 2016 and 2018 in relation to sustainable development. British quality press pays much attention to climate change, Hungarian quality press combines the same attention to climate change with a concern for migration, the Italian quality press concentrates on more general environmental issues. This is due to the nature of news discourse on sustainable development in general, but also to the specific ideological standpoints of the broadsheets represented in the SusCorp. In the case of the SusCorp (English), for instance, the keyness of the lexemes *climate* and *change* can be partly explained by the fact that the most represented British quality paper in the SusCorp (English) is *The Guardian*, which is particularly eager to treat environmental issues.

Additional similarities in relation to cultural keywords can be found among the climate-related lexemes. The semantics of *climate* in English and of *climate's* Hungarian and Italian equivalents in the SusCorp is constructed by collocation in the following way: climate is a natural element of environment which can be characterized by extreme events, boosted by human actions, and ruled on by politics. The noun is usually found in compounds that function as goals of material processes. These transitivity patterns are frequently filled with lexical items related to the war metaphor. The metaphor of war is often associated with climate change in news discourse, as Deignan et al. (2017) point out. Atanasova and Koteyko (2017) too notice that in the news one of the most commonly employed metaphors for climate change is war, alongside religion and politics.

Moreover, it is interesting to notice that the American president Trump is mentioned in all three sections of the SusCorp among the statistically significant words. This depends on his debatable position with regards to climate change. At the same time, the only subcorpus that features a foreign word among the most frequent or statistically significant lexemes of the collection is the SusCorp (Italian). In the SusCorp (Italian), the English lexemes *GREEN* 'green' and *ECONOMY* 'economy' play an important role for the Italian news discourse on sustainable development.

As far as meanings by collocation are concerned, the semantics of the English adjective *sustainable*, of the Hungarian adjective *FENNTARTHATÓ* 'sustainable' and of the Italian adjective *SOSTENIBILE* 'sustainable' are comparable. The three adjectives display similar semantic preferences and similar evaluative prosodies. The main differences in the meaning by collocation of the English *sustainable*, of the Hungarian *FENNTARTHATÓ* 'sustainable' and of the Italian *SOSTENIBILE* 'sustainable' lie in the language-specific colligational patterns that involve the adjectives. If we compare collocation networks across languages, we might easily notice a divergence in the ratio between content and function words, which is reflected by colligational patterns. In the collocation network retrieved for the English *sustainable* between five words to the left and five words to the right of the node, the purely content words make up for the 43% of the total number of collocates. With a  $\pm 5$ -word span, the collocation network of the Hungarian *FENNTARTHATÓ* 'sustainable' consists of purely content words for 60% of the total number of collocates, and of function words or words halfway between function and content words for 40% of the total. In the case of the collocation network of the Italian *sostenibile* 'sustainable' within a  $\pm 5$ -word collocation window, purely content words make up for only 22% of the total number of collocates. These differences depend on the structural characteristics of the three languages and on the fact that the Sustainable development Corpus is analysed in its raw version in search for meaning by collocation (as it has already been elaborated on in § 4.5).



## 6. Conclusion

The sixth chapter of the dissertation offers some final remarks. First, I summarise the results obtained from the analysis of the discursive construction of sustainable development in the 2030 Agenda Corpus and in the Sustainable development Corpus. Then, I compare the results cross-linguistically, and I interpret them under the light of the ecological framework adopted in the current research. Finally, I outline some of the main strengths and weaknesses of the study, and I suggest future developments.

### 6.1. Summary of the results

The current research has outlined the discursive construction of sustainable development in the English, Hungarian and Italian versions of the United Nations's *2030 Agenda for Sustainable Development* (2030 Agenda Corpus; for a thorough analysis of the corpus see § 4) and in the British, Hungarian and Italian news discourse on sustainable development published after the release of the 2030 Agenda (Sustainable development Corpus; for the complete analysis of the corpus see § 5). The discursive construction of sustainable development has been investigated in two steps: first, through the retrieval of the most frequent and statistically significant lexemes of the collections, with a consequent identification of cultural keywords among them (for more on cultural keywords, frequency and statistical keywords see § 1.1 and § 1.3.3.3 respectively); second, through the extraction of the meaning by collocation of the identified cultural keywords and of lexical items chosen *a priori*, namely *sustainable development*, *sustainable*, and *sustainability* for English, and their Hungarian and Italian translational equivalents (i.e. *FENNTARTHATÓ FEJLŐDÉS*, *FENNTARTHATÓ, FENNTARTHATÓSÁG* for Hungarian, and *SVILUPPO SOSTENIBILE, SOSTENIBILE, SOSTENIBILITÀ* for Italian; for more on meaning by collocation see § 1.2 and § 1.3.3.3).

#### 6.1.1. The 2030 Agenda Corpus

As far as the discursive construction of sustainable development in the 2030 Agenda Corpus is concerned (for a thorough description see § 4), the most frequent content words of the English version of the 2030 Agenda shape the resolution as a document that assesses and encourages the world's *sustainable development*. Sustainable development is defined within the framework of the 2030 Agenda (*2030, agenda,*

*Nations, United*) as a complex enterprise that can be achieved only if it is subdivided in circumscribed tasks (*goals*). This enterprise engages the world's countries according to their level of development (*countries, developing, developed*) and it forces them both into a national and into an international dimension (*global, national, international, levels*). The *economic* dimension of sustainable development is mentioned more frequently than the social and the environmental ones. All these dimensions require a certain degree of promotion to be achieved (*support, access, ensure, implementation, promote*).

The lists of the most frequent content words of the Hungarian and of the Italian versions of the 2030 Agenda share most of their content with the English content word list. Therefore, the aboutness of the three documents is comparable. The differences observed across the lists depend on the linguistic characteristics of the three languages under inquiry and on peculiar stylistic choices that might have been done by the translators of the resolution, but they seldom include additions to the aboutness of the 2030 Agenda Corpus.

Thus, the cultural keywords identified among the most frequent content words of the 2030 Agenda Corpus are the English *development, sustainable, global* and *goals*, paralleled by the Hungarian *fejlesztés/fejlesztési* 'development', *fenntartható* 'sustainable', *globális-globale* 'global' and *célok/cél* 'goals', and the Italian *sviluppo* 'development', *sostenibile* 'sustainable', *globale* 'global' and *obiettivi* 'goals'.

The most salient statistical keywords of the 2030 Agenda (English) correspond in part to the most frequent content words of the document. Those lexemes that are not shared contribute both to the aboutness and to the style of the resolution. They directly address the 2030 Agenda (*goal, targets*) and they identify some of the institutional actors that should cooperate for its implementation (*forum*); they specify the kind of countries that will profit from the advantages of sustainability (*landlocked*) and the kind of advantages that could come with it (*inclusive*), while also pointing at one of the sectors that should improve to allow for these advantages to happen (*technology*).

Most of the most significant statistical keywords of the Hungarian and of the Italian Agendas match the corresponding translational equivalents in the list of the most salient statistical keywords of the English Agenda. Hungarian and Italian especially add elements of style and a further classification of the countries involved in sustainable practices (i.e. the Italian *insulari* 'island').

In addition to the cultural keywords highlighted among the most frequent content words of the 2030 Agenda Corpus, namely *sustainable, development, global, goals* and their Hungarian and Italian translational equivalents, the statistical keywords of

the English Agenda introduce also the cultural keywords *goal, inclusive* and *technology*. These are not included in the lists of the most salient statistical keywords of the Hungarian and Italian Agendas.

The study of the economic, social and environmental lexemes included in the 2030 Agenda shows that economic lexemes are by far more frequent than social and environmental lexemes. Furthermore, human beings are the most represented category among social and environmental words, followed by natural elements, plants and animals.

In this context, the meaning by collocation that the adjective *sustainable* acquires in the 2030 Agenda (English) makes the lexeme represent a positive quality associated with other positive qualities like inclusivity, trustworthiness and stability, and mainly characterising processes of change, depletion, improving and supporting. This semantic tendency is encoded in a specific semantic preference that fills colligational patterns and that features as the basis of a positively connoted evaluative prosody. The semantic preference of *sustainable* involves qualities of inclusiveness, trustworthiness and stability encoded through the lexemes *sustained, inclusive, resilient, modern, reliable, affordable*; it witnesses the meaningful co-occurrence of material processes of change (represented by the lexemes *development, innovation, growth*) and depletion (represented by the lexemes *consumption, production, management, use*); it experiences the frequent appearance of material and mental processes of improving and supporting encoded through the lexemes *promote, ensure, recognize, enhance*; it is triggered by human activities and products (as with *agriculture, economic, technology, policies*). These lexical preferences fill the colligational patterns that involve the adjective *sustainable*. The adjective is mainly involved in adjective phrases that modify noun phrases; these noun phrases are often included in prepositional phrases, but they can also function as direct objects of transitive verbs. The evaluative prosody gathered for the adjective through its semantic preference and its colligational tendencies proves overall positive. This positivity is established in ecolinguistic terms by considering the property of sustainability as one that tries to safeguard environmental balance.

In the 2030 Agenda Corpus, the meaning by collocation of the Hungarian and Italian translational equivalents of *sustainable*, namely *FENNTARTHATÓ* for Hungarian and *SOSTENIBILE* for Italian, is similar to the English adjective's. This similarity is due to comparable evaluative prosodies and comparable semantic preferences: all adjectives are positively connoted and the collocates retrieved for the three adjectives can be grouped alike. The only difference emerges in Italian, where the  $\pm 5$ -word collocation network features the collocate *oceani* 'oceans', which is totally absent from

the English and the Hungarian collocational patterns. On the other hand, the colligational patterns of the adjectives differ cross-linguistically because of the diverse grammatical characteristics of the languages under inquiry.

The meaning by collocation of the noun *sustainability* in the 2030 Agenda Corpus is straightforward. In the 2030 Agenda (English), sustainability is a property associated to the possibility for borrowing countries to pay back the debts they contract. This semantics is shared also by the Hungarian noun *FENNTARTHATÓSÁG* and by the Italian noun *SOSTENIBILITÀ* in the 2030 Agenda (Hungarian) and in the 2030 Agenda (Italian).

The meaning by collocation of the English *sustainable development* in the 2030 Agenda Corpus is one of a condition that is intertwined with the 2030 Agenda, which compels its receivers to commit to the achievement of sustainability for the wellbeing of people worldwide and with a mutual enterprise necessary to overcome eventual challenges that might occur in the process. The semantics of the lexical item is contributed to by the semantic preference that can be spotted among the collocates of *sustainable development*. These collocates portray a semantic preference for lexemes related to the 2030 Agenda (*goals, targets, goal, agenda*), for political and social conditions requiring cooperation (*peace, partnership, global, relevant*), for human products (*innovation, policies, technology*) and for difficulties of any kind (*poverty, challenges*), as well as for material processes of achieving (*achieving*) and mental processes of understanding (*recognize*). These lexical choices occupy the slots of the colligational patterns involving *sustainable development*. These colligational patterns witness *sustainable development* be part of noun phrases modified by adjective phrases and used either as modifiers of prepositional phrases or as direct objects of transitive verbs. The positive connotation of the lexemes co-occurring with *sustainable development* and of the grammatical patterns that organise them contribute to positively charging the evaluative prosody of the lexical item. The evaluative prosody of *sustainable development* is deemed positive in ecolinguistic terms, as the meaning that the lexical item acquires renders it a condition to be searched for the wellbeing of the world.

The meaning by collocation of the Hungarian and Italian translational equivalents of *sustainable development* (i.e. *FENNTARTHATÓ FEJLŐDÉS* in Hungarian and *SVILUPPO SOSTENIBILE* in Italian) equates the meaning by collocation of the English lexical item. This equivalence depends on a general correspondence between semantic preferences and between evaluative prosodies. On the other hand, colligational

patterns differ across languages because of their morphological and syntactic peculiarities.

### 6.1.2. The *Sustainable development Corpus*

As far as the discursive construction of sustainable development in the SusCorp is concerned (for a thorough description see § 5), the most frequent lexemes of the SusCorp (English) construct for the corpus an aboutness involving sustainable development, national and international politics, social and environmental issues. The style of the collection is characterised by an abundant use of reported speech and by a frequent reference to time. Among the most frequent lexemes shaping this aboutness and style for the SusCorp (English), *development*, *global*, *sustainable*, *goals*, *climate* and *change* are esteemed cultural keywords because of their impact in mirroring and building contemporary British political, social and cultural commitment in relation to sustainable development. *development*, *global*, *sustainable*, *goals* mirror the overall concerns of the collection, while *climate* and *change* highlight a paramount topic of this specimen of news discourse. The most frequent lexemes of the SusCorp (Hungarian) portray for the collection an aboutness that regards sustainable development, national and international politics, a focus on economic issues. The style of the subcorpus is marked by a meaningful presence of verbs reproducing existential processes and verbal processes. The most frequent lexemes of the SusCorp (Hungarian) host cultural keywords that reflect and contribute to the political, social and cultural engagement of Hungary for sustainable development: *fenntartható* 'sustainable' and *fejlesztés* 'development'. These two cultural keywords point at one of the most debated issues in contemporary Hungarian politics. The most frequent lexemes of the SusCorp (Italian) tackle the issues of sustainable development, of national and international politics, of the social, environmental and economic dimensions of sustainable development. The style of the SusCorp (Italian) is characterised by a significant reference to time. The lexemes *sostenibile* 'sustainable', *sviluppo* 'development', *obiettivi* 'goals', *sostenibilità* 'sustainability' can be regarded cultural keywords for the SusCorp (Italian), as they emphasise some of the most urgent political, social and cultural concerns of the Italian society in relation to sustainable development.

In relation to the statistically significant lexemes of the SusCorp, in the SusCorp (English), the most meaningful statistical keywords semantically revolve around the notion of sustainable development and of its social and environmental dimensions, as they are managed internationally. The economic dimension of sustainable

development is only hinted at. These notions are referred to in more specific terms compared to the equivalent references among the most frequent content words of the SusCorp (English). Nevertheless, the most meaningful statistical keywords of the SusCorp (English) do not contribute to the identification of new cultural keywords for the corpus. The most significant statistical keywords of the SusCorp (Hungarian) provide the collection with an aboutness characterised by a focus on sustainable development, on national and international politics, on migration and on climate change. The importance of these topics impacts on the selection of cultural keywords for the SusCorp (Hungarian): among the most salient statistical keywords of the collection, *FENNTARTHATÓSÁGI* '(of) sustainability', *FEJLŐDÉSI* '(of) development', *FENNTARTHATÓSÁG* 'sustainability' prove cultural keywords in relation to sustainable development; *MIGRÁCIÓ* 'migration', *MIGRÁCIÓS* '(of) migration', and *MIGRÁNS* 'migrant' feature as cultural keywords in relation to migration; *KLÍMAVÁLTOZÁS* 'climate change' and *ÉGHAJLATVÁLTOZÁS* 'climate change' are culturally key in relation to climate change. The most salient statistical keywords of the SusCorp (Italian) devote particular attention to sustainable development, to the political actors that sustainable development involves, and to the environmental, economic and social dimensions of sustainable development. The prominence of the environmental dimension of sustainable development among the statistical keywords of the SusCorp (Italian) impacts on the cultural keywords of the collection: the statistical keyword *GREEN* 'green' enhances the list of the cultural keywords identified in the SusCorp (Italian).

As far as the meaning by collocation of *sustainable* in the SusCorp is concerned, the semantics of the English *sustainable* makes it a quality associated to properties of novelty and internationality, and which characterises political and economic enterprises, as well as a general notion of growth. The meaning by collocation of *sustainable* is shaped through a semantic preference for lexemes related to the following matters: sustainable development and the 2030 Agenda (*development, goals, SDGs, goal, 17, 2030, agenda*); international organisations in an international dimension (*UN's, UN, global, world*); economy (*investing, capital, business, investment*); environment (*water*); development in general (*growth*); material processes (*adopted, achieve*); verbal processes (*said, says*). This semantic preference characterises content words that populate several colligational patterns. In all colligational patterns the adjective *sustainable* modifies noun phrases that are used independently as verb arguments or that modify prepositions in prepositional phrases that function as modifiers or circumstantials. Thanks to these lexical and grammatical preferences, *sustainable* acquires a neutral-to-positive evaluative prosody under the ecological framework

adopted in this study. The meaning by collocation of the Hungarian adjective *FENNTARTHATÓ* 'sustainable' makes it a property bound to national and international politics and characterising the economic and environmental spheres of life. The meaning by collocation of the Hungarian adjective is constructed thanks to a semantic preference for lexemes referring to sustainable development (*fejlődés* 'development', *fejlődési* 'development', *fejlődést* 'development', *célok* 'goals', *fejlesztési* 'developmental'), to politics (*országgyűlés* 'parliament', *bizottsága* 'its commission', *bizottságának* '(to) its commission', *ENSZ* 'United Nations', *tanács* 'council', *ülésén* 'on its sitting', *elnöke* 'its president'), to geo-political dynamics of nationality and internationality (*nemzeti* 'national', *Magyarország* 'Hungary', *világ* 'world', *európai* 'European'), to the economic and environmental dimensions of sustainable development (*gazdasági* 'economic', *környezeti* 'environmental'). The lexemes embodying these semantic preferences fill colligational patterns where *FENNTARTHATÓ* 'sustainable' belongs to adjective phrases consisting of the sole adjective or of combined adjectives, seldom left-modified by noun phrases, and usually modifying noun phrases marked by a wide variety of cases. The evaluative prosody of the adjective can be regarded neutral-to-positive if it is interpreted under the light of the ecological framework used in this research. The semantics of the Italian adjective *SOSTENIBILE* 'sustainable' is modelled alike thanks to its featuring a semantic preference for lexemes gathering in the following semantic areas: sustainable development and the 2030 Agenda (*sviluppo* 'development', *obiettivi* 'goals', *2030* '2030', *17* '17'); political actors (*fondazione* 'foundation', *italiana* 'Italian', *Nazioni* 'Nations' and *Unite* 'United'); general ethics (*equo* 'equitable'); places and transport system (*città* 'city; cities', *territorio* 'territory', *mobilità* 'mobility'); forward-directed aspiration (*futuro* 'future'); plan (*modello* 'model'). These collocating lexemes are found in several colligational patterns, which encompass the adjective *SOSTENIBILE* 'sustainable' in an adjective phrase that modifies a noun phrase which modifies a prepositional phrase in turn. Thanks to these collocational, colligational and semantic preferences, the evaluative prosody of the adjective *SOSTENIBILE* 'sustainable' is neutrally assessed under the light of the ecological framework of the current work.

### **6.1.3. The 2030 Agenda Corpus vs. the Sustainable development Corpus**

As it might have been noticed so far, the 2030 Agenda Corpus and the Sustainable development Corpus share similarities and differences in the discursive construction of sustainable development.

As far as the most frequent lexemes of the two corpora are concerned, it can be observed that the most frequent lexemes of the two collections are function words. This depends on structural features of Standard Average European (on this see, for instance, Heine and Kuteva 2006). The highest-ranking function words of the two corpora are the same and they differ only in their position in the list: for instance, in the 2030 Agenda Corpus *and* precedes *the* and the order of *to* and *of* is reversed. This is due to the different stylistic trends of the two collections.

In relation to the most frequent and statistically significant lexemes of the 2030 Agenda Corpus and of the Sustainable development Corpus, both corpora identify sustainable development as the essence of their aboutness and both of them consider sustainable development in a national and in an international perspective. Nevertheless, while the 2030 Agenda differentiates between developed and developing countries, in SusCorp there is no explicit mention to this classification of the world's states. In addition, the 2030 Agenda treats the social, environmental and economic dimensions of sustainable development in a more balanced way compared to SusCorp, where the economic and environmental dimensions of sustainability seems to predominate. Furthermore, the 2030 Agenda lexically deals with these issues in a broader and less specific way, whereas the SusCorp is more precise in detailing the dimensions of sustainable development. This higher degree of detail is reached through the use of more specialised lexicon. For example, elaborating on this point in the case of English, the most frequent content words of the SusCorp (English) partially overlap with the most frequent content words of the 2030 Agenda. Both collections feature *development, countries, global, sustainable, goals* and *international* among their highest-ranking content words. This overlap shapes a predictable common concern for *sustainable development* and for *sustainable development goals* but it also hints at a more interesting involvement for international dynamics (with the word types *global* and *international*) and for the condition of the world's *countries*. In addition to this, the SusCorp (English) introduces lexemes like *Ngo, DfID* and *Macquarie* among the statistically significant words of the corpus and it contributes further to the construction of sustainable development as an international, political enterprise but it does so in more specific terms compared to the 2030 Agenda. The same can be observed also for Hungarian, where the generic reference to political actors of the 2030 Agenda is substituted with a specific reference to political actors like *ÁDER 'Áder', SALLAI 'Sallai', TRUMP 'Trump', SZIJÁRTÓ 'Szijjártó'* and *MAJTÉNYI 'Majtényi'*. *Trump 'Trump'* is the protagonist also of the specificity of the Italian lexicon in the SusCorp (Italian), together with the Italian association *ASviS 'ASviS'*.

In terms of the semantics of *sustainable* and its Hungarian and Italian translational equivalents, both the 2030 Agenda Corpus and the SusCorp frame the adjective as a quality that goes in pair with other properties like inclusion and equality in characterizing lexemes that have to do with sustainable development, with the 2030 Agenda or with human activities that impact on the social, environmental and economic dimensions of sustainability. However, while in the 2030 Agenda Corpus the adjectives usually collocate with material processes and mental processes, in the Sustainable development Corpus they collocate with material processes and with verbal processes.

This difference might be due to the diverse nature of the two discourses represented by the 2030 Agenda Corpus and by the SusCorp: the former reproduces political discourse and the latter exemplifies news discourse. Political discourse aims at ideologically engaging the addressee in reflecting on political issues and in acting accordingly (see, among others, Fairclough 1989); news discourse intends to inform its addressees about newsworthy topics and it ends up stimulating ideologically charged actions as a consequence of this information process (see, among others, Fowler 1991).

The difference between the 2030 Agenda Corpus and the Sustainable development Corpus can also depend on the goals of the texts included in the two collections. The 2030 Agenda aspires to brief the world's nations about the resolution taken at the end of 2015 UN's summit on sustainable development and it wishes to convince world leaders and citizens of the importance of undertaking the recommended measures in favour of sustainability (Kanie et al. 2017). The articles published in the broadsheets attempt to inform their addressees about matters that regard sustainable development by and large; they frequently do so by reporting the declarations of official documents or of authorities who intervene or who are interviewed in order to strengthen the article's core points.

## **6.2. An ecological interpretation of the results**

An ecological interpretation of the outcome of a linguistic study consists in explaining the reasons that might have led to use specific linguistic patterns and in evaluating these reasons under the light of a precise ecological framework. In ecological discourse analysis these reasons are usually called stories, and they feature as cognitive narratives that guide individuals to perceive the world. Stories can have different functions and forms. Some of them are salience, ideology and evaluation (Stibbe 2015).

In the current study, cultural keywords and meaning by collocation are interpreted under the light of the ecolinguistic notions of salience, ideology and evaluation. As it has already been written in § 1.3.2, salience is a story that an entity or a concept is important or worthy of consideration, ideology is the belief system that founds and influences one's ideas as a consequence of their belonging to a certain community, evaluation is a story according to which an entity or a concept is good or bad. Cultural keywords are endowed with the quality of salience as they pinpoint political, social and cultural notions that are regarded as particularly worthy of consideration in the political, social and cultural context that has given rise to the texts where cultural keywords are identified. Meaning by collocation reflects salience, ideology and evaluation as it encompasses the system of beliefs of a person or of a community and their judging an entity or a concept as important or as good or bad by means of semantic preference and evaluative prosody.

As the results of the corpus-assisted study of the 2030 Agenda Corpus and of the Sustainable development Corpus have shown, the ideological stance reflected by the collections revolves around a positive attitude towards sustainable development. In both corpora, this positive attitude is bolstered by encouraging governments and citizens to act in favour of sustainability, which is regarded as the arriving point of a journey filled with challenges to face. Governments and citizens are asked to become conscious agents of the path they should take in order to reach sustainable development and they are compelled to strive in order to make their countries follow the same path. This path goes especially along an economic route, although it touches upon social and environmental fields too. This touching upon social and environmental fields revolves around human beings, while it almost completely neglects the other beings of the environment, namely animals, plants and natural elements. Human beings and societies are written about in relation to forms of wellbeing that encompass health and the eradication of poverty, but that exclude issues like education. The connotation of sustainable development and of sustainable actions, actors and values is positive thanks to the positive evaluative prosody that the related lexical items are imbued with.

These observations result from the salience of the lexemes highlighted among the most frequent and statistically significant lexemes of the corpora. The frequency and statistical prominence of these lexemes make them salient in ecolinguistic terms because they render them important and worthy of consideration from a linguistic point of view. The same applies also to the collocation networks of the studied lexemes: the strongest collocates of these lexemes can be deemed salient in ecological

terms. Furthermore, collocation networks provide also with ecolinguistic evaluation of the concepts embodied in the lexemes under inquiry, and both salience and evaluation contribute to shape the ideology underlying the corpora.

Under the light of the ecological framework of this study, the support and engagement with sustainable development is regarded as valuable. A sustainable kind of development, in fact, is the only chance that the world has to develop without causing major damage to its ecology. This implies that the forms that development might take significantly depend on one of the dichotomies put forward by the 2030 Agenda, namely the distinction between developed and developing countries: developed and developing countries ought to address sustainability matters with different approaches that consider the initial development conditions of the countries and the desired arriving point of the developmental process. As a consequence, the *growth* that appears frequently among the most prominent lexemes of the two corpora should be pondered according to the initial level of development of the countries: least developed countries ought to be entitled more rights and opportunities to improve and grow towards full development, whereas more developed countries should halt their race towards full development not to deprive developing countries of necessary resources (Meadows et al. 1972).

The commitment to sustainable development is ecologically valuable. However, the focus on the economic dimension of sustainability causes a neglect of the social and of the environmental aspects of sustainable development. At the same time, when social and environmental impacts of sustainable development are taken into account, human beings are considered more often and with more care compared to animals, plants and natural elements.

In ecolinguistic terms, economical and social lexemes and concepts are salient in the political discourse of the 2030 Agenda Corpus and in the news discourse of the Sustainable development Corpus, while environmental lexemes and concepts are erased. Within society and environment, human beings are salient while animals, plants and natural elements are erased. This is in line with the observations made by Naeem et al. (2016) in relation to news discourse, and by Stibbe (2012), but it clashes with the ecological framework adopted for the present research. The almost complete absence of animals, plants and natural elements in the corpora appears to be detrimental and harmful for a balanced development of the world. Moreover, the recurrent reference to animals, plants and natural elements as if they were sheer resources to be exploited for the wellbeing of humans reflects a distorted picture of the world's order.

### 6.3. Discussion

The study of the discursive construction of sustainable development in the 2030 Agenda Corpus and in the SusCorp by means of cultural keywords and meaning by collocation, within the theoretical framework of corpus approaches to the ecological analysis of discourse and from a cross-linguistic perspective has shown both strengths and weaknesses.

The study of the discursive construction of sustainable development has been carried out within the theoretical framework of ecological analysis of discourse and by relying on Williams's (1983) theory on cultural keywords and Firth (1957a, 1957b) and Sinclair's (1991) theories on meaning by collocation. The theory provided by Williams (1983) on cultural keywords has offered a valuable opportunity to collect lexical items that have a meaningful role in a culture. Within an ecological framework, Williams's (1983) cultural keywords have highlight salient terms that contribute to shaping the culture originating the investigated texts. In the present research, cultural keywords have shown that the British, Hungarian and Italian cultures, as they are mirrored in the 2030 Agenda and in the press, are rooted in concepts like development, technology, economy, and climate, and in values like inclusivity and sustainability. In addition, frequency lists have aided in recognizing salient lexical items among economic and social terms, and they have led to characterize environmental terms referring to animals, plants and natural elements as erased.

Firth (1957a, 1957b) and Sinclair's (1991) meaning by collocation have contributed to outlining the ideological construction and the evaluation of the lexemes investigated in a study. Meaning by collocation has sketched the semantics that lexical items are endowed with as a result of the ideological stance of the authors of the texts under investigation: the study of corpus semantics has allowed to assess the ideology lying in discourse. The ideological construction of the semantics of lexical items has resulted from the peculiar collocational behaviour of the lexemes; this collocational behaviour is made of sheer collocates, but also of ideologically marked colligational patterns, semantic preferences, and evaluative prosodies. Evaluative prosodies are hallmarks of evaluation. Thus, corpus semantics has contributed to identify the evaluation that lexical items are repeatedly assigned in discourse. In the current study, the meaning by collocation of the English *sustainable development*, *sustainable*, and *sustainability*, and of their Hungarian and Italian equivalents has strengthened an ideological and evaluational framework according to which sustainable actions, actors

and values are positively connoted and regard international endeavours characteristically bound to economic matters.

The methodology employed to match ecological discourse analysis with corpus analysis has come in handy. It has enabled to carry out quantitative and qualitative analyses of ideology, evaluation and salience, while showing how these ecolinguistic concepts could be explored with the aid of corpora and corpus tools. Quantitative and qualitative research methods were continuously intertwined in the study: the quantitative retrieval of the most frequent and statistically significant lexemes of the corpora and the quantitative computation of statistically significant collocates have been constantly matched with qualitative, in-depth descriptions, explanations and interpretations of the results of the quantitative studies. To this extent, one of the most meaningful strengths of this research consists in its contributing to the still underexplored field of corpus-assisted, ecological analysis of discourse.

The second strength of this research regards its involving three languages (i.e. English, Hungarian and Italian) that are rarely analysed together in an all-encompassing study. In relation to this, the methodological choices made to expand on the 2030 Agenda Corpus and on the Sustainable development Corpus have proven fruitful despite the different structural characteristics of the three languages investigated in the research.

However, the different structural characteristics of the three languages investigated in the research have also triggered one of the most meaningful weaknesses of the study, namely the lack of lemmatisation for the extraction of collocation networks. As it has been elaborated on also in § 4.5, in the Hungarian and Italian cases, the lack of lemmatisation for the extraction of collocation networks has resulted in the identification of collocational patterns including the most frequent and statistically significant forms of lemmas and excluding the least frequent and least statistically significant forms of the same lemmas. While this has been interesting for discourse analytical purposes (as Baker 2006 states), it might have partially skewed the outcome of the extraction of collocation networks: in fact, it could have left out lemmas whose word forms were not frequent or statistically significant enough to appear in the collocational pattern, but it could have also assigned an incorrect statistical or frequency weight to a lemma included in the collocation network (Kübler and Zinsmeister 2015; see also Lenci et al. 2005).

## 6.4. Future developments

The present dissertation can be regarded as a starting point for further studies of the discursive construction of sustainable development. Within the theoretical framework of ecological discourse analysis, the research questions and the approach to the study could be expanded: they might integrate a thorough analysis of the transitivity of the protagonists of sustainable development, namely human beings, animals, plants and natural elements; they might involve a fine-grained study of the metaphors appearing in relation to sustainable development in the 2030 Agenda Corpus and in the Sustainable development Corpus; they might include an overview of the modality of the texts belonging to the collections.

Research on corpus semantics could be enhanced with the aid of *distributional semantics*. Distributional semantics is a research field that explores the semantics of a language's lexicon by computationally assessing the meaning of lexical items with a study of their distributional properties and with an evaluation of their similarity and difference (Lenci 2008). This approach would fit the study of the meaning of lexemes related to sustainable development extremely well.

Also the data could be boosted by adding other examples of political discourse on sustainable development and by enriching the collection of news discourse on sustainable development; in addition, data could be improved with systematic annotation of morphological, syntactic and pragmatic features, especially if the research questions were expanded to cover transitivity.

The methodology employed for the study could be enhanced in the following ways: it could turn into a more sophisticated, computational analysis of the corpora; it could add the independent study of SusCorp's newspapers; it could improve by enriching the analysis and the interpretation of the results through methodological triangulation.

Moreover, the current research could be enhanced with a study of the textual features of the corpora or with a study of the perception that addressees have of discursive construction of sustainable research. As it is right now, in fact, the dissertation offers only an overview of the textual and lexical semantics of *sustainable* and of the related items *sustainable development* and *sustainability*. It does so by focusing on what has happened from a linguistic point of view with the release of the UN's 2030 Agenda for Sustainable Development in the British, Hungarian, and Italian political and news panorama. This reflects recurrent linguistic uses that portray the ideological stance of the singular and collective authors of the 2030 Agenda and of the British,

Hungarian, and Italian broadsheet articles published between 2016 and 2018 on sustainable development. As Fowler (1991) notices, this might be influenced by the ideological stance of the addressees of the documents, and it might influence it back. This ideological impact can only be guessed with a functionalist study of the kind of this dissertation. It might be tested with an assessment of the perception that addressees gather from the discursive construction of sustainable development especially in the press, which is more easily accessed by laymen. This kind of research would require carrying out perception studies on the reception of discourse by readers (as in Nightingale 2011, Ross and Nightingale 2003).

Another potential development of the research could be of a pedagogical kind. This pedagogical development of the research follows several studies on the application of corpus-assisted discourse analysis to teaching and it is well explained by Scott and Tribble (2006). Scott and Tribble (2006) state that

an attention to *texts* is the best starting point for a corpus informed language pedagogy. Rather than seeing corpus data as entirely abstracted from its linguistic and social context, the studies stress the obligation on the researcher to re-connect with the text (and, where possible, with its context of production) in order to build accounts of language in use which will have value for teachers and students of language alike. (Scott and Tribble 2006: x)

Following Scott and Tribble (2006), the current ecological, corpus-assisted study of discourse could be adopted for the teaching of languages but also for the teaching of environmental education. In line with Stibbe's (2015) ideas on salience and erasure and with Hoey's (2005) idea of collocational priming, in fact, guidance in the recognition of the ways environmental issues and world dynamics are constructed and represented in discourse by means of significant lexemes and through their meaning by collocation could be a paramount way to stimulate new generations to reflect and act towards a sustainable future.



## References

- Abondolo, D. (1998). Hungarian. In D. Abondolo (Ed.). *The Uralic Languages* (pp. 428-456). London/New York: Routledge.
- Aijmer, K. (2009). Parallel and comparable corpora. In A. Lüdeling, and M. Kytö (Eds.). *Corpus linguistics: an international handbook* (Volume 1; pp. 275-292). Berlin: Walter de Gruyter.
- Aijmer, K., and Altenberg, B. (1996). Introduction. In K. Aijmer, B. Altenberg, and M. Johansson (Eds.). *Languages in Contrast: Papers from a symposium on text-based cross-linguistic studies, Lund 4-5 March 1994* [Lund Studies in English 88] (pp. 11-16). Lund: Lund University Press.
- Aijmer, K., and Rühlemann, C. (2015). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press.
- Alexander, R.J. (2002). Everyone is talking about 'sustainable development'. Can they all mean the same thing? Computer discourse analysis of ecological texts. In A. Fill, H. Penz, e W. Trampe (ed.). *Colourful Green Ideas: Papers from the Conference 30 Years of Language and Ecology* (pp. 239-254). Bern: Peter Lang.
- Alexander, R.J. (2009). *Framing Discourse on the Environment: A Critical Discourse Approach*. New York/London: Routledge.
- Alexander, R.J. (2018). Investigating Texts about Environmental Degradation Using Critical Discourse Analysis and Corpus Linguistics Techniques. In A.F. Fill and H. Penz (Eds.). *The Routledge Handbook of Ecolinguistics* (pp. 196-210). New York/London: Routledge.
- Alexander, R., and Stibbe, A. (2014). From the analysis of ecological discourse to the ecological analysis of discourse. *Language Sciences*, 41: 104-110.
- Algina, J., Keselman, H., and Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size. A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, 10(3): 317-328.
- Allora, A., and Barbera, M. (2007). Il problema legale dei corpora: Prime approssimazioni. In M. Barbera, E. Corino, and C. Onesti (Eds.). *Corpora e linguistica in rete* (pp. 109-118). Perugia: Guerra Edizioni.
- Altenberg, B., and Granger, S. (Eds.). (2002a). *Lexis in Contrast. Corpus-based approaches*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Altenberg, B, and Granger, S. (2002b). Recent trends in cross-linguistic lexical studies. In B. Altenberg, and S. Granger (Eds.). *Lexis in Contrast: Corpus-based approaches* (pp. 3-48). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Antelmi, D. (2018). *Verdiparole. Un'analisi linguistica del discorso green*. Milano: Mimesis.
- Anthony, L. (2017a). AntFileConverter (Version 1.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Anthony, L. (2017b). AntPConc (Version 1.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Anthony, L. (2020). *AntConc (Version 3.5.9)* [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>.
- Aragrande, G. (2018). Migrants in Translation: A Corpus-Based Approach to the Representation of Migrants by Four News Broadcasting Channels. In I. Ranzato and S. Zanotti (Eds.). *Linguistic and Cultural Representation in Audiovisual Translation* (pp. 121-146). New York/London: Routledge.
- Archer, D. (Ed.). (2009). *What's in a Word List? Investigating word frequency and key word extraction*. London: Ashgate.
- Aston, G., and Burnard, L. (1998). *The BNC handbook. Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
- Atanasova, D., and Koteyko, N. (2017). Metaphors in Guardian Online and Mail Online Opinion-page Content on Climate Change: War, Religion, and Politics. *Environmental Communication*, 11(4): 452-469.
- Atkins, S., Clear, J., and Ostler, N. (1992). Corpus Design Criteria. *Literary and Linguistic Computing*, 7(1): 1-16.
- Baeza-Yates, R., and Ribeiro-Neto, B. (2010). *Modern Information Retrieval: The concepts and technology behind search*. London: Addison Wesley.
- Baker, M. (1995). Corpora in translation studies. An overview and some suggestions for future research. *Target*, 7(2): 223-242.
- Baker, P. (2004). Querying Keywords. Questions of Difference, Frequency, and Sense in Keywords Analysis. *Journal of English Linguistics*, 32(4): 346-359.
- Baker, P. (2005). *Public Discourses of Gay Men*. Abingdon/New York: Routledge.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London/New York: Continuum.
- Baker, P. (2010a). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.

- Baker, P. (2010b). Representations of Islam in British broadsheet and tabloid newspapers 1999-2005. *Journal of Language and Politics*, 9(2): 310-338.
- Baker, P. (2016). The shapes of collocation. *International Journal of Corpus Linguistics*, 21(2): 139-164.
- Baker, P., Hardie, A., and McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Baker, P., Gabrielatos, C., Khosravini, M., Krzyzanowski, M., McEnery, T., and Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society*, 19(3): 273-306.
- Baker, P., Gabrielatos, C., and McEnery, T. (2013a). *Discourse Analysis and Media Attitudes. The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.
- Baker, P., Gabrielatos, C. and McEnery, T. (2013b). Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998-2009. *Applied Linguistics*, 34(3): 255-278.
- Baker, P., and Levon, E. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication*, 9(2): 221-236.
- Baker, P., and McEnery, T. (2015). Introduction. In P. Baker and T. McEnery (Eds.). *Corpora and Discourse Studies. Integrating Discourse and Corpora* (pp. 1-19). Houndmills/ New York: Palgrave Macmillan.
- Baker, P., and Egbert, J. (2016). Introduction. In P. Baker, and J. Egbert (Eds.). (2016). *Triangulating Methodological Approaches in Corpus-Linguistic Research* (pp. 1-19). New York/London: Routledge.
- Baker, P., and Vessey, R. (2018). A corpus-driven comparison of English and French Islamist extremist texts. *International Journal of Corpus Linguistics*, 23(3): 255-278.
- Baker, S. (2006). *Sustainable Development*. London/New York: Routledge.
- Balfour, J. (2019). 'The mythological marauding violent schizophrenic': Using the WordSketch tool to examine representations of schizophrenic people as violent in the British press. *Journal of Corpora and Discourse Studies*, 2: 40-64.
- Barát, E. (2017). Populist discourses in the Hungarian public sphere: From right to left (and beyond)?. *Journal of Language and Politics*, 16(4): 535-555.
- Barnbrook, G. (1996). *Language and Computers. A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.

- Baroni, M., and Bernardini, S. (2004). *BootCaT: Bootstrapping Corpora and Terms from the Web*. Proceedings of LREC 2004. Available from <https://pdfs.semanticscholar.org/a1ea/c69123e1acbe3f248e1ce85e94ae67b0bbe8.pdf>.
- Baroni, M., and Ueyama, M. (2004). *Retrieving Japanese specialized terms and corpora from the World Wide Web*. Proceedings of KONVENS 2004. Available from <http://sslmit.unibo.it/~baroni/publications/konvens2004/jpbootKONV.pdf>.
- Bartlett, T. (2012). *Hybrid Voices and Collaborative Change: Contextualising Positive Discourse Analysis*. London/New York: Routledge.
- Bassi, E. (2010). A contrastive analysis of keywords in newspaper articles on the “Kyoto Protocol”. In M. Bondi, and M. Scott (Eds). *Keyness in Texts* (pp. 207-218). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Baudrillard, J. (1994). *Simulacra and simulation*. Ann Arbor: University of Michigan Press.
- Bayley, P., and Williams, G. (Eds.). (2012). *European Identity: What the Media Say*. Oxford: Oxford University Press.
- Beaugrande, R. de (1999). Discourse studies and the ideology of ‘liberalism’. *Discourse Studies*, 1(3): 259-295.
- Beaugrande, R. de (2001). Interpreting the Discourse of H.G. Widdowson: A Corpus-Based Critical Discourse Analysis. *Applied Linguistics*, 22(1): 104-121.
- Beaugrande, R. A. de, and Dressler, W. U. (1981). *Introduction to text linguistics*. London/New York: Longman.
- Bednarek, M. (2006). *Evaluation in Media Discourse*. London/New York: Continuum.
- Bednarek, M. and Caple, H. (2012). *News Discourse*. London/New York: Continuum.
- Benwell, B., and Stokoe, E. (2006). *Discourse and Identity*. Edinburgh: Edinburgh University Press.
- Berber Sardinha, T. (2004). *Lingüística de Corpus*. Barueri, SP: Editora Manole.
- Best, S. (1998). Murray Bookchin’s Theory of Social Ecology: An Appraisal of “The Ecology of Freedom”. *Organization & Environment*, 11(3): 334-353.
- Bevitori, C., and Johnson, J. H. (2017). Human Mobility and Climate Change at the Crossroad. A Diachronic Corpus-Assisted Discourse Analysis of the Nexus in UK and US Newspaper Discourse. *Anglistica AION An Interdisciplinary Journal*, 21: 1-19.
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4): 243- 257.

- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D. Conrad, S., and Leech, G. (2002). *Longman Student Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Biber, D., Conrad, S., and Cortes, V. (2004). *If you look at...: Lexical Bundles in University Teaching and Textbooks*. *Applied Linguistics*, 25(3): 371-405.
- Bookchin, M. (1989). *Per una società ecologica*. Milano: Elèuthera.
- Bookchin, M. (1995). *L'ecologia della libertà: Emergenza e dissoluzione della gerarchia*. Milano: Elèuthera.
- Bookchin, M. (2005). Society and Ecology. In J.S. Dryzek and D. Schlosberg (Eds.). *Debating the Earth: The Environmental Politics Reader* (pp. 385-398). Oxford: Oxford University Press.
- Bondi, M., and Scott, M. (Eds.). (2010). *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bondi, M. (2010). Perspectives on keywords and keyness. An introduction. In M. Bondi, and M. Scott (Eds.). *Keyness in Texts* (pp. 1-18). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bourdieu, P. (1983). *La distinzione: Critica sociale del gusto*. Bologna: Il Mulino.
- Bourdieu, P. (1986). The forms of capital. In J.G. Richardson (Ed.). *Handbook of Theory and Research for the Sociology of Education* (pp. 241-258). New York: Greenwood Press.
- Bourdieu, P. (1988). *La parola e il potere: L'economia degli scambi linguistici*. Napoli: Guida.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., and Yergeau, F. (Eds.). (2008). Extensible Markup Language (XML) 1.0. Fifth edition. World Wide Web Consortium. Available from <https://www.w3.org/TR/xml/>.
- Breeze, R. (2013). Critical Discourse Analysis and its critics. *Pragmatics*, 21(4): 493-525.
- Brezina, V. (2018a). *Statistics in Corpus Linguistics. A Practical Guide*. Cambridge: Cambridge University Press.
- Brezina, V. (2018b). Statistical choices in corpus-based discourse analysis. In C. Taylor, and A. Marchi (Eds.). *Corpus Approaches to Discourse: A Critical Review* (pp. 259-280). Abingdon/New York: Routledge.
- Brezina, V., McEnery, T. and Wattam, S. (2015). Collocations in context. A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2): 139-173.
- Brezina, V., Weill-Tessier, P., and McEnery, A. (2020). #LancsBox v.5.x [Computer Software]. Available from <http://corpora.lancs.ac.uk/lancsbox>.

- Bublitz, W. (2003). Emotive prosody: How attitudinal frames help construct context. In E. Mengel, H. J. Schmid, and M. Steppat (Eds.). *Proceedings of Anglistenstag 2002, Bayreuth* (pp. 381-391). Trier: Wissenschaftlicher Verlag.
- Burnard, L. (2005). Metadata for corpus work. In M. Wynne (Ed.). (2005). *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 40-57). Available from <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm> .
- Bušta, J., Herman, O., Jakubíček, M., Krek, S., and Novak, B. (2017). *JSI Newsfeed Corpus*. Paper presented at 9<sup>th</sup> International Corpus Linguistics Conference. University of Birmingham (United Kingdom), 25-28 July 2017.
- Caimotto, M.C., and Molino, A. (2011). Anglicisms in Italian as Alerts to Greenwashing: A Case Study. *Critical Approaches to Discourse Analysis across Disciplines*, 5(1): 1-16.
- Caldas-Coulthard, C.R., and Coulthard, M. (Eds.). (1996). *Texts and Practices: Readings in Critical Discourse Analysis*. London/New York: Routledge.
- Carapinha, C., and Plag, C. E. (2019). Tradução de Marcadores Discursivos portugueses para alemão: ‘na verdade’, o que se ganha e o que se perde?. *REDIS: Revista de Estudos do Discurso*, 8: 12-40.
- Cartoni, B., Zufferey, S., and Meyer, T. (2013). Using the Europarl corpus for cross-linguistic research. *Belgian Journal of Linguistics*, 27: 23-42.
- Castello, E., and Gesuato, S. (2019). Pope Francis’s *Laudato si’*. A corpus study of environmental and religious discourse. *Lingue e Linguaggi*, 29: 121-145.
- Chilton, P., and Wodak, R. (2005). Preface. In R. Wodak and P. Chilton (Eds.). *A New Agenda in (Critical) Discourse Analysis: Theory, methodology and interdisciplinarity* (pp. XI-XVIII). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Chouliaraki, L., and Fairclough, N. (1999). *Discourse in Late Modernity. Rethinking Critical Discourse Analysis*. Edinburgh: Edinburgh University Press.
- Chowdhury, G. G. (2004). *Introduction to Modern Information Retrieval*. London: Facet Publishing.
- Church, K. W., and Hanks, P. (1990). Word association norms, mutual information and lexicography. *Computational Linguistics*, 16(1): 22-29.
- Church, K. W., Gale, W., Hanks, P., and Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.). *Lexical Acquisition* (pp. 115-164). Englewood Cliff, NJ: Erlbaum.
- Cimadamore, A.D., Koehler, G., and Pogge, T. (Eds.). (2016). *Poverty and the Millennium Development Goals: A Critical Look Forward*. London: Zed Books.

- Clear, J. (1993). From Firth principles. Computational tools for the study of collocation. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.). *Text and Technology. In Honour of John Sinclair* (pp. 271-292). Amsterdam/New York: John Benjamins.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Coleman, J.S. (1988). Social Capital in the Creation of Human Capital. *The American Journal of Sociology*, 94: S95-S120.
- Comrie, B. (1981). *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Basil Blackwell.
- Conte, M. E. (Ed.). (1977). *La linguistica testuale*. Milano: Feltrinelli.
- Convington, M. A., and McFall, J. D. (2010). Cutting the Gordian Knot. The moving average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2): 94-100.
- Cortes, V. (2002). Lexical bundles in Freshman composition. In R. Reppen, S.M. Fitzmaurice and D. Biber (Eds.). *Using Corpora to Explore Linguistic Variation* (pp. 131-145). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum. Origins and functional analysis developments. In V. Cortes and E. Csomay (Eds.). *Corpus-based Research in Applied Linguistics. Studies in Honor of Doug Biber* (pp. 197-216). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Croft, W. (2003 [1990]). *Typology and Universals* (2<sup>nd</sup> Edition). Cambridge: Cambridge University Press.
- Culpeper, J. (2001). Computers, language and characterisation: An analysis of six characters in Romeo and Juliet. *Conversation in life and in literature. Papers from the ASLA symposium*, Uppsala, 11-30.
- Daille, B. (1995). *Combined approach for terminology extraction. Lexical statistics and linguistic filtering*. UCREL Technical Papers, 15. Lancaster: Lancaster University.
- Dardano, M., and Trifone, P. (1997). *La nuova grammatica della lingua italiana*. Bologna: Zanichelli.
- Dayrell, C. (2019). Discourses around climate change in Brazilian newspapers: 2003-2013. *Discourse & Communication*, 13(2): 149-171.
- Dayrell, C., and Urry, J. (2015). Mediating climate politics: The surprising case of Brasil. *European Journal of Social Theory*, 18(3): 257-273.
- Deignan, A., Semino, E., and Paul, S.E. (2017). Metaphors of Climate Science in Three Genres: Research Articles, Educational Texts, and Secondary School Student Talk. *Applied Linguistics*: 1-26.

- Del Fante, D. (2018). Where Are They From? A Corpus-Assisted Study of the Geographical Representation of Migrants in the UK and Italian Quality Press in 2016. *I-LanD Journal – Identity, Language and Diversity*, 1: 28-52.
- Devall, B., and Sessions, G. (1985). *Deep ecology*. Salt Lake City: Gibbs Smith.
- Diessel, H. (2019). *The Grammar Network. How Linguistic Structure is Shaped by Language Use*. Cambridge: Cambridge University Press.
- Dijk, T.A. van (1977). *Text and Context. Explorations in the semantics and pragmatics of discourse*. London/New York: Longman.
- Dijk, T.A. van (1991). *Racism and the Press*. London/New York: Routledge.
- Dijk, T.A. van (1998). *Ideology: A Multidisciplinary Approach*. London/Thousand Oaks/New Delhi: SAGE Publications.
- Dijk, T.A. van (2005). *Racism and Discourse in Spain and Latin America*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Dijk, T. A. van (2009). *Society and Discourse. How Social Contexts Influence Text and Talk*. Cambridge: Cambridge University Press.
- Doyle, P. G. (2003) *Replicating corpus linguistics. A corpus-driven investigation of lexical networks in text*. PhD dissertation, Lancaster University.
- Dressler, W. U. (1972). *Einführung in die Textlinguistik*. Tübingen: Niemeyer.
- Duguid, A. (2010). Newspaper discourse informalisation: a diachronic comparison from keywords. *Corpora*, 5(2): 109-138.
- Duguid, A., and Partington, A. (2018). Absence. You don't know what you're missing. Or do you? In C. Taylor, and A. Marchi (Eds.). *Corpus Approaches to Discourse: A Critical Review* (pp. 38-59). Abingdon/New York: Routledge.
- Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1), 61-74.
- Durrant, P., and Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2): 125-155.
- Ebeling, J., and Ebeling, S. O. (2013). *Patterns in Contrast*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Egbert, J., and Baker, P. (2020). Introduction. In J. Egbert and P. Baker (Eds.). *Using Corpus Methods to Triangulate Linguistic Analysis* (pp. 1-21). New York/London: Routledge.
- Everitt, B. S. (2002). *The Cambridge Dictionary of Statistics* (2<sup>nd</sup> Edition). Cambridge: Cambridge University Press.

- Evert, S. (2004). *The Statistics of Word Cooccurrences. Word Pairs and Collocations*. PhD dissertation, Universität Stuttgart.
- Evert, S. (2008). Corpora and collocations. In A. Lüdeling, and M. Kytö. *Corpus Linguistics: An International Handbook* (Volume 2; pp. 1212-1248). Berlin/New York: Mouton de Gruyter.
- Fairclough, N. (1985) Critical and descriptive goals in discourse analysis. *Journal of Pragmatics*, 9(6): 739-763.
- Fairclough, N. (1989). *Language and power*. London: Longman.
- Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity Press.
- Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. London/New York: Longman.
- Fairclough, N. (2003). *Analysing Discourse: Textual analysis for social research*. London/New York: Routledge.
- Fairclough, N., and Wodak, R. (1997). Critical Discourse Analysis. In T. van Dijk (Ed.). *Discourse as Social Interaction* (pp. 258-284).
- Fiammenghi, C., and Pinnavaia, L. (2019). Complementing and Contradictory Meanings in the Discourse of Environment: The Case of *global warming* and *climate change*. *Textus*, XXXII(1): 37-53.
- Fill, A., and Mülhäusler, P. (2001). Introduction. In A. Fill and P. Mülhäusler, P. (Eds.). *The Ecolinguistics Reader. Language, Ecology and Environment* (pp. 1-9). London/New York: Continuum.
- Firth, J. R. (1957a). A synopsis of linguistic theory. 1930-1955. In J. R. Firth. *Studies in Linguistic Analysis. Special Volume of the Philological Society* (pp. 1-32). Oxford: Basil Blackwell.
- Firth, J. R. (1957b). Modes of Meaning. In J. R. Firth. *Papers in Linguistics. 1935-1951* (pp. 190-215). London/New York/Toronto: Oxford University Press.
- Flowerdew, J., and Richardson, J.E. (Eds.). (2017a). *The Routledge Handbook of Critical Discourse Studies*. London/New York: Routledge.
- Flowerdew, J., and Richardson, J.E. (2017b). Introduction. In J. Flowerdew and J.E. Richardson (Eds.). *The Routledge Handbook of Critical Discourse Studies* (pp. 1-10). London/New York: Routledge.
- Forest, R. (2007). *'Tell us what you really think'. Collocation and discourse in an intertextual corpus*. PhD dissertation, City University of Hong Kong.
- Formato, F. (2019). *Gender, Discourse and Ideology in Italian*. Cham: Palgrave Macmillan.
- Foucault, M. (1966). *Les mots et les choses: Une archéologie des sciences humaines*. Paris: Gallimard.

- Foucault, M. (1972). *The Archaeology of Knowledge*. London: Tavistock.
- Foucault, M. (1981). The order of discourse. In R. Young (Ed.). *Untying the text: A post-structural anthology*. Boston: Routledge & Kegan Paul, pp. 48-78.
- Fowler, R. (1981). *Literature as Social Discourse: The Practice of Linguistic Criticism*. London: Batsford Academic and Educational LTD.
- Fowler, R. (1991). *Language in the News. Discourse and Ideology in the Press*. London/New York: Routledge.
- Fowler, R. (1996a [2009]). *Linguistic Criticism [2<sup>nd</sup> Edition]*. Oxford: Oxford University Press.
- Fowler, R. (1996b). On critical linguistics. In C.R. Caldas-Coulthard and M. Coulthard. *Texts and Practices: Readings in Critical Discourse Analysis* (pp. 3-14).
- Fowler, R., Hodge, R., Kress, G., and Trew, T. (1979[2019]). *Language and control [2<sup>nd</sup> Edition]*. London/Boston/Henley: Routledge & Kegan Paul.
- Francis, G. (1993). A Corpus-Driven Approach to Grammar. Principles, Methods and Examples. In M. Baker, G. Francis, and E. Tognini-Bonelli. *Text and Technology. In Honour of John Sinclair* (pp. 137-156). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Freake, R. (2012). A cross-linguistic corpus-assisted discourse study of language ideologies in Canadian newspapers. *Proceedings of the 2011 Corpus Linguistics Conference, Birmingham University*: 1-11.
- Freake, R, Gentil, G., and Sheyholislami, J. (2011). A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society*, 22(1): 21-47.
- Freddi, M. (2013). Corpus Methods in Syntax. In S. Luraghi and C. Parodi (Eds.). *The Bloomsbury Companion to Syntax* (pp. 51-62). London/New Delhi/New York/Sydney: Bloomsbury.
- Freddi, M. (2014). *Linguistica dei corpora*. Roma: Carocci.
- Friginal, E., and Hardy, J. (2014). *Corpus-Based Sociolinguistics: A Guide for Students*. London/New York: Routledge.
- Froehlich, A. (Ed.). (2019). *Embedding Space in African Society: The United Nations Sustainable Development Goals 2030 Supported by Space Applications*. Cham: Springer.
- Fruttaldo, A., and Cambria, M. (2017). Combining Qualitative and Quantitative Approaches to Discourse Analysis: In Conversation with Gerlinde Mautner and Alan Partington. *Imago: A Journal of the Social Imaginary*, 9: 284-304.

- Fusari, S. (2016). The permeable context of institutional and newspaper discourse: A corpus-based functional case study of the European sovereign debt crisis. In D.R. Miller and P. Bayley (Eds.). *Hybridity in Systemic Functional Linguistics: Grammar, Text and Discursive Context* (pp. 306-333). Sheffield/Bristol: Equinox.
- Gabrielatos, C. (2007). Selecting query terms to build a specialized corpus from a restricted-access database. *ICAME Journal*, 31: 5-43.
- Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor, and A. Marchi (Eds.). *Corpus Approaches to Discourse: A Critical Review* (pp. 225-258). Abingdon/New York: Routledge.
- Gabrielatos, C., and Baker, P. (2008). Fleeing, sneaking, flooding: a corpus analysis of discourse constructions of refugees and asylum seekers in the UK Press 1996-2005. *Journal of English Linguistics*, 36(1): 5-38.
- Gabrielatos, C., and Marchi, A. (2011). *Keyness. Matching metrics to definitions*. Paper presented at Corpus Linguistics in the South 1. University of Portsmouth (United Kingdom), 5 November 2011. Available from [https://eprints.lancs.ac.uk/id/eprint/51449/4/Gabrielatos\\_Marchi\\_Keyness.pdf](https://eprints.lancs.ac.uk/id/eprint/51449/4/Gabrielatos_Marchi_Keyness.pdf).
- Gabrielatos, C., and Marchi, A. (2012). *Keyness. Appropriate metrics and practical issues*. Paper presented at CADS International Conference. University of Bologna (Italy), 13-15 September 2012.
- Gatto, M. (2014). *The Web As Corpus: Theory and Practice*. London/New Dehli/New York/Sydney: Bloomsbury.
- Ghivirigă, T. (2020). The gerund challenge. English gerund forms and their Romanian equivalents in the translation of EU documents. *Babel*: 1-13.
- Goatly, A. (1996). Green grammar and grammatical metaphor, or Language and the myth of power, or Metaphors we die by. *Journal of Pragmatics*, 25: 537-560.
- Gramsci, A. (2009). *Quaderni dal carcere: Edizione anastatica dei manoscritti*. Roma: Istituto dell'Enciclopedia Italiana.
- Granath, S. (2007). Size matters – or thus can meaningful structures be revealed in large corpora. In R. Facchinetti (Ed.). *Corpus Linguistics 25 Years on* (pp. 169-185). Amsterdam/New York: Rodopi.
- Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4): 403-437.
- Gries, S. Th. (2010). Useful statistics for corpus linguistics. In A. Sánchez, and M. Almela (Eds.). *A Mosaic of Corpus Linguistics. Selected Approaches* (pp. 269-291). Frankfurt am Main: Peter Lang.

- Gries, S. Th. (2013). 50-something years of work on collocations: What is or should be next... *International Journal of Corpus Linguistics*, 18(1): 137-166.
- Habermas, J. (1971). *Teoria e prassi nella società tecnologica*. Bari: Laterza.
- Halliday, M.A.K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (Eds.). *In Memory of J. R. Firth* (pp. 150-161). London: Longman.
- Halliday, M.A.K. (1975). *Learning how to Mean. Explorations in the Development of Language*. London: Edward Arnold.
- Halliday, M.A.K. (1978). *Language as social semiotic. The social interpretation of language and meaning*. London/Victoria/Baltimore: Edward Arnold.
- Halliday, M.A.K. (1992). New ways of meaning: The challenge to applied linguistics. In M. Pütz (Ed.). *Thirty Years of Linguistic Evolution: Studies in Honour of René Dirven* (pp. 59-95). Amsterdam/Philadelphia: John Benjamins.
- Halliday, M.A.K. (1994). *An Introduction to Functional Grammar*. London: Arnold.
- Halliday, M.A.K. (2004). Lexicology. In M. A. K. Halliday, W. Teubert, C. Yallop, and A. Čermáková. *Lexicology and Corpus Linguistics. An Introduction* (pp. 1-22). London/New York: Continuum.
- Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English*. London/New York: Longman.
- Halliday, M.A.K., and Matthiessen, C. (2004). *An Introduction to Functional Grammar*. London: Hodder Education.
- Halliday, M. A. K., and Webster, J. J. (2014). *Text Linguistics. The How and Why of Meaning*. Sheffield/Bristol: Equinox.
- Hardie, A. (2014a). *Log ratio. An informal introduction*. Post on the website of the ESRC Centre for Corpus Approaches to Social Science CASS.
- Hardie, A. (2014b). Modest XML for Corpora: Not a standard, but a suggestion. *ICAME Journal*, 38: 73-103.
- Hardt-Mautner, G. (1995). "Only Connect". Critical discourse analysis and corpus linguistics. Lancaster University. Available from <http://ucrel.lancs.ac.uk/papers/techpaper/vol6.pdf>.
- Heid, U. (2008). Corpus linguistics and lexicography. In A. Lüdeling, and M. Kytö. *Corpus Linguistics: An International Handbook* (Volume 1; pp. 131-153). Berlin/New York: Mouton de Gruyter.
- Heine, B., and Kuteva, T. (2006). *The Changing Languages of Europe*. Oxford: Oxford University Press.

- Hill, J. (2016). *The British Newspaper Industry. The Future of the Regional Press*. Basingstoke/New York: Palgrave Macmillan.
- Hodge, R., and Kress, G. (1979[1993]). *Language as Ideology*. London/New York: Routledge.
- Hoey, M. (2004). Lexical priming and the properties of text. In A. Partington, J. Morley, and L. Haarman (Eds.). *Corpora and Discourse* (pp. 385-412). Bern: Peter Lang.
- Hoey, M. (2005). *Lexical Priming. A new theory of words and language*. London/New York: Routledge.
- Hoey, M. (2017). Cohesion and coherence in content-specific corpus. In M. Pace-Sigge, and K. J. Patterson (Eds.). (2017). *Lexical Priming. Applications and advances* (pp. 3-40). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hofland, K., and Johansson, S. (1982). *Word Frequencies in British and American English*. Bergen: Norwegian Computing Centre for the Humanities, London/New York: Longman.
- Hoover, D. L. (2009). Word frequency, statistical stylistics and authorship attribution. In D. Archer (Ed.), *What's in a Word List? Investigating word frequency and key word extraction* (pp. 35-51). London: Ashgate.
- Hughes, J. J. (2017). *The processing of collocations by native and non-native speakers of English: Evidence from ERP studies*. Paper presented at 9<sup>th</sup> International Corpus Linguistics Conference. University of Birmingham (United Kingdom), 24-27 July 2017.
- Hughes, J. J. (2018). *The psychological validity of collocation: evidence from event-related brain potentials*. PhD dissertation, Lancaster University.
- Hughes, J.J., and Hardie, A. (2020). Corpus Linguistics and Event-Related Potentials. In J. Egbert and P. Baker (Eds.). *Using Corpus Methods to Triangulate Linguistic Analysis* (pp. 185-218). New York/London: Routledge.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2004). Counting the uncountable: Problems of identifying evaluation in a text and in a corpus. In A. Partington, J. Morley, and L. Haarman (Eds.). *Corpora and Discourse* (pp. 157-188). Bern: Peter Lang.
- Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12(2), 249-268.
- Hunston, S. (2010). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Routledge.

- Hunston, S., and Thompson, G. (2000). Evaluation: An Introduction. In S. Hunston and G. Thompson (Eds.). *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 1-27). Oxford: Oxford University Press.
- Ide, N., Calzolari, N., Eckle-Kohler, J., Gibbon, D., Hellmann, S., Lee, K., Nivre, J., and Romary, L. (2017). Community Standards for Linguistically-Annotated Resources. In N. Ide, and J. Pustejovsky (Eds.). *Handbook of Linguistic Annotation* (pp. 113-165). Dordrecht: Springer.
- International Union for Conservation of Nature and natural resources (IUNC). (1980). *World conservation strategy: Living resources conservation for sustainable development*. Gland: IUNC.
- Jakubíček, M., Kilgarrif, A., Kovář, V., Rychlý, P., and Suchomel, V. (2013). *The TenTen Corpus Family*. Paper presented at 7<sup>th</sup> International Corpus Linguistics Conference. Lancaster University (United Kingdom), 23-26 July 2013.
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(1): 87-106.
- Jaworska, S., and Kinloch, K. (2018). Using multiple data sets. In C. Taylor, and A. Marchi (Eds.). (2018). *Corpus Approaches to Discourse: A Critical Review* (pp. 110-129). London/New York: Routledge.
- Jeffries, L., and Walker, B. (2017). *Keywords in the Press. The New Labour Years*. London/New York: Bloomsbury Academic.
- Johansson, S. (1998). On the role of corpora in cross-linguistic research. In S. Johansson, and S. Oksefjell (Eds.). *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies* (pp. 3-24). Amsterdam/Atlanta: Rodopi.
- Johansson, S. (2007). *Seeing through Multilingual Corpora. On the use of corpora in contrastive studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Johansson, S., Leech, G. N., and Goodluck, H. (1978). *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Retrieved from: <http://www.hd.uib.no/icame/lob/lob-dir.htm> .
- Johansson, S., and Oksefjell, S. (Eds.). (1998). *Corpora and Cross-Linguistic Research. Theory, Method, and Case Studies*. Amsterdam/Atlanta: Rodopi.
- Jones, S., and Sinclair, J. (1974). English lexical collocations. *Cahiers de Lexicologie*, 24, 15-61.
- Kanie, N., Bernstein, S., Biermann, F., and Haas, P.M. (2017). Introduction: Global Governance through Goal Setting. In N. Kanie and F. Biermann (Eds.).

- Governing through Goals. Sustainable Development Goals as Governance Innovation (pp. 1-27). Cambridge, Massachusetts/London: The MIT Press.
- Kenesei, I., Vago, R. M., and Fenyvesi, A. (1998). *Hungarian Descriptive Grammars*. Abingdon/New York: Routledge.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London/New York: Longman.
- Kiefer, F. (1967). *On emphasis and word order in Hungarian*. Bloomington: Indiana University.
- Kilgarriff, A. (2009). Simple maths for keywords. In M. Mahlberg, V. González-Díaz, and C. Smith (Eds.). *Proceedings of the Corpus Linguistics Conference, CL2009*. Liverpool: University of Liverpool. Available from <https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf>.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., e Suchomel, V. (2014). The Sketch Engine: 10 years on. *Lexicography*, 1: 7-36.
- Knight, D., Evans, D., Carter, R, and Adolphs, S. (2009). HeadTalk, HandTalk and the corpus: Towards a framework for multi-modal, multi-media corpus development. *Corpora*, 4(1): 1-32.
- Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. *MT summit*, 5: 79-86.
- Koller, V., and Mautner, G. (2004). Computer Applications in Critical Discourse Analysis. In C. Coffin, A. Hewings, and K. O'Halloran (Eds.). *Applying English Grammar: Functional and Corpus Approaches* (pp. 216-228). London: Hodder Education.
- Koltay-Kastner, J., and Juhász, Zs. (2000). *Magyar-olasz szótár*. Budapest: Akadémiai Kiadó.
- Kosem, I. (2015). Interrogating a Corpus. In P. Durkin (Ed.). *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press.
- Kovács, A., and Wodak, R. (Eds.). (2003). *NATO, Neutrality and National Identity: The Case of Austria and Hungary*. Wien/Köln/Weimar: Böhlau Verlag.
- König, E. (1994). English. In E. König and J. van der Auwera (Eds.). *The Germanic Languages* (pp. 532-565). London/New York: Routledge.
- Kress, G. (Ed.). (1976). *Halliday: System and Function in Language*. Oxford: Oxford University Press.
- Kress, G., and Leeuwen, T. van (1996). *Reading Images: A grammar of visual design*. London/New York: Routledge.

- Kress, G., and Leeuwen, T. van (2001). *Multimodal Discourse: The modes and media of contemporary communication*. London: Arnold.
- Krizsán, A. (2011). *"The EU is Not Them, But Us!": The First Person Plural and the Articulation of Collective Identities in European Political Discourse*. Newcastle upon Tyne: Cambridge Scholars.
- Kübler, S., and Zinsmeister, H. (2015). *Corpus Linguistics and Linguistically Annotated Corpora*. London/New York: Bloomsbury Academic.
- Lakoff, G, and Wehling, E. (2012). *The Little Blue Book: The Essential Guide to Thinking and Talking Democratic*. New York: Free Press.
- Langford, M., Sumner, A., and Yamin, A.E. (Eds.). (2013). *The Millennium Development Goals and Human Rights: Past, Present and Future*. Cambridge: Cambridge University Press.
- Laviosa, S. (2002). *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam: Rodopi.
- Lazar, M.M. (Ed.). (2005). *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*. Houndmills/New York: Palgrave Macmillan.
- Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer, and B. Altenberg (Eds.). *English Corpus Linguistics* (pp. 8-29). Abingdon/New York: Routledge.
- Leech, G. (1992). Corpora and Theories of Linguistic Performance. In J. Svartvik (Ed.). *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991* (pp. 105-122). Berlin/New York: Mouton de Gruyter.
- Leech, G. (1997). Introducing Corpus Annotation. In R. Garside, G. Leech, and T. McEnery (Eds.). *Corpus Annotation: Linguistic Information from Computer Text Corpora* (pp. 1-18). London/New York: Longman.
- Leech, G., and Short, M. (1981). *Style in fiction. A linguistic introduction to English fictional prose*. London/New York: Longman.
- Leeuwen, T. van (2006). Critical Discourse Analysis. In K. Brown (Ed.). *Encyclopedia of Language and Linguistics [2<sup>nd</sup> edition]* (vol. III; pp. 290-294). Oxford: Elsevier.
- Leeuwen, T. van (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford: Oxford University Press.
- Lehmberg, T., and Wörner, K. (2009). Annotation standards. In A. Lüdeling, and M. Kytö (Eds.). *Corpus linguistics: an international handbook* (Volume 1; pp. 484-501). Berlin: Walter de Gruyter.
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, 20(1): 1-31.

- Lenci, A., Pirrelli, V., and Montemagni, S. (2005). *Testo e computer: Elementi di linguistica computazionale*. Roma: Carocci.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago/London: University of Chicago.
- Lo Cascio, V. (1997). Semantica lessicale e i criteri di collocazione nei dizionari bilingui a stampa ed elettronici. In T. De Mauro and V. Lo Cascio (Eds.). *Lessico e Grammatica: Teorie linguistiche e applicazioni lessicografiche. Atti del Convegno interannuale della Società di Linguistica Italiana, Madrid, 21-25 febbraio 1995* (pp. 63-88). Roma: Bulzoni.
- Louw, B. (1993). Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.). *Text and technology: In honour of John Sinclair* (pp. 157-176). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Machin, D. (2016). *Introduction to Multimodal Analysis*. London/Oxford/New York/New Delhi/Sydney: Bloomsbury.
- Machin, D., and Leeuwen, T. van (2007). *Global Media Discourse: A critical introduction*. London/New York: Routledge.
- Macgilchrist, F. (2007). Positive Discourse Analysis: Contesting Dominant Discourses by Reframing the Issues. *Critical Approaches to Discourse Analysis Across Disciplines*, 1(1): 74-94.
- Mahlberg, M. (2007). Lexical items in discourse: identifying local textual functions of *sustainable development*. In M. Hoey, M. Mahlberg, M. Stubbs, and W. Teubert (Eds.). *Text, Discourse and Corpora. Theory and Analysis* (pp. 191-218). London/New York: Continuum.
- Mahlberg, M., and McIntyre, D. (2011). A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction*, 4(2): 204-227.
- Maiden, M., and Robustelli, C. (2000). *A Reference Grammar of Italian*. Chicago: NTC Publishing.
- Manzelli, G. (1990). Il plurale cappatico in armeno e ungherese. In M.E. Conte, A. Giacalone Ramat and P. Ramat (Eds.). *Dimensioni della linguistica* (pp. 61-116). Milano: Franco Angeli.
- Manzelli, G. (1993). Le lingue uraliche (ugrofinniche e samoiede). In E. Banfi (Ed.). *La formazione dell'Europa linguistica: Le lingue d'Europa tra la fine del I e del II millennio* (pp. 491-551). Firenze: La Nuova Italia Editrice.
- Marchi, A. (2010). 'The moral in the story': a diachronic investigation of lexicalised morality in the UK press. *Corpora*, 5(2): 161-189.

- Marchi, A. (2018). Dividing up the data: Epistemological, methodological and practical impact of diachronic segmentation. In C. Taylor, and A. Marchi (Eds.). *Corpus Approaches to Discourse: A Critical Review* (pp. 174-196). Abingdon/New York: Routledge.
- Marchi, A., and Taylor, C. (2009). If on a Winter's Night Two Researchers... A Challenge to Assumptions of Soundness of Interpretation. *Critical Approaches to Discourse Analysis across Disciplines*, 3(1): 1-20.
- Marchi, A., and Taylor, C. (2018). Introduction. Partiality and reflexivity. In C. Taylor and A. Marchi. *Corpus Approaches to Discourse. A Critical Review* (pp. 1-15). Oxon/New York: Routledge.
- Martin, J. (2004) Positive discourse analysis: Solidarity and change. *Revista Canaria de Estudios Ingleses*, 49: 179-202.
- Martin, J., and Rose, D. (2003). *Working with Discourse: Meaning beyond the clause*. London/New York: Continuum.
- Martin, J., and White, P. (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan.
- Mautner, G. (2000). British National Identity in the European Context. In A. Musolff, C. Good, P. Points, and R. Wittlinger (Eds.). *Attitudes Towards Europe: Language in the unification process* (pp. 3-22). Aldershot/Burlington: Ashgate.
- Mautner, G. (2007). Mining large corpora for social information: The case of *elderly*. *Language in Society*, 36, 51-72.
- Mautner, G. (2009). Corpora and Critical Discourse Analysis. In P. Baker (Ed.). *Contemporary Corpus Linguistics* (pp. 32-46). London/New York: Continuum.
- Mautner, G. (2016). Checks and balances: How corpus linguistics can contribute to CDA. In R. Wodak and M. Meyer (Eds.). *Methods of Critical Discourse Studies [3<sup>rd</sup> Edition]* (pp. 154-179). London/Thousand Oaks/New Delhi/Singapore: SAGE Publications.
- McCarthy, M., and Carter, R. (2001). Size Isn't Everything: Spoken English, Corpus, and the Classroom. *TESOL Quarterly*, 35(2): 337-340.
- McCowan, T. (2019). *Higher Education for and beyond the Sustainable Development Goals*. Cham: Palgrave Macmillan.
- McEnery, T. (2006). *Swearing in English. Bad Language, Purity and Power from 1586 to the Present*. Abingdon/New York: Routledge.
- McEnery, T., and Wilson, A. (2001). *Corpus Linguistics. An Introduction* (Second Edition). Edinburgh: Edinburgh University Press.

- McEnery, T., Xiao, R., and Tono, Y. (2006). *Corpus-Based Language Studies: An advanced resource book*. Abingdon/New York: Routledge.
- McEnery, T., and Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- McEnery, T., and Baker, H. (2017). *Corpus Linguistics and 17th-Century Prostitution. Computational Linguistics and History*. London/New York: Bloomsbury Academic.
- McEnery, T., Brezina, V., and Baker, H. (2019). Usage Fluctuation Analysis: A new way of analysing shifts in historical discourse. *International Journal of Corpus Linguistics*, 24(4): 413-444.
- McGillivray, M. (Ed.). (2008). *Achieving the Millennium Development Goals*. Houndsmills/New York: Palgrave Macmillan.
- Meadowcroft, J. (2005). Sustainable Development: A New(ish) Idea for a New Century?. In J.S. Dryzek and D. Schlosberg (Eds.). *Debating the Earth: The Environmental Politics Reader* (pp. 267-284). Oxford: Oxford University Press.
- Meadows, D.H., Meadows, D.L., Randers, J., and Behrens III, W.W. (1972). *The Limits to Growth. A Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books.
- Meyer, M. (2001). Between theory, method, and politics: Positioning of the approaches to CDA. In R. Wodak, and M. Meyer (Eds.). *Methods of Critical Discourse Analysis* (pp. 14-31). London/Thousand Oaks/New Delhi: Sage.
- Meyer, C.F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Mikhailov, M., and Cooper, R. (2016). *Corpus Linguistics for Translation and Contrastive Studies. A guide for research*. London/New York: Routledge.
- Milizia, D. (2010). Keywords and phrases in political speeches. In M. Bondi, and M. Scott (Eds). *Keyness in Texts* (pp. 127-145). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Moon, R. (2010). What can a corpus tell us about lexis? In A. O' Keeffe, and M. McCarthy (Eds). *The Routledge Handbook of Corpus Linguistics* (pp. 197-211). Abingdon/New York: Routledge.
- Morley, J., and Partington, A. (2009). A few *Frequently Asked Questions* about semantic – or *evaluative* – prosody. *International Journal of Corpus Linguistics*, 14: 139-158.
- Mortara Garavelli, B. (1979). *Il filo del discorso*. Torino: Giappichelli.
- Mühlhäusler, P. (2003). *Environment of Language. A Course in Ecolinguistics*. London: Battlebridge Publications.

- Müller, C. (2008). *Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View*. Chicago, IL: University of Chicago Press.
- Naeem, S., Chazdon, R., Duffy, J.E., Prager, C., and Worm, B. (2016). Biodiversity and human well-being: an essential link for sustainable development. *Proceedings: Biological Sciences*, 283(1844): 1-10.
- Naess, A. (1995). The shallow and the long range, deep ecology movement. In A. Drengson and Y. Inoue (Eds.). *The deep ecology movement: An introductory anthology* (pp. 3-10). Berkeley: North Atlantic Books.
- Naess, A. (2005). The Shallow and the Deep, Long-Range Ecology Movement: A Summary. In J.S. Dryzek and D. Schlosberg (Eds.). *Debating the Earth: The Environmental Politics Reader* (pp. 343-347). Oxford: Oxford University Press.
- Nelson, M. (2010). Building a written corpus: what are the basics? In A. O' Keeffe, and M. McCarthy (Eds.). *The Routledge Handbook of Corpus Linguistics* (pp. 53-65). Abingdon/New York: Routledge.
- Nightingale, V. (Ed.). (2011). *The Handbook of Media Audiences*. Malden/Oxford/Chichester: Wiley-Blackwell.
- Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- O'Donnell, M. (2008). The UAM CorpusTool: Software for corpus annotation and exploration. In C.M. Bretones Callejas, J.F.F. Sánchez, J.R. Ibáñez Ibáñez, M.E.G. Sánchez, E. Cortés de los Ríos, S.S. Ramiro, S.C. Martínez, N.P. Honeyman and B.C. Márquez (Eds.). *Applied Linguistics Now: Understanding Language and Mind/La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente* (pp. 1433-1447). Almería: Universidad de Almería.
- O'Halloran, K. (2010). How to use corpus linguistics in the study of media discourse. In A. O' Keeffe, and M. McCarthy (Eds.). *The Routledge Handbook of Corpus Linguistics* (pp. 536-577). Abingdon/New York: Routledge.
- Orpin, D. (2005). Corpus Linguistics and Critical Discourse Analysis. Examining the ideology of sleaze. *International Journal of Corpus Linguistics*, 10(1): 37-61.
- Pace-Sigge, M. (2018). *Spreading Activation, Lexical Priming and the Semantic Web. Early Psycholinguistic Theories, Corpus Linguistics and AI Applications*. Cham: Palgrave Macmillan.
- Pace-Sigge, M., and Patterson, K. J. (Eds.). (2017). *Lexical Priming. Applications and advances*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Partington, A. (1998). *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Partington, A. (2004). Corpora and discourse, a most congruous beast. In A. Partington, J. Morley and L. Haarman (Eds.). *Corpora and Discourse* (pp. 11-20). Bern: Peter Lang.
- Partington, A. (2007). Irony and reversal of evaluation. *Journal of Pragmatics*, 39(9): 1547-1569.
- Partington, A. (2009). Evaluating evaluation and some concluding thoughts on CADS. In J. Morley, and P. Bayley (Eds.). *Corpus-Assisted Discourse Studies on the Iraq Conflict. Wording the War* (pp. 261-303). Abingdon/New York: Routledge.
- Partington, A. (2014). Mind the gaps. The role of corpus linguistics in researching absences. *International Journal of Corpus Linguistics*, 19(1): 118-146.
- Partington, A. (2015) Evaluative prosody. In K. Aijmer, and C. Rühlemann (Eds.). *Corpus pragmatics: a handbook* (pp. 279-303). Cambridge: Cambridge University Press.
- Partington, A. (2017). Evaluative clash, evaluative cohesion and how we actually read evaluation in texts. *Journal of Pragmatics*, 117: 190-203.
- Partington, A. (Ed.). (2010). Modern Diachronic Corpus-assisted Studies (MD-CADS). *Corpora*, 5(2), special issue.
- Partington, A., Duguid, A., and Taylor, C. (2013). *Patterns and Meanings in Discourse. Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Peters, C., Picchi, E., and Biagini, L. (2000). *Parallel and comparable bilingual corpora in language teaching and learning*. In S.P. Botley, A. M. McEnery, and A. Wilson (Eds.). *Multilingual Corpora in Teaching and Research* (pp. 73-85). Amsterdam/Atlanta: Rodopi.
- Peterson del Mar, D. (2012). *Environmentalism*. Harlow: Pearson Education Limited.
- Phillips, M. (1989). *Lexical Structure of Text*. Birmingham: University of Birmingham.
- Phillipson, R. (1992). *Linguistic Imperialism*. Oxford: Oxford University Press.
- Pojanapunya, P., and Watson Todd, R. (2018). Log-likelihood and odds ratio: Keynes statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1): 133-167.
- Poole, R. (2016a). A corpus-aided approach for the teaching and learning of rhetoric in an undergraduate composition course for L2 writers. *Journal of English for Academic Purposes*, 21: 99-109.

- Poole, R. (2016b). A corpus-aided ecological discourse analysis of the Rosemont Copper Mine debate of Arizona, USA. *Discourse & Communication*, 10(6): 576-595.
- Poole, R. (2019). *Expanding the Scope of Corpus-Aided Ecolinguistics*. Paper presented at 4<sup>th</sup> International Conference on Ecolinguistics. University of Southern Denmark (Denmark), 12-15 August 2019.
- Potter, J. (1996). *Representing Reality: Discourse, Rhetoric and Social Construction*. London: Sage.
- Potts, A. (2018). *Making a patchwork quilt or Frankenstein's monster? Interdisciplinarity and collaboration in CADs*. Paper presented at 4<sup>th</sup> International Corpora and Discourse Conference. Lancaster University (United Kingdom), 22-24 June 2018.
- Potts, A., Bednarek, M., and Caple, H. (2015). How can computer-based methods help researchers to investigate news values in large datasets? A corpus linguistic study of the construction of newsworthiness in the reporting on Hurricane Katrina. *Discourse & Communication*, 9(2): 149-172.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A comprehensive grammar of the English language (Indexed by David Crystal)*. London/New York: Longman.
- Ragazzini, G. (2019). *Il Ragazzini: Dizionario inglese-italiano, italiano-inglese*. Bologna: Zanichelli.
- Rayson, P., Berridge, D., and Francis, B. (2004). *Extending the Cochran rule for the comparison of word frequencies between corpora*. Paper presented at 7<sup>th</sup> International Conference on Statistical Analysis of Textual Data, JADT 2004 (pp. 926-36). Louvain-la-Neuve.
- Rissanen, M. (2008). Corpus linguistics and historical linguistics. In A. Lüdeling, and M. Kytö (Eds.). *Corpus linguistics: an international handbook* (Volume 1; pp. 53-68). Berlin: Walter de Gruyter.
- Rissanen, M., Kytö, M., and Palander, M. (Eds.). (1993). *Early English in the computer age. Explorations through the Helsinki corpus*. Berlin: Mouton de Gruyter.
- Ross, K., and Nightingale, V. (2003). *Media and Audiences: New Perspectives*. Maidenhead: Open University Press.
- Rossini Favretti, R. (1998). *La linguistica applicata: Aspetti Problemi Percorsi*. Bologna: Pàtron Editore.
- Rossini Favretti, R. (2002). *Un'introduzione alla linguistica applicata*. Bologna: Pàtron Editore.

- Rossini Favretti, R., Tamburini, F., and Martelli, E. (2007). Words from Bononia Legal Corpus. In W. Teubert (Ed.). *Text Corpora and Multilingual Lexicography* (pp. 11-30). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Rychlý, P. (2008). A Lexicographer-Friendly Association Score. In P. Sojka, and A. Horák (Eds.). *Proceedings of Recent Advances in Slavonic Natural Language Processing. RASLAN* (pp. 6-9). Brno: Masaryk University.
- Saeed, J. I. (2015). *Semantics* (Fourth Edition). Malden/Oxford/Chichester: John Wiley & Sons.
- Salvi, R., and Turnbull, J. (2010). Appraisal theory as a methodological proposal for stylistic analysis. *Textus*, 23(1): 103-138.
- Schröter, M., and Taylor, C. (Eds.). (2018). *Exploring Silence and Absence in Discourse. Empirical Approaches*. London: Palgrave Macmillan.
- Scott, M. (1997). PC Analysis of Key Words – and Key Key Words. *System*, 25(2): 233-245.
- Scott, M. (1999). *WordSmith Tools Manual: Version 3.0*. Oxford: Oxford University Press.
- Scott, M. (2002). Picturing the Key Words of a very Large Corpus and their Lexical Upshots or Getting at the *Guardian's* View of the World. In B. Kettemann, and G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000* (pp. 43-50). Amsterdam/New York: Rodopi.
- Scott, M. (2009). In Search of a Bad Reference Corpus. In D. Archer (Ed.). *What's in a Word List? Investigating word frequency and key word extraction* (pp. 79-92). London: Ashgate.
- Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails... In M. Bondi, and M. Scott (Eds). *Keyness in Texts* (pp. 43-57). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Scott, M. (2015). *WordSmith Tools Manual: Version 6.0*. Available from <https://lexically.net/downloads/version6/wordsmith6.pdf>.
- Scott, M. (2016). *WordSmith Tools version 7* [Computer Software]. Stroud: Lexical Analysis Software. Retrieved from: [http://lexically.net/downloads/version7/HTML/index.html?getting\\_started.htm](http://lexically.net/downloads/version7/HTML/index.html?getting_started.htm).
- Scott, M., and Tribble, C. (2006). *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Sealey, A. (2018). Animals, animacy and anthropocentrism. *International Journal of Language and Culture*, 5(2): 224-247.
- Semino, E., and Short, M. (2014). *Corpus Stylistics. Speech, writing and thought presentation in a corpus of English writing*. Abingdon/New York: Routledge.
- Sharoff, S. (2006). Creating General-Purpose Corpora Using Automated Search Engine Queries. In M. Baroni, and S. Bernardini (Eds). *WaCky! Working papers on the Web as Corpus* (pp. 63-98). Bologna: GEDIT.
- Silverman, D. (Ed.). (2004). *Qualitative Research: Theory, Method and Practice [2<sup>nd</sup> Edition]*. London/Thousand Oaks/New Delhi: SAGE Publications.
- Silverman, D. (2014). *Interpreting Qualitative Data [5<sup>th</sup> Edition]*. Los Angeles/London/Washington DC/New Delhi/Singapore: SAGE Publications.
- Silverman, D., and Marvasti, A. (2008). *Doing Qualitative Research: A Comprehensive Guide*. Los Angeles/London/New Delhi/Singapore: SAGE Publications.
- Sinclair, J. (Ed.). (1987). *Collins Cobuild English language dictionary*. London/Glasgow: Collins.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1995). Paper presented at *XI Encontro da Associação Portuguesa de Linguística*. Lisbon, 2-4 October 1995.
- Sinclair, J. (1996). The search for units of meaning. *Textus*, IX: 75-106.
- Sinclair, J. (1998). The Lexical Item. In E. Weigand (Ed.). *Contrastive Lexical Semantics* (pp. 1-24). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Sinclair, J. (2004). *Trust the Text. Language, corpus and discourse*. London/New York: Routledge.
- Sinclair, J., Jones, S., and Daley, R. (2004). *English Collocation Studies. The OSTI Report*. London/New York: Continuum.
- Sinclair, J., and Mauranen, A. (2006). *Linear Unit Grammar. Integrating speech and writing*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Smadja, F. (1993). Retrieving Collocations from Text. *Xtract. Computational Linguistics*, 19(1): 143-177.
- Smadja, F., McKeown, K. R., Hatzivassiloglou, V. (1996). Translating Collocations for Bilingual Lexicons. A Statistical Approach. *Computational Linguistics*, 22(1): 1-38.
- Sperberg-McQueen, C. M., and Burnard, L. (2019). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Available from <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> .
- Spina, S. (2001). *Fare i conti con le parole: Introduzione alla linguistica dei corpora*. Perugia: Guerra Edizioni.

- Spina, S. (2010). *The Dictionary of Italian Collocations: Design and Integration in an Online Learning Environment*. Available from [http://www.lrec-conf.org/proceedings/lrec2010/pdf/681\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/681_Paper.pdf).
- Spina, S. (2014). Come cambia la lingua dei politici nell'era di *Twitter*. In E. Garavelli and E. Suomela-Härmä (Eds.). *Dal manoscritto al Web: Canali e modalità di trasmissione dell'italiano. Tecniche, materiali e usi nella storia della lingua. Atti del XII Congresso SILFI, Società Internazionale di Linguistica e Filologia Italiana (Helsinki, 18-20 giugno 2012)* (Volume II; pp. 707-715). Firenze: Franco Cesati Editore.
- Steffensen, S.V., and Fill, A. (2014). Ecolinguistics: the state of art and future horizons. *Language Sciences*, 41: 6-25.
- Steward, D. (2010). *Semantic Prosody. A Critical Evaluation*. London: Routledge.
- Stibbe, A. (2012). *Animals Erased: Discourse, Ecology, and Reconnection with the Natural World*. Middletown: Wesleyan University Press.
- Stibbe, A. (2015). *Ecolinguistics. Language, Ecology and the Stories We Live By*. London/New York: Routledge.
- Stibbe, A. (2017). Critical discourse analysis and ecology. In J. Flowerdew and J.E. Richardson (Eds.). *The Routledge Handbook of Critical Discourse Studies* (pp. 497-509). London/New York: Routledge.
- Stibbe, A. (2018). Positive Discourse Analysis: Re-thinking human ecological relationships. In A. Fill and H. Penz (Eds.). *The Routledge Handbook of Ecolinguistics* (pp. 165-178). London: Routledge.
- Stokols, D., Perez Lejano, R., and Hipp, J. (2013). Enhancing the Resilience of Human-Environment Systems: A Social-Ecological Perspective. *Ecology and Society*, 18(1): 1-12.
- Stubbs, M. (1995a). Collocations and semantic profiles. On the cause of the trouble with quantitative methods. *Functions of Language*, 2(1): 1-33.
- Stubbs, M. (1995b). Corpus evidence for norms of lexical collocation. In G. Cook, and B. Seidlhofer (Eds.). *Principle & Practice in Applied Linguistics: Studies in honour of H. G. Widdowson* (pp. 245-256). Oxford: Oxford University Press.
- Stubbs, M. (1996). *Text and Corpus Analysis: Corpus-assisted Studies of Language and Culture*. Oxford/Malden: Blackwell Publishers.
- Stubbs, M. (1997). Whorf's children: Critical comments on Critical Discourse Analysis (CDA). In A. Ryan, and A. Wray (Eds.). *Evolving Models of Language* (pp. 100-116). Clevedon: Multilingual Matters.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford/Malden: Blackwell Publishers.

- Stubbs, M. (2010). Three concepts of keywords. In M. Bondi, and M. Scott (Eds). *Keyness is Texts* (pp. 21-42). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Stubbs, M. (2015). Corpus semantics. In N. Riemer (Ed.). *The Routledge Handbook of Semantics* (pp. 106-121). London/New York: Routledge.
- Taylor, C. (2012). *And there it isn't. (How) can we access the absent using CADS?*. Paper presented at CADS International Conference. University of Bologna (Italy), 13-15 September 2012.
- Taylor, C. (2013a). *What is absent in UK press discourses of migration?*. Paper presented at International Language in the Media Conference. Queen Mary University of London (United Kingdom), 23-30 September 2013.
- Taylor, C. (2014). Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics*, 19(3): 368-400.
- Taylor, C. (2016). *Mock Politeness in English and Italian. A corpus-assisted metalanguage analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Taylor, C., and Marchi, A. (2018a). *Corpus Approaches to Discourse. A Critical Review*. Oxon/New York: Routledge.
- Taylor, C. and Marchi, A. (2018b). *Blind spots and dusty corners. (Self)-reflections on partiality in corpus & discourse studies*. Paper presented at 4<sup>th</sup> Corpora and Discourse International Conference. Lancaster University (United Kingdom), 22-24 June 2018.
- Taylor, C., and Del Fante, D. (2020). Comparing across languages in corpus and discourse analysis: some issues and approaches. *Meta: Journal des traducteurs / Meta: Translators' Journal*, 65(1): 29-50.
- Teubert, W. (2000). A Province of a Federal Superstate, Ruled by an Unelected Bureaucracy – Keywords of the Euro-sceptic Discourse in Britain. In A. Musolff, C. Good, P. Points, and R. Wittlinger (Eds.). *Attitudes Towards Europe: Language in the unification process* (pp. 45-88). Aldershot/Burlington: Ashgate.
- Teubert, W. (2002). The role of parallel corpora in translation and multilingual lexicography. In B. Altenberg, and S. Granger (Eds.). *Lexis in Contrast: Corpus-based approaches* (pp. 189-214). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Teubert, W. (2007). Corpus linguistics and lexicography. In W. Teubert (Ed.). *Text Corpora and Multilingual Lexicography* (pp. 109-133). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Thompson, G. (1996). *Introducing Functional Grammar*. London: Arnold.
- Thompson, G., and Hunston, S. (2000). Evaluation: An Introduction. In S. Hunston, and G. Thompson (Eds.). *Evaluation in Text. Authorial Stance and the Construction of Discourse* (pp. 1-27). Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tognini-Bonelli, E. (2002). Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. In B. Altenberg and S. Granger (Eds.). *Lexis in Contrast. Corpus-based approaches* (pp. 73-95). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Trampuš, M., and Novak, B. (2012). *The Internals of An Aggregated Web News Feed*. Proceedings of 15<sup>th</sup> Multiconference on Information Society 2012.
- United Nations. (1973). *Report of the United Nations conference on the human environment: Stockholm, 5-16 June 1972*. New York: United Nations.
- United Nations. (2015a). *Transforming our world: The 2030 Agenda for Sustainable Development*. Available from <https://sustainabledevelopment.un.org/post2015/transformingourworld> .
- United Nations. (2015b). *Világunk átalakítása: Fenntartható Fejlődési Keretrendszer 2030*. Available from <http://ensz.kormany.hu/download/7/06/22000/Vil%C3%A1gunk%20%C3%A1talak%C3%ADt%C3%A1sa%20Fenntarthat%C3%B3%20Fejl%C5%91d%C3%A9si%20Keretrendszer%202030.pdf> .
- United Nations. (2015c). *Trasformare il nostro mondo: l'Agenda 2030 per lo Sviluppo Sostenibile*. Available from [https://www.unric.org/it/images/Agenda\\_2030\\_ITA.pdf](https://www.unric.org/it/images/Agenda_2030_ITA.pdf) .
- van Dusseldorp, D., e Wigboldus, S. (1994). Interdisciplinary Research for Integrated Rural Development in Developing Countries: The Role of Social Sciences. *Issues in Integrative Studies*, 12: 93-138.
- Verschuere, J. (2001). Predicaments of criticism. *Critique of Anthropology*, 21(1): 59-81.
- Vessey, R. (2013). Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora*, 8(1): 1-26.
- Vessey, R. (2016). *Language and Canadian Media. Representations, Ideologies, Policies*. London: Palgrave Macmillan.
- Vincent, N. (1988). Italian. In M. Harris and N. Vincent (Eds.). *The Romance Languages* (pp. 279-313). London/Sydney: Croom Helm.

- Walter, E. (2010). Using a corpus to write dictionaries. In A. O' Keffe, and M. McCarthy (Eds). *The Routledge Handbook of Corpus Linguistics* (pp. 428-443). Abingdon/New York: Routledge.
- Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics*, 10: 283-305.
- Widdowson, H.G. (1998). The Theory and Practice of Critical Discourse Analysis. *Applied Linguistics*, 19(1): 136-151.
- Widdowson, H.G. (2004). *Text, Context, Pretext: Critical Issues in Discourse Analysis*. Malden/Oxford/Victoria: Blackwell Publishing.
- Wierzbicka, A. (1996). *Semantics. Primes and Universals*. Oxford/New York: Oxford University Press.
- Wierzbicka, A. (1997). *Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese*. Oxford/New York: Oxford University Press.
- Williams, G. (1998). Collocation networks. Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics*, 3(1): 151-171.
- Williams, R. (1958). *Culture and Society: 1780-1950*. New York: Anchor Books.
- Williams, R. (1961). *The Long Revolution*. Harmondsworth Middlesex/Victoria: Penguin Books.
- Williams, R. (1983). *Keywords. A vocabulary of culture and society*. London: Fontana Paperbacks.
- Wilson, A. (2013). Embracing Bayes factors for key item analysis in corpus linguistics. In M. Bieswanger, and A. Koll-Stobbe (Eds.). *New Approaches to the Study of Linguistic Variability. Language Competence and Language Awareness in Europe* (Vol. 4; pp. 3-11). Frankfurt am Main: Peter Lang.
- Wodak, R. (Ed.). (1989). *Language, Power and Ideology: Studies in Political Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Wodak, R. (2001). What CDA is about: A summary of its history, important concepts and its developments. In R. Wodak and M. Meyer (Eds.). *Methods of Critical Discourse Analysis [1<sup>st</sup> Edition]* (pp. 1-13). London/Thousand Oaks/New Delhi: SAGE Publications.
- Wodak, R., and Krzyżanowski, M. (Eds.). (2008). *Qualitative Discourse Analysis in the Social Sciences*. Houndmills/New York: Palgrave Macmillan.
- Wodak, R., and Meyer, M. (2016). Critical discourse studies: history, agenda, theory and methodology. In R. Wodak and M. Meyer (Eds.). *Methods of Critical*

- Discourse Studies [3<sup>rd</sup> Edition]* (pp. 1-22). London/Thousand Oaks/New Delhi/Singapore: SAGE Publications.
- World Commission on Environment and Development (WCED). (1987). *Our common future*. Oxford/New York: Oxford University Press.
- Zipf, G. K. (1935). *The Psycho-Biology of Language*. Boston: Houghton Mifflin.
- Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. London/New York: Hafner.
- Zottola, A. (2019). (Trans)gender in the news: specialized language in the UK press. A corpus-based discourse analysis. *Lingue e Linguaggi*, 29: 461-480.
- Zufferey, S., and Cartoni, B. (2012). English and French causal connectives in contrast. *Language in Contrast*, 12(2): 232-250.

## Sitography

- A description of Sketch Engine's TenTen Corpora can be found at <https://www.sketchengine.eu/documentation/tenten-corpora/> (last accessed on 15<sup>th</sup> May 2021).
- Sketch Engine's presentation of word sketches can be found at <https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/#toggle-id-9> (last accessed on 15<sup>th</sup> May 2021).
- A description of the function and structure of the United Nations Economic and Social Council can be found at <https://www.un.org/ecosoc/en/about-us> (last accessed on 15<sup>th</sup> May 2021).



## Appendix

### A1. Outline of the statistics for the computation of keywords and collocations

*Statistical keywords.* The computation of statistical keywords can be carried out with statistical significance tests and with effect size tests. Among the statistical significance tests, chi-square and Log Likelihood are the most common. As a statistical test, the chi-square helps uncovering similarities and differences between two data set by comparing “frequencies found experimentally with those expected on the basis of some theoretical model” (Oakes 1998). In fact, the frequency differences of the items under inquiry (*dependent variables*) may be caused by the intervention of *independent variables* such as age or gender of a speaker or writer, genre or topic of the texts included in the corpora, etc. The application of this statistical test starts with the call of a *null hypothesis*, according to which differences in frequency are merely due to chance. The hypothesis needs thus to be tested in order to be fully accepted or refuted. It can be substituted by an *alternative hypothesis*, which asserts that differences are due to a correlation between independent and dependent variables.

The formula of the chi-square statistical test is:

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Where  $X^2$  stands for chi-square,  $O$  stands for observed frequency and  $E$  for expected frequency, namely the frequency the item would display if words were evenly distributed in a text or corpus (Oakes 1998: 25).

When doing keyword analysis, the chi-square statistical score is calculated for every type in the corpus within the frame of a *contingency table*, “a table in which the outcomes of an experiment are classified according to two criteria” (Oakes 1998: 25).

These two criteria convert into *degrees of freedom*. Degrees of freedom are the number of possible independent observations that can be made on the data and they are calculated with the formula

$$df = (n_{rows} - 1) \times (n_{columns} - 1)$$

where  $df$  stands for degrees of freedom,  $n_{rows}$  for the number of rows and  $n_{columns}$  for the number of columns of the contingency table (Freddi 2014: 95). Every degree of freedom has a specific critical value for chi-square which is also tied to the accuracy level of the test (Freddi 2014). The critical value for one degree of freedom is, for example, 3.84 with 95% accuracy (i.e. only a 5% chance of being mistaken), 6.63 with a 99% level of accuracy, 10.83 and 15.13 with 99.9% and 99.99% accuracy rates (Rayson et al. 2004). Accuracy rates are also known as probability values ( $p$ ) and they can be expressed with numbers comprised between 0 and 1 (with 95% accuracy corresponding to a 0.05 probability value etc.; Baker 2006). If the outcome of a chi-square calculation is higher than the critical value then it is likely for the difference in frequency not to be due to chance (Freddi 2014).

According to Oakes (1998), a striking problem with the use of chi-square is that it presupposes a normal distribution of the data, i.e. one that is bell shaped (high in the centre and asymptotically approaching the zero on the x axis of a coordinate system). Normal distribution of data can be encountered in many spheres of life, from the comparison of the heights of human beings to the results of psychological tests (Oakes 1998). As Dunning (1993) observes, this kind of distribution seems to hold for high-frequency linguistic patterns but it collapses with rare events which can be better understood with a loglinear model of distribution such as the Log Likelihood.

In the same way as the chi-square does, also Log Likelihood allows to work on contingency tables so as to detect relationships between variables; what Log Likelihood adds to chi-square is the possibility to deal with bigger tables involving more independent variables to influence frequency and, consequently, more degrees of freedom (Oakes 1998). Even if Log Likelihood compares observed data with model data retrieved with iterative proportional scaling in more complex cases, for the same contingency tables seen for the chi-square, the mathematical formula of Log Likelihood is reduced to:

$$G^2 = 2 \sum O (\log_e O - \log_e E).$$

As far as critical values of Log Likelihood are concerned, a study by Rayson et al. (2004) shows that with all degrees of accuracy (e.g. 5%, 1%, 0.1%, 0.01%, etc.) a cut-off threshold of 15.13 should be adopted to ensure that the results be more robust.

Recent research on keyword extraction, nonetheless, maintains that statistical measures like Log Likelihood should be combined with effect-size metrics to bolster reliability (among others, Gabrielatos 2018 and Pojanapunya and Watson Todd 2018).

Effect-size metrics measures frequency differences across corpora. Effect size, in fact, is the “% difference of the frequency of a word in the study corpus when compared to that in the reference corpus” (Gabrielatos and Marchi 2011: 9). Gabrielatos (2018) lists some of these metrics and concludes that *Ratio* (Kilgarriff 2009), *Odds Ratio* (Everitt 2002, Pojanapunya and Watson Todd 2018), *Log Ratio* (Hardie 2014a), *%Diff* (Gabrielatos and Marchi 2011), and *Difference Coefficient* (Hofland and Johansson 1982) are those that come in handy when exploring statistical keywords (Gabrielatos 2018).

All these effect-size metrics compare to some extent the raw frequency or the relative frequency of the types of a study corpus to those of the same items in a reference corpus. While effect sizes differ depending on the specific formula of the metrics, all those metrics produce equally ranking keyword lists and “the selection of one rather than another hinges on the availability in corpus tools, and the extent to which researchers find their values easy to interpret” (Gabrielatos 2018: 237).

What effect-size metrics (except for Difference Coefficient) also share is a limitation: due to the fact that they consist in a division, when the frequency of a type in the reference corpus is zero, the ratio cannot be calculated. This can be extremely problematic when a researcher is interested in studying presence and absence of linguistic items in a corpus. Nonetheless, this problem is overcome by computer software in various ways: software, for example, might “replace zero frequencies with an infinitesimally small number” (Gabrielatos 2018: 237; see also Gabrielatos and Marchi 2011 and Scott 2015).

Moreover, most effect-size metrics are directional. This means that the outcome of keyword extraction significantly changes according to the direction of the comparison: if two corpora C1 and C2 are available, a keyword list obtained by comparing the word frequencies of C1 to those of C2 will differ from one obtained by comparing C2 to C1.

As far as thresholds are concerned, Gabrielatos and Marchi (2012) believe that setting a limit for the extraction of statistical keywords is less straightforward with effect-size metrics than with statistical significance. Using %DIFF, they recommend that a threshold “has to be relative to the resulting range of %DIFF values”; according to them, for instance, “a 50%DIFF is relatively small [...] if most values are larger than 200%”, it is relatively “large, if most values are smaller than 20%” (Gabrielatos and Marchi 2012: 30). In other terms, a threshold whereby the last statistical keyword to be selected is characterised by a 50% effect size will be too low if most statistical keywords

display effect sizes around 200%, too high if most statistical keywords measure effect sizes around 20%.

In addition, Gabrielatos and Marchi (2012) claim that if one wants to explore in depth the first N statistical keywords, they should not stop at the Nth key item but they ought to take into consideration all those statistical keywords that share the same effect size value as the Nth.

Another difference between effect-size metrics and statistical significance regards the ranking of key items within a keyword list (Gabrielatos and Marchi 2011, 2012). The reason for this is that “statistical significance does not reveal the size of a frequency difference, but, indirectly, the level of confidence we can have that the difference we have observed (however large or small) is dependable” (Gabrielatos 2018: 230). Effect-size metrics, on the other hand, ranks statistical keywords according to frequency differences.

Furthermore, not only does statistical significance disregard frequency differences, but it is also “sensitive to the size of the sample: the larger the sample, the higher the statistical significance of all effect-sizes, however small they may be” (Gabrielatos 2018: 233).

Effect-size metrics and statistical significance seem to diverge also in the research purposes they serve. Pojanapunya and Watson Todd (2018), for example, compare keyword extraction using Log Likelihood (LL) as a statistical significance measure and Odd Ratio (OD) as an effect-size metric. They assume that

there is a range of purposes in research which can be represented as a continuum from genre-oriented research at one hand to critical research at the other. Using LL as the statistic in keyword analysis is more appropriate for research purposes at the genre-oriented end of the continuum, and using OR more appropriate for the critical end. For purposes falling near the middle of the continuum, one or the other might be chosen (Pojanapunya and Watson Todd 2018: 161).

*Collocational patterns.* Brezina (2018a: 70) lists some association measures employed for the retrieval of collocational patterns with their main characteristics:

Some collocation measures such as MI highlight rare exclusivity of the collocational relationship, favouring collocates which occur almost exclusively in the company of the node, even though this may be only once or twice the entire corpus. Other metrics, such as Dice and log Dice, and MI2 favour collocates which occur exclusively in each other's

company but do not have to be rare. Others can take into account directionality (Delta P) or dispersion (Cohen's *d*).

Mutual Information or MI (first reported by Church and Hanks 1990, Church et al. 1991, Clear 1993) “compares the probability of observing *x* and *y* together (the joint probability) with the probabilities of observing *x* and *y* independently (chance)” (Church and Hanks 1990: 23). In other terms, it compares the probability of observing the node together with the collocate, with the probability of observing them independently.

Mutual Information establishes thus “how strongly two words seem to associate in a corpus”, hence it tends to pop up rare words (generally content words; Hunston 2002: 72). As a result of this, it is the measure adopted in much corpus-assisted studies of discourse since it favours links between lexemes and since it showcases “highly specialised terms”, typical of the investigated corpus (Evert 2008: 1238).

Mutual Information-2 (MI2, the square version of MI) and Mutual Information-3 (MI3, the cube version of MI, or MI cubic, put forward by Daille 1995) lower MI's bias towards rare words by assigning a bigger importance to frequency. The result is a list of collocates featuring both content words and function words, where exclusivity of association is matched with usual appearance in the corpus (Brezina et al. 2015). Mutual Information cubic, in particular, “appears as a good compromise between the fact to retain only rare events and to neglect them too much” (Daille 1995: 21).

Dice (Smadja 1993) and Dice coefficient (Smadja et al. 1996) are symmetric association measures that outperform Mutual Information in many instances but that “cannot be used to identify word pairs with strong negative association”, namely when node and collocate tend to be firmly independent from each other (Evert 2008: 1235). In addition, the scores they produce are generally low. These low scores are fixed in the log Dice version (suggested by Rychlý 2008), that “has a reasonable interpretation, scales well on a different corpus size, is stable on subcorpora, and the values are in reasonable range” in comparison with Dice and Dice coefficient (Rychlý 2008: 7).

While being useful for the extraction of collocational patterns, these association measures disregard the notion of directionality in computing collocations (Brezina et al. 2015). Delta P seems to solve this (Gries 2013). According to Gries (2013), Delta P “provides directionality information” and it “is more sensitive than all traditional measures because, unlike them, it can tease apart which collocates in a collocation

exhibit the strongest or weakest amounts of attraction or repulsion to the other collocate(s)” (Gries 2013: 152).

Cohen’s *d* (originally elaborated by Cohen 1988, but see also Algina et al. 2005), on the other hand, measures the effect of corpus size on collocation networks and it foregrounds the dispersion of collocations; in fact, it “takes into account the distribution of collocates in different texts (or subcorpora) by comparing the values of collocate frequencies in the collocation window and outside of the window in each text/subcorpus” (Brezina et al. 2015: 163).

Log Likelihood too is used for the computation of collocates. Unlike Mutual Information, it highlights terms with high frequency. In most European languages it usually makes function words emerge first, since they are the most common items in communication (Evert 2008).

## A2. Analysis of the 2030 Agenda Corpus

Rank	Keyword	Frequency	Keyness measure	Effect size measure
1	<i>sustainable</i>	177	1076.85	0.0128
2	<i>and</i>	1237	1008.95	0.0015
3	<i>2030</i>	73	881.2	0.009
4	<i>development</i>	202	685.04	0.0058
5	<i>including</i>	113	464.41	0.0062
6	<i>developing</i>	95	431.16	0.0064
7	<i>countries</i>	184	402.29	0.0033
8	<i>goals</i>	64	378.46	0.0062
9	<i>inclusive</i>	40	329.55	0.0048
10	<i>global</i>	78	310.57	0.005
11	<i>developed</i>	60	292.89	0.0052
12	<i>agenda</i>	71	282.93	0.0048
13	<i>nations</i>	58	233.81	0.0043
14	<i>goal</i>	40	207.14	0.004
15	<i>targets</i>	41	203.33	0.004
16	<i>forum</i>	34	199.66	0.0037
17	<i>landlocked</i>	15	189.56	0.0019
18	<i>all</i>	175	188.2	0.0018
19	<i>levels</i>	45	183.37	0.0037
20	<i>technology</i>	40	165.87	0.0035
21	<i>promote</i>	45	156.56	0.0033
22	<i>recognize</i>	22	152.07	0.0026
23	<i>access</i>	52	151.6	0.0031
24	<i>least</i>	51	143.47	0.003
25	<i>national</i>	71	140.79	0.0025
26	<i>resilient</i>	12	140.13	0.0015

27	<i>universal</i>	28	139.79	0.0029
28	<i>implementation</i>	49	139.56	0.0029
29	17	29	133.65	0.0029
30	<i>united</i>	56	123.74	0.0025
31	<i>reviews</i>	14	118.69	0.0017
32	<i>reaffirm</i>	16	116.6	0.0019
33	<i>island</i>	22	115.63	0.0024
34	3	37	113.45	0.0027
35	<i>capacity</i>	30	108.02	0.0026
36	<i>international</i>	61	107.12	0.0021
37	<i>b</i>	17	105.72	0.002
38	<i>affordable</i>	15	101.98	0.0018
39	<i>ababa</i>	10	101.97	0.0013
40	<i>addis</i>	10	100.56	0.0013
41	5	34	97.85	0.0024
42	1	42	97.16	0.0023
43	<i>relevant</i>	28	95.57	0.0024
44	<i>vulnerable</i>	21	95.17	0.0022
45	<i>girls</i>	15	93.95	0.0018
46	4	30	93.82	0.0024
47	<i>mobilize</i>	9	91.13	0.0011
48	<i>impacts</i>	13	90.49	0.0016
49	<i>substantially</i>	16	90.33	0.0019
50	<i>enhance</i>	18	88.93	0.002
51	<i>organization</i>	17	86.99	0.0019
52	<i>equitable</i>	13	86.47	0.0016
53	<i>strengthen</i>	26	85.47	0.0022
54	<i>ensure</i>	50	85.03	0.002
55	2	33	84.32	0.0022
56	2020	22	84.26	0.0021
57	15	27	83.04	0.0022
58	<i>debt</i>	20	82.28	0.002
59	<i>innovation</i>	25	82.01	0.0022
60	<i>revitalized</i>	6	81.77	0.0008
61	7	25	81.75	0.0022
62	<i>follow</i>	25	80.57	0.0022
63	<i>empowerment</i>	9	80.47	0.0011
64	<i>resources</i>	38	80.06	0.0021
65	8	24	78.61	0.0021
66	<i>poverty</i>	28	78.24	0.0021
67	<i>stakeholder</i>	9	76.85	0.0011
68	<i>science</i>	16	76.8	0.0018
69	<i>forms</i>	23	76.61	0.0021
70	<i>review</i>	25	75.11	0.0021
71	6	24	73.42	0.002
72	<i>dimensions</i>	12	73.29	0.0014
73	<i>degradation</i>	10	72.69	0.0012

74	<i>through</i>	44	71.82	0.0018
75	<i>vol</i>	5	71.42	0.0006
76	<i>c</i>	12	71.33	0.0014
77	<i>ecosystems</i>	11	71.32	0.0013
78	<i>planet</i>	13	70.56	0.0015
79	<i>sustainably</i>	8	70.06	0.001
80	<i>reduce</i>	25	69.88	0.002
81	<i>oda</i>	8	68.38	0.001
82	<i>world</i>	49	68.09	0.0017
83	<i>by</i>	164	66.69	0.0011
84	<i>safe</i>	18	66.45	0.0018
85	<i>stakeholders</i>	13	66.37	0.0015
86	<i>cent</i>	12	65.36	0.0014
87	<i>16</i>	18	65.17	0.0018
88	<i>per</i>	20	64.23	0.0018
89	<i>building</i>	20	63.73	0.0018
90	<i>economic</i>	60	63.69	0.0015
91	<i>capacities</i>	11	63.16	0.0013
92	<i>policies</i>	32	60.4	0.0018
93	<i>achieve</i>	30	60.26	0.0018
94	<i>water</i>	22	59.94	0.0018
95	<i>processes</i>	15	59.63	0.0016
96	<i>integrated</i>	18	59.39	0.0017
97	<i>mobilization</i>	6	58.61	0.0008
98	<i>domestic</i>	16	58.32	0.0016
99	<i>implement</i>	21	57.58	0.0018
100	<i>9</i>	19	57.45	0.0017
101	<i>regional</i>	28	56.65	0.0018
102	<i>10</i>	23	56.25	0.0018
103	<i>action</i>	41	56.16	0.0016
104	<i>climate</i>	26	56.08	0.0018
105	<i>gender</i>	17	56.01	0.0016
106	<i>platform</i>	11	56	0.0013
107	<i>accordance</i>	19	54.77	0.0017
108	<i>realization</i>	6	53.43	0.0007
109	<i>annex</i>	12	53.29	0.0014
110	<i>resilience</i>	6	53.18	0.0007
111	<i>subregional</i>	5	52.15	0.0006
112	<i>patterns</i>	8	52.09	0.001
113	<i>challenges</i>	20	51.96	0.0017
114	<i>build</i>	16	51.86	0.0016
115	<i>productive</i>	11	51.72	0.0013
116	<i>14</i>	16	51.38	0.0016
117	<i>oceans</i>	7	49.04	0.0009
118	<i>level</i>	39	47.83	0.0015
119	<i>marine</i>	11	47.13	0.0012
120	<i>particular</i>	39	47.08	0.0015

121	2015	9	46.01	0.0011
122	<i>communicable</i>	6	44.61	0.0007
123	<i>land</i>	14	43.9	0.0014
124	<i>sanitation</i>	6	43.79	0.0007
125	<i>facilitate</i>	13	43.65	0.0013
126	<i>full</i>	24	43.29	0.0015
127	<i>situations</i>	14	43.2	0.0014
128	<i>strategies</i>	12	43.07	0.0013
129	<i>sustained</i>	9	42.8	0.0011
130	<i>small</i>	26	42.06	0.0015
131	<i>assembly</i>	13	41.08	0.0013
132	<i>support</i>	53	40.49	0.0012
133	<i>decent</i>	9	39.8	0.001
134	<i>ethnicity</i>	5	39.68	0.0006
135	13	14	39.66	0.0013
136	<i>women</i>	32	39.38	0.0014
137	<i>significantly</i>	11	38.92	0.0012
138	<i>high</i>	27	38.8	0.0014
139	<i>income</i>	13	38.19	0.0013
140	<i>multi</i>	9	38.09	0.001
141	<i>reliable</i>	10	37.99	0.0011
142	<i>determined</i>	11	37.84	0.0012
143	<i>effective</i>	24	37.79	0.0014
144	<i>biodiversity</i>	10	37.52	0.0011
145	12	17	37.31	0.0014
146	<i>provide</i>	25	37.24	0.0014
147	<i>progress</i>	27	36.61	0.0014
148	<i>consumption</i>	12	36.57	0.0012
149	<i>desertification</i>	6	36.34	0.0007
150	<i>education</i>	21	36.18	0.0014
151	<i>indivisible</i>	5	35.96	0.0006
152	<i>conserve</i>	5	35.81	0.0006
153	<i>african</i>	12	35.78	0.0012
154	<i>hunger</i>	8	35.78	0.0009
155	<i>commit</i>	9	35.57	0.001
156	<i>mitigation</i>	5	35.44	0.0006
157	<i>management</i>	20	35.23	0.0014
158	<i>primary</i>	10	34.7	0.0011
159	<i>private</i>	16	34.55	0.0013
160	<i>enhanced</i>	9	34.51	0.001
161	<i>upgrade</i>	5	34.44	0.0006
162	<i>conferences</i>	7	34.31	0.0008
163	<i>equal</i>	17	34.29	0.0013
164	<i>millennium</i>	9	33.92	0.001
165	<i>auspices</i>	6	33.86	0.0007
166	<i>partnership</i>	16	33.86	0.0013
167	<i>related</i>	14	33.79	0.0013

168	<i>gni</i>	5	33.61	0.0006
169	<i>11</i>	15	32.63	0.0013
170	<i>scientific</i>	14	32.63	0.0012
171	<i>online</i>	7	32.23	0.0008
172	<i>persons</i>	12	32.23	0.0012
173	<i>boys</i>	5	32.17	0.0006
174	<i>quality</i>	21	31.35	0.0013
175	<i>eradication</i>	6	30.93	0.0007
176	<i>youth</i>	10	30.58	0.001
177	<i>health</i>	28	30.21	0.0012
178	<i>race</i>	7	30.13	0.0008
179	<i>natural</i>	14	30.08	0.0012
180	<i>societies</i>	9	30.07	0.001
181	<i>foster</i>	7	29.9	0.0008
182	<i>declaration</i>	12	29.57	0.0011
183	<i>conference</i>	18	29.42	0.0012
184	<i>adverse</i>	7	29.21	0.0008
185	<i>forums</i>	5	29.02	0.0006
186	<i>peoples</i>	11	28.97	0.0011
187	<i>taking</i>	21	28.84	0.0012
188	<i>everywhere</i>	7	28.32	0.0008
189	<i>outcomes</i>	5	27.93	0.0006
190	<i>gross</i>	7	27.71	0.0008
191	<i>indigenous</i>	6	27.7	0.0007
192	<i>partnerships</i>	7	27.65	0.0008
193	<i>planning</i>	10	27.17	0.001
194	<i>accountable</i>	6	27.14	0.0007
195	<i>practices</i>	11	26.83	0.0011
196	<i>reporting</i>	7	26.81	0.0008
197	<i>halt</i>	6	26.74	0.0007
198	<i>extreme</i>	8	26.69	0.0009
199	<i>mortality</i>	5	26	0.0006
200	<i>facilitation</i>	5	25.96	0.0006
201	<i>reproductive</i>	6	25.95	0.0007
202	<i>data</i>	17	25.71	0.0012
203	<i>coherence</i>	7	25.4	0.0008
204	<i>inequality</i>	6	25.36	0.0007
205	<i>human</i>	36	25.35	0.0011
206	<i>forests</i>	7	25.22	0.0008
207	<i>equality</i>	12	24.56	0.0011
208	<i>acknowledge</i>	8	24.32	0.0009
209	<i>infrastructure</i>	11	24.29	0.001
210	<i>social</i>	38	23.98	0.0011
211	<i>address</i>	13	23.83	0.0011
212	<i>change</i>	22	23.55	0.0011
213	<i>increase</i>	20	23.36	0.0011
214	<i>ownership</i>	6	23.18	0.0007

215	<i>framework</i>	24	23.13	0.0011
216	<i>appropriate</i>	16	23.05	0.0011
217	<i>child</i>	10	23.03	0.001
218	<i>disabilities</i>	6	23	0.0007
219	<i>capabilities</i>	5	21.8	0.0006
220	<i>protect</i>	13	21.68	0.001
221	<i>0</i>	7	21.56	0.0008
222	<i>resource</i>	6	21.27	0.0007
223	<i>migrants</i>	6	21.05	0.0007
224	<i>organizations</i>	6	21.03	0.0007
225	<i>disability</i>	5	20.74	0.0006
226	<i>diseases</i>	8	20.67	0.0008
227	<i>69</i>	5	20.58	0.0006
228	<i>knowledge</i>	11	20.2	0.001
229	<i>timely</i>	6	20.2	0.0007
230	<i>species</i>	7	20.11	0.0008
231	<i>other</i>	50	20.1	0.0009
232	<i>inter</i>	6	19.57	0.0007
233	<i>will</i>	123	19.45	0.0008
234	<i>encourage</i>	11	19.22	0.0009
235	<i>integral</i>	5	19.21	0.0006
236	<i>ongoing</i>	7	19.07	0.0008
237	<i>adaptation</i>	5	18.54	0.0006
238	<i>agreed</i>	13	18.4	0.001
239	<i>sources</i>	9	18.34	0.0009
240	<i>learning</i>	7	18.33	0.0007
241	<i>growth</i>	17	18.24	0.001
242	<i>institutions</i>	20	18.17	0.001
243	<i>participation</i>	10	18.17	0.0009
244	<i>shared</i>	8	18.1	0.0008
245	<i>reduction</i>	11	18.07	0.0009
246	<i>combat</i>	10	17.97	0.0009
247	<i>restore</i>	6	17.7	0.0007
248	<i>productivity</i>	5	17.38	0.0006
249	<i>sound</i>	8	17.33	0.0008
250	<i>vocational</i>	5	17.2	0.0006
251	<i>achievement</i>	6	16.93	0.0007
252	<i>enabling</i>	6	16.8	0.0007
253	<i>existing</i>	12	16.77	0.0009
254	<i>actors</i>	5	16.75	0.0006
255	<i>statistical</i>	5	16.7	0.0006
256	<i>doha</i>	5	16.62	0.0006
257	<i>humanity</i>	5	16.62	0.0006
258	<i>programmes</i>	15	16.45	0.001
259	<i>post</i>	7	16.39	0.0007
260	<i>cities</i>	6	16.26	0.0007
261	<i>poor</i>	9	16.16	0.0008

262	<i>disasters</i>	7	16.05	0.0007
263	<i>technological</i>	7	15.91	0.0007
264	<i>well</i>	30	15.82	0.0009
265	<i>communities</i>	8	15.81	0.0008
266	<i>develop</i>	11	15.75	0.0009
267	<i>production</i>	14	15.69	0.0009
268	<i>chemicals</i>	5	15.55	0.0006
269	<i>identify</i>	6	15.54	0.0007
270	<i>generation</i>	6	15.28	0.0007
271	<i>finance</i>	8	15.21	0.0008
272	<i>sustainability</i>	6	15.18	0.0007

**Table 44.** Positive keywords from the 2030 Agenda Corpus (English).

Rank	English keyword	Hungarian keyword	Italian keyword
1	<i>sustainable</i>	<i>fenntartható</i> ‘sustainable’	<i>sviluppo</i> ‘development’
2	<i>and</i>	<i>2030</i> ‘2030’	<i>2030</i> ‘2030’
3	<i>2030</i>	<i>agenda</i> ‘agenda’	<i>sostenibile</i> ‘sustainable’
4	<i>development</i>	<i>és</i> ‘and’	<i>e</i> ‘and’
5	<i>including</i>	<i>ig</i> ‘by’	<i>entro</i> ‘by’
6	<i>developing</i>	<i>fejlődő</i> ‘developing’	<i>sviluppatti</i> ‘developed’
7	<i>countries</i>	<i>beleértve</i> ‘including’	<i>agenda</i> ‘agenda’
8	<i>goals</i>	<i>fejlett</i> ‘developed’	<i>via</i> (part of <i>in via di sviluppo</i> ‘developing’)
9	<i>inclusive</i>	<i>fejlesztési</i> ‘development’	<i>traguardi</i> ‘targets’
10	<i>global</i>	<i>alcélok</i> ‘targets’	<i>paesi</i> ‘countries’
11	<i>developed</i>	<i>legkevesbé</i> ‘least’	<i>forum</i> ‘forum’
12	<i>agenda</i>	<i>fejlesztés</i> ‘development’	<i>globale</i> ‘global’
13	<i>nations</i>	<i>országok</i> ‘countries’	<i>nazioni</i> ‘nations’
14	<i>goal</i>	<i>globális</i> ‘global’	<i>unite</i> ‘united’
15	<i>targets</i>	<i>valamint</i> ‘and’	<i>obiettivi</i> ‘goals’
16	<i>forum</i>	<i>szárazfölddel</i> (part of <i>szárazfölddel körülvevett</i> ‘landlocked’)	<i>sostenibili</i> ‘sustainable’
17	<i>landlocked</i>	<i>elismerjük</i> ‘(we) recognize; (we) acknowledge’	<i>insulari</i> ‘island’
18	<i>all</i>	<i>forum</i> ‘forum’	<i>riconosciamo</i> ‘(we) recognize; (we) acknowledge’
19	<i>levels</i>	<i>körülvevett</i> (part of <i>szárazfölddel körülvevett</i> ‘landlocked’)	<i>capacità</i> ‘CAPACITY; capabilities’
20	<i>technology</i>	<i>célok</i> ‘goals’	<i>accesso</i> ‘access; facilitation’
21	<i>promote</i>	<i>biztosítása</i> ‘ensure; provide’	<i>implementare</i> ‘implement’
22	<i>recognize</i>	<i>cél</i> ‘goal’	<i>17</i> ‘17’
23	<i>access</i>	<i>szakpolitikák</i> ‘policies’	<i>sbocco</i> (part of <i>senza sbocco sul mare</i> ‘landlocked’)

24	<i>least</i>	<i>befogadó</i> 'inclusive'	<i>lo</i> 'the'
25	<i>national</i>	<i>nemzeti</i> 'national'	<i>promuovere</i> 'promote; foster'
26	<i>resilient</i>	<i>hozzáférés</i> 'access'	3 '3'
27	<i>universal</i>	<i>minden</i> 'all'	<i>universale</i> 'universal'
28	<i>implementation</i>	<i>elősegítése</i> 'promote; facilitate'	<i>supportare</i> 'support'
29	17	<i>mindenki</i> 'all'	<i>livello</i> 'level'
30	<i>united</i>	<i>megerősítjük</i> '(we) reaffirm'	<i>tutti</i> 'all'
31	<i>reviews</i>	<i>elhatározásunk</i> 'we are determined'	<i>inclusiva</i> 'inclusive'
32	<i>reaffirm</i>	17 '17'	<i>piccoli</i> 'small'
33	<i>island</i>	3 '3'	<i>risorse</i> 'resources; sources'
34	3	<i>egyetemes</i> 'universal'	<i>b</i> 'b'
35	<i>capacity</i>	<i>b</i> 'b'	<i>ribadiamo</i> 'reaffirm'
36	<i>international</i>	<i>szigetállamok</i> 'island [...] states'	1 '1'
37	<i>b</i>	<i>szintű</i> 'level'	<i>internazionale</i> 'international'
38	<i>affordable</i>	<i>csökkentése</i> 'reduce; reduction'	5 '5'
39	<i>ababa</i>	<i>országokban</i> 'in (the) countries'	4 '4'
40	<i>addis</i>	<i>ensz</i> 'United Nations'	<i>implementazione</i> 'implementation'
41	5	<i>abebai</i> 'Ababa'	<i>resilienti</i> 'resilient'
42	1	<i>megfizethető</i> 'affordable'	<i>livelli</i> 'levels'
43	<i>relevant</i>	<i>megvalósítása</i> 'its implementation; its achievement'	<i>addis</i> 'Addis'
44	<i>vulnerable</i>	<i>releváns</i> 'relevant'	<i>abeba</i> 'Ababa'
45	<i>girls</i>	<i>nyomon</i> (in <i>nyomon követés</i> 'follow-up')	<i>meno</i> 'least'
46	4	4 '4'	<i>attraverso</i> 'through'
47	<i>mobilize</i>	<i>addisz</i> 'Addis'	7 '7'
48	<i>impacts</i>	<i>növelése</i> 'enhance; increase'	2 '2'
49	<i>substantially</i>	<i>szigetállamokban</i> 'in (the) island [...] states'	<i>monitoraggio</i> 'follow-up'
50	<i>enhance</i>	<i>világot</i> 'world'	15 '15'
51	<i>organization</i>	<i>nemzetközi</i> 'international'	<i>vol</i> 'vol'
52	<i>equitable</i>	<i>felszámolása</i> 'eradicate; eradication'	<i>povertà</i> 'poverty'
53	<i>strengthen</i>	5 '5'	8 '8'
54	<i>ensure</i>	<i>érdekcsoportot</i> 'multi-stakeholder'	2020 '2020'
55	2	<i>fejlesztési</i> 'development'	i
56	2020	<i>ellenállóképes</i> 'resilient'	<i>inclusive</i> 'inclusive'
57	15	<i>előrehaladás</i> 'progress'	<i>inclusivi</i> 'inclusive'
58	<i>debt</i>	1 '1'	6 '6'
59	<i>innovation</i>	<i>c</i> 'c'	<i>conformità</i> 'accordance'
60	<i>revitalized</i>	<i>erősítése</i> 'strengthen'	<i>nazionale</i> 'national; domestic'
61	7	<i>méltányos</i> 'decent'	<i>oceani</i> 'oceans'

62	<i>follow (in follow-up)</i>	6 '6'	delle 'of the'
63	<i>empowerment</i>	7 '7'	<i>aps</i> 'oda'
64	<i>resources</i>	2 '2'	<i>le</i> 'the'
65	8	<i>inkluzív</i> 'inclusive'	<i>tutte</i> 'all'
66	<i>poverty</i>	15 '15'	<i>mare</i> (part of <i>senza sbocco sul mare</i> 'landlocked')
67	<i>stakeholder</i>	<i>megnyilvánulásában</i> 'in its dimensions'	<i>processi</i> 'processes'
68	<i>science</i>	<i>szanitációhoz</i> 'to (the) sanitation'	<i>innovazione</i> 'innovation'
69	<i>forms</i>	<i>fejlődéshez</i> 'to the development'	<i>ragazze</i> 'girls'
70	<i>review</i>	<i>összhangban</i> 'in accordance with'	<i>supporto</i> 'support'
71	6	<i>jelentős</i> 'substantially; significantly'	<i>emancipazione</i> 'empowerment'
72	<i>dimensions</i>	<i>támogatása</i> 'support'	<i>realizzazione</i> 'realization'
73	<i>degradation</i>	<i>erőforrásokhoz</i> 'to (the) resources'	<i>partnership</i> 'partnership'
74	<i>through</i>	<i>többek (in többek között 'including')</i>	16 '16'
75	<i>vol</i>	<i>ugyancsak</i> 'also'	<i>considerevolmente</i> 'substantially'
76	<i>c</i>	<i>technológiai</i> 'technological'	<i>impegniamo</i> 'we commit'
77	<i>ecosystems</i>	8 '8'	<i>vulnerabili</i> 'vulnerable'
78	<i>planet</i>	<i>lány</i> 'girl'	<i>sfide</i> 'challenges'
79	<i>sustainably</i>	16 '16'	<i>pianeta</i> 'planet'
80	<i>reduce</i>	<i>jólét</i> 'well-being'	<i>ed</i> 'and'
81	<i>oda</i>	<i>érdekelt (in érdekelt fél 'actors')</i>	<i>piattaforma</i> 'platform'
82	<i>world</i>	<i>alcél</i> 'target'	<i>acqua</i> 'water'
83	<i>by</i>	<i>tudományos</i> 'scientific'	<i>forme</i> 'forms'
84	<i>safe</i>	<i>óceánok</i> 'oceans'	9 '9'
85	<i>stakeholders</i>	<i>szakpolitikai</i> 'policy'	<i>nazionali</i> 'national; domestic'
86	<i>cent</i>	<i>illetve</i> 'or'	<i>ridurre</i> 'reduce'
87	16	<i>akcióterv</i> 'action'	<i>educazione</i> 'education'
88	<i>per</i>	<i>történő</i> '-'	<i>tecnologia</i> 'technology'
89	<i>building (in capacity-building)</i>	<i>támogató</i> 'enabling'	<i>mondo</i> 'world'
90	<i>economic</i>	<i>formájának</i> 'forms'	<i>alto</i> 'high'
91	<i>capacities</i>	<i>egyéb</i> 'other'	<i>obiettivo</i> 'goal'
92	<i>policies</i>	<i>a</i> 'the'	<i>raggiungimento</i> 'achievement'
93	<i>achieve</i>	9 '9'	<i>ogni</i> 'all'
94	<i>water</i>	<i>ellenálló</i> 'resilient'	<i>duratura</i> 'sustained'
95	<i>processes</i>	<i>felülvizsgálati</i> 'review'	<i>inclusi</i> 'including'
96	<i>integrated</i>	<i>szolgáltatásokhoz</i> 'to the services'	<i>globali</i> 'global'
97	<i>mobilization</i>	<i>regionális</i> 'regional'	14 '14'
98	<i>domestic</i>	<i>kiszolgáltatott</i> 'vulnerable'	<i>scienza</i> 'science'
99	<i>implement</i>	<i>közgyűlés</i> 'General Assembly'	<i>porre (in porre fine 'end')</i>
100	9	<i>ökoszisztémák</i> 'ecosystems'	<i>climatico</i> 'climate'

101	<i>regional</i>	<i>szubregionális</i> 'subregional'	<i>emergenti</i> 'developing; emerging'
102	10	<i>követési</i> (in <i>nyomon követési</i> 'follow-up')	<i>fine</i> (in <i>al fine di</i> 'to' and <i>porre fine</i> 'end')
103	<i>action</i>	<i>valamennyi</i> 'all'	2015 '2015'
104	<i>climate</i>	<i>biztonságos</i> 'safe'	<i>ecosistemi</i> 'ecosystems'
105	<i>gender</i>	10 '10'	<i>rafforzare</i> 'strengthen'
106	<i>platform</i>	<i>számára</i> 'for'	<i>indivisibili</i> 'indivisible'
107	<i>accordance</i>	<i>megszüntetése</i> 'end, eliminate'	10 '10'
108	<i>realization</i>	<i>nemek</i> 'gender'	<i>attori</i> 'stakeholders; actors'
109	<i>annex</i>	<i>lányok</i> 'girls'	<i>bisogni</i> 'needs'
110	<i>resilience</i>	<i>technológia</i> 'technology'	<i>interconnessi</i> 'integrates'
111	<i>subregional</i>	<i>végrehajtás</i> 'implementation'	<i>regionale</i> 'regional'
112	<i>patterns</i>	<i>megerősítése</i> 'empowerment'	<i>proteggere</i> 'protect'
113	<i>challenges</i>	<i>elérését</i> 'achieve'	<i>raggiungere</i> 'achieve'
114	<i>build</i>	<i>hatékony</i> 'effective'	<i>razza</i> 'race'
115	<i>productive</i>	14 '14'	<i>mitigazione</i> 'mitigation'
116	14	<i>magas</i> 'high'	<i>pacifiche</i> 'peaceful'
117	<i>oceans</i>	<i>biológiai</i> 'biological' (in <i>biodiversity</i> )	<i>sottolineiamo</i> 'we emphasise'
118	<i>level</i>	<i>tartós</i> 'sustained'	<i>online</i> 'online'
119	<i>marine</i>	<i>kis</i> 'small'	<i>tecnologiche</i> 'technological'
120	<i>particular</i>	<i>bolygó</i> 'planet'	<i>aumentare</i> 'increase'
121	2015	<i>minőségi</i> 'quality'	<i>garantire</i> 'provide'
122	<i>communicable</i>	<i>jövedelmű</i> 'income'	<i>reddito</i> 'income'
123	<i>land</i>	<i>hangsúlyozzuk</i> 'we emphasise'	<i>attuazione</i> 'implementation'
124	<i>sanitation</i>	<i>felülvizsgálatok</i> 'reviews'	<i>rnl</i> 'gni'
125	<i>facilitate</i>	<i>szinten</i> 'level'	<i>debito</i> 'debt'
126	<i>full</i>	<i>integrált</i> 'integrated'	<i>mondiale</i> 'world'
127	<i>situations</i>	2020 '2020'	<i>equo</i> 'equitable; equal'
128	<i>strategies</i>	<i>kiépítése</i> 'build'	<i>economico</i> 'economic'
129	<i>sustained</i>	<i>sajátosságokat</i> 'realities'	<i>disastri</i> 'disasters'
130	<i>small</i>	<i>platform</i> 'platform'	<i>degrado</i> 'degradation'
131	<i>assembly</i> (in <i>General Assembly</i> )	<i>kapacitásokat</i> 'capacities'	<i>affidabili</i> 'reliable'
132	<i>support</i>	<i>elkötelezzük</i> 'we commit'	<i>trasmissibili</i> 'communicable'
133	<i>decent</i>	<i>ösztönzése</i> 'foster; encourage'	<i>organizzazione</i> 'organization'
134	<i>ethnicity</i>	<i>fertőző</i> 'communicable'	<i>progresso</i> 'progress'
135	13	<i>fogunk</i> 'we will'	<i>politiche</i> 'policies'
136	<i>women</i>	<i>elkötelezettek</i> 'committed'	<i>tecnologie</i> 'technologies'
137	<i>significantly</i>	<i>tömörítő</i> 'collaborative'	13 '13'
138	<i>high</i>	2016 '2016'	<i>particolare</i> 'particular'
139	<i>income</i>	<i>társadalmi</i> 'social'	<i>donne</i> 'women'

140	<i>multi</i>	11 '11'	<i>valutazioni</i> 'reviews'
141	<i>reliable</i>	12 '12'	<i>africani</i> 'african'
142	<i>determined</i>	<i>égisze</i> 'auspices'	<i>finanza</i> 'finance'
143	<i>effective</i>	<i>béke</i> 'peace'	<i>strategie</i> 'strategies'
144	<i>biodiversity</i>	<i>továbbá</i> 'also'	12 '12'
145	12	<i>kezelése</i> 'management'	<i>scientifica</i> 'scientific'
146	<i>provide</i>	<i>gazdasági</i> 'economic'	<i>dignitoso</i> 'decent'
147	<i>progress</i>	<i>követés</i> (in <i>nyomon követés</i> 'follow-up')	<i>compresi</i> 'including'
148	<i>consumption</i>	<i>fokozása</i> 'enhance'	<i>incoraggiamo</i> '(we) encourage'
149	<i>desertification</i>	<i>gni</i> 'gni'	<i>verifiche</i> 'reviews'
150	<i>education</i>	<i>felek</i> 'stakeholders'	<i>privato</i> 'private'
151	<i>indivisible</i>	<i>magánszektor</i> 'private sector'	<i>specialmente</i> 'especially'
152	<i>conserve</i>	<i>szerepének</i> (in <i>társadalmi szerepének megerősítése</i> 'empowerment')	<i>naturali</i> 'natural'
153	<i>african</i>	<i>termelési</i> 'production'	<i>conferenze</i> 'conferences'
154	<i>hunger</i>	<i>világszervezet</i> 'organization'	<i>nei</i> 'in the'
155	<i>commit</i>	<i>vetünk</i> (in <i>véget vetünk</i> 'end')	<i>scientifiche</i> 'scientific'
156	<i>mitigation</i>	<i>érdekében</i> 'for'	<i>sociale</i> 'social'
157	<i>management</i>	<i>képességeinek</i> 'to its capacities; to its capabilities'	<i>persone</i> 'persons'
158	<i>primary</i>	<i>fejlettségi</i> 'of development'	<i>etnia</i> 'ethnicity'
159	<i>private</i>	<i>elszámoltatható</i> 'accountable'	<i>determinati</i> 'determined'
160	<i>enhanced</i>	<i>cselekvési</i> 'of action'	<i>lavoreremo</i> 'we will work'
161	<i>upgrade</i>	<i>megújított</i> 'revitalized'	<i>allegato</i> 'annex'
162	<i>conferences</i>	<i>egymástól</i> (in <i>egymástól elválaszthatlan(ok)</i> 'indivisible')	<i>tenendo</i> 'taking'
163	<i>equal</i>	<i>megállítása</i> 'halt'	11 '11'
164	<i>millennium</i>	<i>kapacitásának</i> 'capacity'	<i>disabilità</i> 'disability; disabilities'
165	<i>auspices</i>	<i>egyenlő</i> 'equitable; equal'	<i>facilitare</i> 'facilitate'
166	<i>partnership</i>	<i>közepes</i> 'middle'	<i>mortalità</i> 'mortality'
167	<i>related</i>	<i>akárcsak</i> 'as well as'	<i>egida</i> 'auspices'
168	<i>gni</i>	<i>társadalmak</i> 'societies'	<i>multilaterale</i> 'multilateral'
169	11	<i>keretrendszer</i> 'framework'	<i>potenziare</i> 'enhance; strengthen'
170	<i>scientific</i>	13 '13'	<i>indigene</i> 'indigenous'
171	<i>online</i>	<i>országokon</i> 'in (the) countries'	<i>poveri</i> 'poor'
172	<i>persons</i>	<i>természeti</i> 'natural'	<i>crescita</i> 'growth'
173	<i>boys</i>	<i>fogják</i> 'they will'	<i>consumo</i> 'consumption'
174	<i>quality</i>	<i>fogjuk</i> 'we will'	<i>lordo</i> 'gross'
175	<i>eradication</i>	<i>erőforrások</i> 'resources'	<i>esistenti</i> 'existing'
176	<i>youth</i>	<i>kapacitások</i> 'capacities'	<i>fame</i> 'hunger'
177	<i>health</i>	<i>helyreállítása</i> 'restore'	<i>per</i> 'to; for'

178	<i>race</i>	<i>megvalósításának</i> 'to its implementation; to its achievement'	<i>fornire</i> 'provide'
179	<i>natural</i>	<i>minimálisra</i> (in <i>minimálisra CSÖKKENT</i> 'minimize')	<i>millennio</i> 'millenium'
180	<i>societies</i>	<i>hozzáférést</i> 'access'	<i>biodiversità</i> 'biodiversity'
181	<i>foster</i>	<i>faji</i> 'race'	<i>particolare</i> (in <i>in particolare modo</i> 'particularly')
182	<i>declaration</i>	<i>egyenlőtlenség</i> 'inequality'	<i>desertificazione</i> 'desertification'
183	<i>conference</i>	<i>különös</i> 'particular'	<i>mobilitare</i> 'mobilize'
184	<i>adverse</i>	<i>megbízható</i> 'reliable'	<i>bambini</i> 'child'
185	<i>forums</i>	<i>országokat</i> 'countries'	<i>azione</i> 'action'
186	<i>peoples</i>	<i>millenniumi</i> 'millennium'	<i>politico</i> 'political'
187	<i>taking</i>	<i>használatának</i> 'use'	<i>rilevanti</i> 'relevant'
188	<i>everywhere</i>	<i>felülvizsgálat</i> 'review'	<i>saranno</i> 'will be'
189	<i>outcomes</i>	<i>hazai</i> 'domestic'	<i>cambiamento</i> 'change'
190	<i>gross</i>	<i>módok</i> 'patterns'	<i>raddoppiare</i> 'double'
191	<i>indigenous</i>	<i>céljából</i> 'from (its) goal'	<i>comprese</i> 'including'
192	<i>partnerships</i>	<i>foglalkozó</i> '-'	<i>conferenza</i> 'conference'
193	<i>planning</i>	<i>kerülő</i> '-'	<i>moderni</i> 'modern'
194	<i>accountable</i>	<i>migránsok</i> 'migrants'	<i>sicuri</i> 'safe'
195	<i>practices</i>	<i>számának</i> 'of'	<i>dello</i> 'of the'
196	<i>reporting</i>	<i>hátra</i> 'behind'	<i>età</i> 'age'
197	<i>halt</i>	<i>mindenhol</i> 'everywhere'	<i>primaria</i> 'primary'
198	<i>extreme</i>	<i>végrehajtása</i> 'realization'	<i>adattamento</i> 'adaptation'
199	<i>mortality</i>	<i>ahol</i> 'in which; in particular'	<i>aperto</i> 'open'
200	<i>facilitation</i>	<i>partnerségek</i> 'partnerships'	<i>insediamenti</i> 'settlements'
201	<i>reproductive</i>	<i>célokkal</i> 'with (the) goals'	<i>riproduttiva</i> 'reproductive'
202	<i>data</i>	<i>fogyasztási</i> 'consumption'	<i>produttiva</i> 'productive'
203	<i>coherence</i>	<i>egységet</i> 'integrated'	<i>strumenti</i> 'means'
204	<i>inequality</i>	<i>megrendezésre</i> 'to the conference'	<i>modelli</i> 'patterns'
205	<i>human</i>	<i>kormányközi</i> 'intergovernmental'	<i>mobilitazione</i> 'mobilization'
206	<i>forests</i>	<i>partnerség</i> 'partnership'	<i>foreste</i> 'forests'
207	<i>equality</i>	<i>nemzetek</i> 'nations'	<i>sviluppare</i> 'develop'
208	<i>acknowledge</i>	<i>online</i> 'online'	<i>migranti</i> 'migrants'
209	<i>infrastructure</i>	<i>0 '0'</i>	<i>apprendimento</i> 'learning'
210	<i>social</i>	<i>növekedés</i> 'growth'	<i>utilizzo</i> 'use'
211	<i>address</i>	<i>senkit</i> 'no one'	<i>equa</i> 'equitable; equal'
212	<i>change</i>	<i>innovációs</i> 'innovation'	<i>progressivamente</i> 'progressively'
213	<i>increase</i>	<i>afrikai</i> 'african'	<i>costruire</i> 'build'
214	<i>ownership</i>	<i>helyzetben</i> 'in (the) situation'	<i>gestione</i> 'management'
215	<i>framework</i>	<i>megfelelő</i> 'appropriate'	<i>assicurare</i> 'ensure'
216	<i>appropriate</i>	<i>nő</i> 'woman'	<i>0 '0'</i>
217	<i>child</i>	<i>betegségek</i> 'diseases'	<i>riconoscendo</i> 'recognizing'

218	<i>disabilities</i>	<i>munkacsoport</i> 'working group'	<i>pratiche</i> 'practices'
219	<i>capabilities</i>	<i>fogyatékkal</i> 'disability; disabilities'	<i>sussidi</i> 'subsidies'
220	<i>protect</i>	<i>szegénység</i> 'poverty'	<i>uguaglianza</i> 'equality'
221	0	<i>szárazföldi</i> 'terrestrial'	<i>incluso</i> 'including'
222	<i>resource</i>	<i>konfliktus</i> 'conflict'	<i>malattie</i> 'diseases'
223	<i>migrants</i>	<i>fogyasztás</i> 'consumption'	69 '69'
224	<i>organizations</i>	<i>elősegítő</i> 'facilitation'	<i>genere</i> 'gender'
225	<i>disability</i>	<i>megőrzése</i> 'conserve'	<i>il</i> 'the'
226	<i>diseases</i>	<i>gyakorlatok</i> 'practices'	<i>adeguati</i> 'appropriate'
227	69	<i>kidolgozása</i> 'build'	<i>tecnologica</i> 'technological'
228	<i>knowledge</i>	<i>infrastruktúra</i> 'infrastructure'	<i>post</i> 'post'
229	<i>timely</i>	<i>ország</i> 'country'	<i>potenziamento</i> 'strengthen'
230	<i>species</i>	<i>tiszteletben</i> (in <i>tiszteletben tart</i> 'to respect')	<i>dimensioni</i> 'dimensions'
231	<i>other</i>	<i>koherens</i> 'coherent'	<i>qualità</i> 'quality'
232	<i>inter</i>	<i>támogatás</i> 'support'	<i>inclusa</i> 'including'
233	<i>will</i>	<i>körű</i> (in <i>teljes körű</i> 'full')	<i>loro</i> 'their'
234	<i>encourage</i>	<i>konferencia</i> 'conference'	<i>pianificazione</i> 'planning'
235	<i>integral</i>	<i>szerves</i> 'integral'	<i>allo</i> 'to the'
236	<i>ongoing</i>	<i>javitása</i> 'improve'	<i>locali</i> 'local'
237	<i>adaptation</i>	<i>között</i> (in <i>többek között</i> 'including through')	<i>combattere</i> 'combat'
238	<i>agreed</i>	<i>sokféleség</i> 'diversity' (in <i>biodiversity</i> )	<i>generazioni</i> 'generation'
239	<i>sources</i>	<i>hatások</i> 'impacts'	<i>commercio</i> 'trade'
240	<i>learning</i>	<i>különösen</i> 'in particular'	<i>verifica</i> 'review'
241	<i>growth</i>	<i>stratégiák</i> 'strategies'	<i>tecnologico</i> 'technological'
242	<i>institutions</i>	<i>levő</i> '-'	<i>leadership</i> 'leadership'
243	<i>participation</i>	<i>folymatok</i> 'processes'	<i>cibo</i> 'food'
244	<i>shared</i>	<i>megfelelően</i> 'in accordance with'	<i>produttività</i> 'productivity'
245	<i>reduction</i>	<i>emberi</i> 'human'	<i>programmi</i> 'programmes'
246	<i>combat</i>	<i>élők</i> 'persons'	<i>concordato</i> 'agreed'
247	<i>restore</i>	<i>ambiciózus</i> 'ambitious'	<i>economica</i> 'economic'
248	<i>productivity</i>	<i>használata</i> 'use'	<i>situazioni</i> 'situations'
249	<i>sound</i>	<i>civil</i> 'civil'	<i>piena</i> 'full'
250	<i>vocational</i>		<i>infrastrutture</i> 'infrastructure'
251	<i>achievement</i>		<i>sanitaria</i> 'health'
252	<i>enabling</i>		
253	<i>existing</i>		
254	<i>actors</i>		
255	<i>statistical</i>		
256	<i>doha</i>		
257	<i>humanity</i>		
258	<i>programmes</i>		
259	<i>post</i>		
260	<i>cities</i>		

261	<i>poor</i>
262	<i>disasters</i>
263	<i>technological</i>
264	<i>well</i>
265	<i>communities</i>
266	<i>develop</i>
267	<i>production</i>
268	<i>chemicals</i>
269	<i>identify</i>
270	<i>generation</i>
271	<i>finance</i>
272	<i>sustainability</i>

**Table 45.** Positive keywords of the 2030 Agenda (English, Hungarian, and Italian).

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>development</i>	69.89832107661503	109	201
2	L	<i>for</i>	15.94079783615737	28	220
3	R	<i>consumption</i>	15.717960454912856	6	12
4	R	<i>management</i>	11.982524089331582	6	20
5	R	<i>use</i>	11.67025701888814	6	21

**Table 46.** List of collocates of *SUSTAINABLE* in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>development</i>	74.48520979136597	116	201
2	R	<i>goals</i>	38.62158926565351	34	64
3	L	<i>sustained</i>	24.450547902172428	8	9
4	L	<i>for</i>	22.204150297962137	38	220
5	M	<i>and</i>	22.099792046166094	98	1237
6	R	<i>consumption</i>	21.08157489979571	8	12
7	L	<i>inclusive</i>	19.883621142892803	14	40
8	R	<i>modern</i>	18.699572715384445	5	6
9	R	<i>resilient</i>	18.399767677354284	7	12
10	L	<i>partnership</i>	18.149538004526335	8	16
11	R	<i>production</i>	16.977337992747117	7	14
12	R	<i>goal</i>	16.9458484281753	12	40
13	R	<i>growth</i>	15.328362062921123	7	17
14	L	<i>reliable</i>	14.348469930506337	5	10
15	R	<i>management</i>	14.059843057826734	7	20
16	L	<i>on</i>	13.70238763601372	19	139

17	R	<i>energy</i>	13.504511491324648	6	16
18	L	<i>promote</i>	13.126708849987466	10	45
19	L	<i>innovation</i>	12.746932923300829	7	24
20	L	<i>of</i>	11.69061216837749	36	553
21	R	<i>use</i>	11.67025701888814	6	21
22	L	<i>the</i>	11.656929590691657	46	837
23	L	<i>affordable</i>	11.576511402422998	5	15
24	L	<i>to</i>	10.795582903978833	29	428

**Table 47.** List of collocates of *SUSTAINABLE* in the 2030 Agenda Corpus (English) – Z(10.0), 3L-3R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>development</i>	79.07209850611693	123	201
2	R	<i>goals</i>	39.78284589395393	35	64
3	L	<i>and</i>	37.41988270618654	156	1237
4	L	<i>for</i>	32.85184948303024	55	220
5	R	<i>goal</i>	25.759166572327807	18	40
6	L	<i>the</i>	24.822489070299078	87	837
7	L	<i>sustained</i>	24.450547902172428	8	9
8	L	<i>inclusive</i>	24.290280214969055	17	40
9	R	<i>consumption</i>	23.76338212223714	9	12
10	L	<i>resilient</i>	23.76338212223714	9	12
11	R	<i>agriculture</i>	23.009846303894754	5	4
12	R	<i>production</i>	21.943080143555626	9	14
13	L	<i>promote</i>	20.051105558550592	15	45
14	L	<i>of</i>	19.986730167086314	57	553
15	L	<i>to</i>	19.776616125731017	49	428
16	R	<i>patterns</i>	19.402720890955962	6	8
17	R	<i>modern</i>	18.699572715384445	5	6
18	L	<i>innovation</i>	18.43590532282255	10	24
19	R	<i>management</i>	18.21448099481704	9	20
20	L	<i>partnership</i>	18.149538004526335	8	16
21	L	<i>ensure</i>	17.63222268554455	14	50
22	R	<i>growth</i>	17.581530842337962	8	17
23	L	<i>17</i>	16.11819114467975	5	8
24	L	<i>on</i>	16.06630367433059	22	139
25	R	<i>be</i>	14.528362524069134	14	71
26	L	<i>reliable</i>	14.348469930506337	5	10
27	L	<i>are</i>	13.755662008009901	15	89
28	R	<i>use</i>	13.697512799707154	7	21
29	L	<i>peace</i>	13.648273605574431	5	11
30	R	<i>energy</i>	13.504511491324648	6	16
31	R	<i>we</i>	13.41669301802675	21	173

32	R	<i>targets</i>	12.368517491014027	9	41
33	R	<i>that</i>	11.943959063418738	13	87
34	L	<i>affordable</i>	11.576511402422998	5	15
35	R	<i>recognize</i>	11.378989863903305	6	22
36	R	<i>all</i>	11.216745975701407	18	175
37	R	<i>economic</i>	11.15961536370338	10	60
38	R	<i>including</i>	11.09083513266459	14	113
39	L	<i>technology</i>	11.070302998740294	8	40
40	L	<i>policies</i>	10.886938938564091	7	32
41	R	<i>is</i>	10.818469353869139	9	52
42	L	<i>enhance</i>	10.491746228524692	5	18
43	L	<i>2030</i>	10.118256739567357	10	71

**Table 48.** List of collocates of *SUSTAINABLE* in the 2030 Agenda Corpus (English) – Z(10.0), 5L-5R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>fejldési</i> 'development'	54.88107014617467	43	44
2	R	<i>fejldés</i> 'development'	52.387252279463304	46	55
3	L	<i>a</i> 'the'	22.56073835667597	118	1414
4	R	<i>fejldéshez</i> 'for development'	21.901588446496067	9	12
5	R	<i>használatának</i> 'use'	15.91638656489898	5	7

**Table 49.** List of collocates of *FENNTARTHATÓ* 'sustainable' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>fejldési</i> 'development'	57.4696145711861	45	44
2	R	<i>fejldés</i> 'development'	54.70251682087702	48	55
3	L	<i>a</i> 'the'	31.693187339714807	158	1414
4	R	<i>célok</i> 'goals'	27.769648842246077	21	40
5	R	<i>elérését</i>	24.065655633504843	9	10

6	L	'achieve' és	24.014079718562947	103	1017
7	R	'and' használatának	22.40621108778398	7	7
8	R	'use' fejlődéshez	21.901588446496067	9	12
9	R	'for development' támogatása	19.13297352976251	11	23
10	R	'support' fogyasztás	18.936704235722406	5	5
11	R	'consumption' termelés	18.936704235722406	5	5
12	L	'production' befogadó	17.342829193056314	10	23
13	R	'inclusive' használat	17.239213546220835	5	6
14	L	'use' tartós	16.944903134833755	7	12
15	R	'sustained' érdekében	16.841101202866323	15	53
16	L	'for' megbízható	13.959280970795392	5	9
17	R	'reliable' növekedés	12.556404412689883	5	11
18	R	'growth' foglalkozó	11.485601350967713	5	13
19	L	'-' megfizethető	10.632363528410586	5	15
		'affordable'			

**Table 50.** List of collocates of *FENNTARTHATÓ* 'sustainable' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>fejlődés</i> 'development'	59.33304590370445	52	55
2	R	<i>fejlődési</i> 'development'	58.76388678369182	46	44
3	L	<i>a</i> 'the'	48.81652918291262	233	1414
4	L	és 'and'	35.85932800093603	147	1017
5	R	<i>célok</i>	27.769648842246077	21	40

6	R	'goals' <i>erdőgazdálkodás</i> 'forest management'	25.522735729041475	6	4
7	R	<i>érdekében</i> 'for'	25.096006089502357	22	53
8	L	<i>tartós</i> 'sustained'	24.37993110232722	10	12
9	R	<i>elérését</i> 'achieve'	24.065655633504843	9	10
10	M	<i>nélkülözhetetlen</i> 'essential'	22.776136122864184	6	5
11	R	<i>termelés</i> 'production'	22.776136122864184	6	5
12	R	<i>használatának</i> 'use'	22.40621108778398	7	7
13	R	<i>fejlődéshez</i> 'for development'	21.901588446496067	9	12
14	L	<i>befogadó</i> 'inclusive'	20.923117866468704	12	23
15	R	<i>támogatása</i> 'support'	20.923117866468704	12	23
16	L	<i>megőrzése</i> 'conserve'	20.744119133485658	6	6
17	R	<i>fogyasztás</i> 'consumption'	18.936704235722406	5	5
18	R	<i>fogyasztási</i> 'consumption'	18.936704235722406	5	5
19	R	<i>módok</i> 'patterns'	18.936704235722406	5	5
20	R	<i>támogató</i> 'enabling'	18.850852283514726	9	16
21	R	<i>cél</i> 'goal'	18.267539710689732	14	40
22	R	<i>nélkül</i> 'without'	17.882571531107345	6	8
23	R	<i>használata</i> 'use'	17.239213546220835	5	6
24	L	<i>megbízható</i> 'reliable'	16.82102462660533	6	9
25	R	<i>való</i> '-'	16.73344396567917	14	47
26	R	<i>foglalkozó</i> '-'	16.247830449536597	7	13
27	L	<i>megfizethető</i> 'affordable'	15.065757595563664	7	15

28	L	<i>béke</i> 'peace'	14.4665604790026	6	12
29	L	<i>elhatározásunk</i> 'we are determined'	13.206102340297585	5	10
30	L	<i>partnerség</i> 'partnership'	13.206102340297585	5	10
31	R	<i>termelési</i> 'production'	13.206102340297585	5	10
32	R	<i>szakpolitikák</i> 'policies'	12.917101205023252	7	20
33	R	<i>növekedés</i> 'growth'	12.556404412689883	5	11
34	L	<i>beleértve</i> 'including'	11.786874141550463	13	76
35	L	<i>globális</i> 'global'	11.786874141550463	13	76
36	M	<i>fenntartható</i> 'sustainable'	11.195637308805308	20	181
37	R	<i>az</i> 'the'	10.526464245669311	33	471
38	R	<i>gazdasági</i> 'economic'	10.492017909751347	10	57

**Table 51.** List of collocates of *FENNTARTHATÓ* 'sustainable' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 5L-5R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>sviluppo</i> 'development'	59.3367329528843	106	290
2	L	<i>gestione</i> 'management'	15.75583163945254	7	18
3	R	<i>obiettivo</i> 'goal'	14.323133060788823	9	35
4	R	<i>e</i> 'and'	11.026527245701311	49	1111

**Table 52.** List of collocates of *SOSTENIBILE* 'sustainable' in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>sviluppo</i>	61.641351008409444	110	290

		'development'			
2	L	<i>lo</i>	49.217237066886426	53	107
		'the'			
3	L	<i>uno</i>	29.507744768349905	14	21
		'a'			
4	L	<i>dello</i>	26.419171784626318	14	26
		'of (the)'			
5	L	<i>obiettivi</i>	24.997122678917535	23	76
		'goals'			
6	L	<i>per</i>	24.517975617810244	51	358
		'for'			
7	R	<i>e</i>	21.91790099837037	86	1111
		'and'			
8	L	<i>sullo</i>	19.778104658030315	5	6
		'on (the)'			
9	L	<i>inclusiva</i>	18.681176169246974	7	13
		'inclusive'			
10	L	<i>gestione</i>	18.06843795266137	8	18
		'management'			
11	R	<i>obiettivo</i>	17.640044961367522	11	35
		'goal'			
12	L	<i>produzione</i>	17.338596172174043	7	15
		'production'			
13	R	<i>oceani</i>	16.046834214344806	5	9
		'oceans'			
14	L	<i>di</i>	13.31822616720439	45	751
		'of; to'			
15	L	<i>promuovere</i>	13.317488923816653	9	40
		'promote; foster'			

**Table 53.** List of collocates of *SOSTENIBILE* 'sustainable' in the 2030 Agenda Corpus (Italian) – Z(10.0), 3L-3R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>sviluppo</i> 'development'	64.52212357781586	115	290
2	L	<i>lo</i> 'the'	53.95983254786429	58	107
3	L	<i>obiettivi</i> 'goals'	38.50267711073262	35	76
4	R	<i>e</i> 'and'	32.51491329826459	122	1111
5	L	<i>per</i>	32.296329776085095	66	358

		'for'			
6	L	<i>uno</i>	31.648802175979537	15	21
		'a'			
7	L	<i>dello</i>	28.343376569592756	15	26
		'of (the)'			
8	L	<i>degli</i>	27.83802328159356	21	52
		'of (the)'			
9	L	<i>raggiungimento</i>	25.841139315944535	10	14
		'achievement'			
10	L	<i>duratura</i>	24.49919578803268	8	10
		'sustained'			
11	L	<i>di</i>	23.70106222541834	74	751
		'of'			
12	R	<i>obiettivo</i>	22.615412812235565	14	35
		'goal'			
13	L	<i>produzione</i>	22.40526274966738	9	15
		'production'			
14	L	<i>consumo</i>	22.305758402120308	8	12
		'consumption'			
15	L	<i>inclusiva</i>	21.402412763609693	8	13
		'inclusive'			
16	L	<i>sullo</i>	19.778104658030315	5	6
		'on 'the''			
17	L	<i>gestione</i>	18.06843795266137	8	18
		'management'			
18	L	<i>affidabili</i>	17.056272950271612	5	8
		'reliable'			
19	R	<i>la</i>	16.68150901907457	38	397
		'the'			
20	L	<i>promuovere</i>	16.42017592153752	11	40
		'promote; foster'			
21	L	<i>infrastrutture</i>	16.046834214344806	5	9
		'infrastructure'			
22	R	<i>oceani</i>	16.046834214344806	5	9
		'oceans'			
23	L	<i>economica</i>	15.87801564226771	8	23
		'economic'			
24	R	<i>che</i>	14.083763750305833	21	178
		'that'			
25	L	<i>gli</i>	13.857792574433255	15	98
		'the'			
26	L	<i>crescita</i>	13.832164424764889	7	23
		'growth'			
27	R	<i>il</i>	13.278232410603675	31	395
		'the'			
28	L	<i>sistemi</i>	13.238702980521541	5	13

29	M	'systems' <i>garantire</i>	12.664677110499474	8	35
30	R	'provide' <i>risorse</i>	12.18173273530301	9	47
31	R	'resources; sources' <i>a</i>	11.219261520701627	19	215
32	R	'to; in; on' <i>un</i>	11.040302582960951	14	127
33	R	'a' <i>tutti</i>	10.628520074528547	12	102
34	R	'all' <i>traguardi</i>	10.585947668731874	6	28
35	L	'targets' <i>una</i>	10.341577288864693	11	91
36	R	'a' <i>ad</i>	10.222637680522437	9	64
		'to; in; on'			

**Table 54.** List of collocates of *SOSTENIBILE* 'sustainable' in the 2030 Agenda Corpus (Italian) – Z(10.0), 5L-5R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>goals</i>	48.322610902852325	33	64
2	L	<i>for</i>	19.57217328860884	26	220
3	R	<i>goal</i>	10.735852070292614	6	40

**Table 55.** List of collocates of *SUSTAINABLE DEVELOPMENT* in the 2030 Agenda Corpus (English) – Z(10.0), 1L-1R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>goals</i>	51.29207353545156	35	64
2	L	<i>for</i>	25.978609439611162	34	220
3	L	<i>partnership</i>	23.418939506481944	8	16
4	L	<i>the</i>	16.8605469816976	47	837
5	L	<i>innovation</i>	16.55944604067454	7	24
6	R	<i>goal</i>	14.49195819965948	8	40
7	L	<i>on</i>	13.111936653325706	14	139
8	L	<i>global</i>	12.705468680947284	10	78
9	R	<i>targets</i>	12.445979156590916	7	41
10	L	<i>agenda</i>	12.169989204059107	9	69
11	R	<i>we</i>	11.53543680189534	14	173

12	R	<i>be</i>	10.567731925243462	8	71
13	R	<i>and</i>	10.547624055135394	40	1237

**Table 56.** List of collocates of *SUSTAINABLE DEVELOPMENT* in the 2030 Agenda Corpus (English) Z(10.0), 3L-3R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>goals</i>	51.29207353545156	35	64
2	L	<i>for</i>	33.986654628364064	44	220
3	L	<i>the</i>	32.46177282775085	85	837
4	L	<i>peace</i>	24.789918904706415	7	11
5	L	<i>achieving</i>	24.039337646485393	5	6
6	L	<i>partnership</i>	23.418939506481944	8	16
7	L	<i>and</i>	21.69228719800179	73	1237
8	L	<i>17</i>	20.759146222359735	5	8
9	R	<i>targets</i>	19.86601356498455	11	41
10	L	<i>of</i>	18.224087862612926	40	553
11	R	<i>be</i>	17.61593951472726	13	71
12	L	<i>innovation</i>	16.55944604067454	7	24
13	R	<i>goal</i>	16.370011264342914	9	40
14	R	<i>we</i>	16.050718063216944	19	173
15	L	<i>on</i>	15.12686891777301	16	139
16	L	<i>to</i>	14.908246902279881	29	428
17	R	<i>that</i>	14.495995923953355	12	87
18	R	<i>policies</i>	14.221837976765562	7	32
19	L	<i>agenda</i>	13.599914370765976	10	69
20	M	<i>poverty</i>	13.022723195103213	6	28
21	L	<i>challenges</i>	12.903330290685163	5	20
22	L	<i>an</i>	12.780590089172495	6	29
23	L	<i>global</i>	12.705468680947284	10	78
24	L	<i>technology</i>	12.613905134976047	7	40
25	M	<i>a</i>	12.327211357426066	12	116
26	R	<i>recognize</i>	12.266943521204478	5	22
27	M	<i>are</i>	11.796246554325684	10	89
28	R	<i>in</i>	11.452030456222255	19	305
29	R	<i>will</i>	10.847171091452832	11	123
30	R	<i>relevant</i>	10.77802044845303	5	28
31	R	<i>including</i>	10.278786555364404	10	113

**Table 57.** List of collocates of *SUSTAINABLE DEVELOPMENT* in the 2030 Agenda Corpus (English) – Z(10.0), 5L-5R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>célok</i> 'goals'	32.81574238244024	18	40
2	L	<i>a</i> 'the'	22.660339044486086	83	1414
3	R	<i>érdekében</i> 'for'	13.867477820209992	9	53

**Table 58.** List of collocates of *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 1L-1R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	R	<i>célok</i> 'goals'	34.668819347747935	19	40
2	L	<i>a</i> 'the'	30.452156272032692	108	1414
3	R	<i>elérését</i> 'achieve'	29.379409948373965	8	10
4	R	<i>támogató</i> 'enabling'	20.16850341856978	7	16
5	R	<i>érdekében</i> 'for'	18.6970290048097	12	53
6	L	<i>béke</i> 'peace'	16.620626271217855	5	12
7	R	<i>foglalkozó</i> '-'	15.944916288633687	5	13
8	R	<i>szakpolitikák</i> 'policies'	12.721647662457025	5	20
9	L	<i>és</i> 'and'	10.876623216198178	37	1017
10	R	<i>globális</i> 'global'	10.01106009109435	8	76

**Table 59.** List of collocates of *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 3L-3R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>a</i> 'the'	46.34746341622777	159	1414
2	R	<i>célok</i> 'goals'	34.668819347747935	19	40

3	R	<i>elérését</i> 'achieve'	29.379409948373965	8	10
4	R	<i>megújított</i> 'revitalized'	26.015673154673383	5	5
5	R	<i>nélkülözhetetlen</i> 'essential'	26.015673154673383	5	5
6	R	<i>nélkül</i> 'without'	24.620300688572307	6	8
7	R	<i>támogató</i> 'enabling'	23.098475363552346	8	16
8	L	<i>és</i> 'and'	21.534261443932728	66	1017
9	L	<i>béke</i> 'peace'	20.003866453525326	6	12
10	R	<i>foglalkozó</i> '-'	19.195428314266177	6	13
11	M	<i>szegénység</i> 'poverty'	19.195428314266177	6	13
12	R	<i>érdekében</i> 'for'	18.6970290048097	12	53
13	L	<i>partnerség</i> 'partnership'	18.260948156527814	5	10
14	L	<i>globális</i> 'global'	16.732876730877678	13	76
15	R	<i>szakpolitikák</i> 'policies'	15.342294238916342	6	20
16	L	<i>beleértve</i> 'including'	12.699786747007682	10	76
17	L	<i>is</i> 'too'	10.435686891218706	6	41

**Table 60.** List of collocates of *FENNTARTHATÓ FEJLŐDÉS* 'sustainable development' in the 2030 Agenda Corpus (Hungarian) – Z(10.0), 5L-5R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>lo</i> 'the'	58.46489562922671	48	107
2	L	<i>dello</i> 'of (the)'	34.67288169625768	14	26
3	L	<i>uno</i> 'a'	33.09089559105246	12	21
4	L	<i>sullo</i> 'on (the)'	25.88263935129955	5	6
5	R	<i>obiectivo</i>	10.332684255346653	5	35

'goal'

**Table 61.** List of collocates of *SVILUPPO SOSTENIBILE* 'sustainable development' in the 2030 Agenda Corpus (Italian) – Z(10.0), 1L-1R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>lo</i> 'the'	62.16956343146102	51	107
2	L	<i>obiettivi</i> 'goals'	49.13618276385647	34	76
3	L	<i>uno</i> 'a'	38.66583622535605	14	21
4	L	<i>dello</i> 'of (the)'	34.67288169625768	14	26
5	L	<i>per</i> 'for'	32.949688361641314	51	358
6	L	<i>degli</i> 'of (the)'	26.006574266842033	15	52
7	L	<i>sullo</i> 'on (the)'	25.88263935129955	5	6
8	R	<i>obiettivo</i> 'goal'	12.49184949640523	6	35
9	L	<i>di</i> 'of; to'	12.304433757768068	31	751
10	R	<i>traguardi</i> 'targets'	11.655854035064808	5	28

**Table 62.** List of collocates of *SVILUPPO SOSTENIBILE* 'sustainable development' in the 2030 Agenda Corpus (Italian) – Z(10.0), 3L-3R, NC5-C5.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>lo</i> 'the'	68.3440097685182	56	107
2	L	<i>obiettivi</i> 'goals'	50.601437410692284	35	76
3	L	<i>uno</i> 'a'	38.66583622535605	14	21
4	L	<i>raggiungimento</i> 'achievement'	37.26042381446698	11	14
5	L	<i>dello</i> 'of (the)'	37.178028828201455	15	26
6	L	<i>per</i>	35.65015230836706	55	358

7	L	'for' <i>degli</i>	33.092200248472516	19	52
8	L	'of (the)' <i>sullo</i>	25.88263935129955	5	6
9	L	'on (the)' <i>pace</i>	23.989009330000513	6	10
10	R	'peace' <i>17</i>	22.359666324513377	5	8
11	L	'17' <i>di</i>	21.62688097269226	51	751
12	M	'of; to' <i>il</i>	17.725690628356126	30	395
13	R	'the' <i>e</i>	15.019338311916407	46	1111
14	R	'and' <i>traguardi</i>	14.069874191719151	6	28
15	L	'targets' <i>gli</i>	13.418843716739557	11	98
16	R	'the' <i>che</i>	13.317089101936261	15	178
17	L	'that' <i>raggiungere</i>	13.24971956971085	5	22
18	R	'achieve' <i>obiettivo</i>	12.49184949640523	6	35
19	R	'goal' <i>riconosciamo</i>	12.38236757157525	5	25
20	R	'(we) recognise, (we) acknowledge' <i>un</i>	11.586169845979892	11	127
21	L	'a' <i>globale</i>	10.644057193824965	7	63
22	R	'global' <i>a</i>	10.17726420282939	13	215
		'to, in, on'			

**Table 63.** List of collocates of *SVILUPPO SOSTENIBILE* 'sustainable development' in the 2030 Agenda Corpus (Italian) – Z(10.0), 5L-5R, NC5-C5.

### A3. Analysis of the Sustainable development Corpus

Rank	Position	Collocate	Statistical value	Frequency (collocation)	Frequency (corpus)
1	R	<i>development</i>	341.8433362120768	647	1608

2	L	<i>un's</i>	76.34287855227241	38	111
3	L	<i>un</i>	21.533613104759393	24	512
4	L	<i>the</i>	17.551262517416003	189	25762
5	L	<i>for</i>	17.534124259027177	66	4605
6	L	<i>and</i>	15.84044511590228	110	12378

**Table 64.** List of the collocates of *SUSTAINABLE* in the SusCorp (English) – Z(10.0), 1L-1R, NC20-C20.

Rank	Position	Collocate	Statistical value	Frequency (collocation)	Frequency (corpus)
1	R	<i>development</i>	344.49963954560985	652	1608
2	R	<i>goals</i>	252.9663659896769	284	567
3	L	<i>un's</i>	86.45306660425643	43	111
4	R	<i>sdgs</i>	69.04401275775213	49	224
5	L	<i>the</i>	55.11314351086744	472	25762
6	R	<i>goal</i>	52.98877131329708	33	172
7	L	<i>17</i>	48.79177466877682	20	75
8	L	<i>agenda</i>	44.19877792625355	26	153
9	L	<i>and</i>	42.26480842372848	248	12378
10	R	<i>investing</i>	40.57046302062235	22	130
11	L	<i>for</i>	38.2536496653198	132	4605
12	R	<i>capital</i>	33.30625194863819	20	158
13	L	<i>2030</i>	29.606866054920754	21	218
14	L	<i>of</i>	28.581457258882505	177	12434
15	L	<i>un</i>	27.182552932199723	30	512
16	L	<i>to</i>	25.56883219367451	169	13473
17	L	<i>on</i>	21.734124334065054	66	3311
18	L	<i>a</i>	19.457484770607877	103	8509
19	L	<i>business</i>	18.881561144424097	20	459
20	L	<i>in</i>	16.683031798691115	96	9298
21	R	<i>at</i>	14.852338554345328	33	1748
22	L	<i>new</i>	14.807287971814944	23	912
23	L	<i>more</i>	13.664382478265567	28	1486
24	L	<i>global</i>	13.336464220885201	22	1000

**Table 65.** List of the collocates of *SUSTAINABLE* in the SusCorp (English) – Z(10.0), 3L-3R, NC20-C20.

Rank	Position	Collocate	Statistical value	Frequency (collocation)	Frequency (corpus)
1	R	<i>development</i>	353.53107087962223	669	1608
2	R	<i>goals</i>	258.33434006247836	290	567
3	L	<i>un's</i>	96.56325465624046	48	111
4	L	<i>the</i>	89.3568371374062	730	25762
5	R	<i>sdgs</i>	71.890810978295	51	224

6	L	<i>and</i>	62.56178313843555	354	12378
7	R	<i>goal</i>	61.11064896558535	38	172
8	L	<i>for</i>	54.57812422785338	184	4605
9	R	<i>adopted</i>	54.186023511379574	20	61
10	L	<i>17</i>	53.71160395271106	22	75
11	L	<i>2030</i>	49.806823926452395	35	218
12	L	<i>to</i>	49.42836206815391	299	13473
13	L	<i>agenda</i>	47.64334731288838	28	153
14	L	<i>of</i>	45.393788001171885	265	12434
15	L	<i>achieve</i>	42.607207866924426	23	129
16	R	<i>investing</i>	42.43890252080234	23	130
17	L	<i>a</i>	41.166457651990314	197	8509
18	R	<i>capital</i>	38.390695009768606	23	158
19	L	<i>un</i>	37.53894261584033	41	512
20	L	<i>on</i>	32.47078168171554	95	3311
21	L	<i>in</i>	31.70631403484191	164	9298
22	L	<i>global</i>	26.809972083698746	42	1000
23	R	<i>is</i>	24.86056795554614	96	5238
24	R	<i>growth</i>	24.044026171196904	20	294
25	L	<i>new</i>	23.97786397243209	36	912
26	L	<i>business</i>	23.853369029307533	25	459
27	R	<i>by</i>	22.408403392100375	58	2492
28	R	<i>water</i>	21.633860221014228	23	468
29	R	<i>at</i>	21.4763843116498	46	1748
30	L	<i>will</i>	21.24533997946401	44	1640
31	L	<i>world</i>	20.674251100018076	34	1064
32	R	<i>which</i>	19.59526821684515	35	1232
33	R	<i>investment</i>	19.089357843906274	20	450
34	L	<i>be</i>	18.75403091491904	48	2363
35	L	<i>more</i>	17.53285323749759	35	1486
36	R	<i>that</i>	17.429517168509726	67	4768
37	L	<i>as</i>	17.223630762082642	49	2805
38	R	<i>said</i>	17.1131069050095	34	1467
39	R	<i>are</i>	16.365953196626105	47	2819
40	R	<i>we</i>	15.37967848315995	37	2030
41	L	<i>with</i>	14.607123950970655	40	2523
42	R	<i>it</i>	14.31704335656413	41	2715
43	R	<i>says</i>	13.878729231893338	21	860
44	R	<i>not</i>	12.11899801012774	26	1570
45	M	<i>its</i>	11.523994054751535	20	1065
46	R	<i>this</i>	10.565717360139828	25	1799
47	R	<i>has</i>	10.440410887404346	26	1958

**Table 66.** List of the collocates of *SUSTAINABLE* in the SusCorp (English) – Z(10.0), 5L-5R, NC20-C20.

Rank	Position	Collocate	Statistical value	Frequency (collocation)	Frequency (corpus)
1	R	<i>fejlődés</i> 'development'	246.15209499143336	303	362
2	R	<i>fejlődési</i> 'development'	184.03335326045374	157	174
3	R	<i>fejlődést</i> 'development'	89.34328474039512	47	66
4	L	<i>országgyűlés</i> 'parliament'	75.65827174325415	58	139
5	L	<i>a</i> 'the'	42.846541103857696	542	25176
6	L	<i>nemzeti</i> 'national'	41.54694335907079	46	281
7	L	<i>ENSZ</i> 'UN'	28.683725042227724	32	279

**Table 67.** List of collocates of *FENNTARHATÓ* 'sustainable' in the SusCorp (Hungarian) – Z(10.0), 1L-1R, NC20-C20.

Rank	Position	Collocate	Statistical value	Frequency (collocation)	Frequency (corpus)
1	R	<i>fejlődés</i> 'development'	256.76562734501385	316	362
2	R	<i>fejlődési</i> 'development'	191.09892807840018	163	174
3	R	<i>fejlődést</i> 'development'	98.90352771912389	52	66
4	L	<i>országgyűlés</i> 'parliament'	84.88104481002108	65	139
5	R	<i>bizottsága</i> 'its commission'	82.06952879039525	34	41
6	R	<i>célok</i> 'goals'	80.80383579742625	50	91
7	L	<i>a</i> 'the'	78.38381746941786	905	25176
8	R	<i>bizottságának</i> 'to its commission'	77.15275706856097	40	64
9	R	<i>tanács</i> 'council'	56.257719011534924	37	102
10	L	<i>ENSZ</i> 'UN'	47.28311813351612	52	279
11	L	<i>és</i> 'and'	45.04533991071421	248	5935

12	L	<i>nemzeti</i> 'national'	42.473597588071286	47	281
13	R	<i>szempontjából</i> 'from its point of view'	33.939624518835615	20	81
14	L	<i>környezeti</i> 'environmental'	27.29829804785634	20	123
15	R	<i>fejlesztési</i> 'developmental'	26.411201833957065	21	144
16	L	<i>az</i> 'the'	26.354523374797107	215	10310
17	R	<i>gazdasági</i> 'economic'	22.91310673388787	33	446
18	R	<i>érdekében</i> 'for'	21.838183695281153	20	187
19	L	<i>van</i> '(he/she/it) is'	10.789076628166924	21	684
20	L	<i>hogy</i> 'that'	10.60032000644756	61	4133

**Table 68.** List of collocates of *FENNTARHATÓ* 'sustainable' in the SusCorp (Hungarian) – Z(10.0), 3L-3R, NC20-C20.

Rank	Position	Collocate	Statistical value	Frequency (collocation)	Frequency (corpus)
1	R	<i>fejlődés</i> 'development'	261.6641807389741	322	362
2	R	<i>fejlesztési</i> 'development'	196.98690709335554	168	174
3	L	<i>a</i> 'the'	122.53619113571989	1356	25176
4	R	<i>fejlődést</i> 'development'	104.63967350636115	55	66
5	R	<i>célok</i> 'goals'	95.45906697419737	59	91
6	L	<i>országgyűlés</i> 'parliament'	86.1985838195592	66	139
7	R	<i>bizottsága</i> 'its commission'	82.06952879039525	34	41
8	R	<i>bizottságának</i> '(to) its commission'	77.15275706856097	40	64
9	L	<i>és</i> 'and'	65.20858912037885	348	5935
10	L	<i>ENSZ</i>	63.09260226111126	69	279

		'UN'			
11	R	<i>tanács</i> 'council'	59.333821306026195	39	102
12	L	<i>az</i> 'the'	58.48083309582437	425	10310
13	L	<i>nemzeti</i> 'national'	48.96017719107473	54	281
14	R	<i>ülésén</i> 'on its sitting'	48.60638947450696	23	53
15	R	<i>szempontjából</i> 'from its point of view'	33.939624518835615	20	81
16	L	<i>hosszú</i> 'long'	33.01911317585402	22	103
17	L	<i>környezeti</i> 'environmental'	32.9007522212457	24	123
18	R	<i>gazdasági</i> 'economic'	31.00399373632694	44	446
19	R	<i>érdekében</i> 'for'	29.789666615625716	27	187
20	R	<i>szóló</i> 'about'	29.084784466585223	20	109
21	R	<i>fejlesztési</i> 'developmental'	29.000127829223995	23	144
22	R	<i>elnöke</i> 'its president'	28.885407251991897	31	259
23	L	<i>által</i> 'through'	25.475895568837377	26	233
24	R	<i>is</i> 'also'	23.47617931131179	99	3214
25	L	<i>hogy</i> 'that'	21.47335262041316	106	4133
26	R	<i>fontos</i> 'important'	19.90336214050641	22	266
27	L	<i>Magyarország</i> 'Hungary'	18.499895641648376	27	446
28	L	<i>világ</i> 'world'	18.02427099294999	21	291
29	L	<i>amelyek</i> 'that'	17.960838622487373	20	267
30	L	<i>kell</i> 'have to'	17.530139662580243	36	832
31	L	<i>valamint</i> 'as well as'	17.03453888521841	22	351
32	L	<i>van</i> '(he/she/it) is'	16.72848014384159	31	684
33	L	<i>el</i>	15.526716515389667	26	561

34	L	'-' <i>nem</i> 'not'	14.483988521449866	54	2282
35	L	<i>olyan</i> 'such'	14.085139186124772	23	527
36	L	<i>szerint</i> 'according to'	13.908937771369901	32	975
37	L	<i>európai</i> 'European'	13.618745166736032	23	557
38	R	<i>fenntartható</i> 'sustainable'	12.703118496064182	27	830
39	R	<i>ki</i> '-'	11.690929934369757	21	604
40	L	<i>egy</i> 'a'	11.274890495994518	30	1192

**Table 69.** List of collocates of *FENNTARHATÓ* 'sustainable' in the SusCorp (Hungarian) – Z(10.0), 5L-5R, NC20-C20.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>sviluppo</i> 'development'	316.6279331688702	673	1337
2	R	<i>e</i> 'and'	23.5474594365521	167	9772
3	L	<i>più</i> 'more'	10.273367468004299	29	1594

**Table 70.** List of collocates of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian) – Z(10.0), 1L-1R, NC20-C20.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>sviluppo</i> 'development'	331.7834458508119	705	1337
2	L	<i>lo</i> 'the'	113.08450003147553	204	946
3	L	<i>dello</i> 'of (the)'	97.65800755530441	97	290
4	L	<i>obiettivi</i> 'goals'	83.74977418769822	94	368
5	L	<i>uno</i> 'a'	60.48504834904627	70	387
6	L	<i>per</i>	57.561139837734984	252	5011

7	L	'for' <i>equo</i> 'equitable'	55.09991011744805	20	39
8	L	<i>sullo</i> 'on (the)'	51.490678079634876	24	64
9	R	<i>e</i> 'and'	51.051359470555525	324	9772
10	L	<i>allo</i> 'to (the)'	45.769140398293494	39	210
11	L	<i>mobilità</i> 'mobility'	41.631210098468465	24	97
12	L	<i>di</i> 'of'	41.03398438285766	310	12745
13	R	2030 '2030'	19.996237541340754	21	300
14	R	<i>il</i> 'the'	18.507782483413923	109	6617
15	R	<i>in</i> 'in'	18.29186765976693	87	4602
16	R	<i>che</i> 'that'	15.982948614615601	84	5211
17	L	<i>più</i> 'more'	15.044647271009135	40	1594
18	L	<i>un</i> 'a'	15.001798955100274	62	3414
19	R	<i>la</i> 'the'	14.394525935520475	82	5727
20	R	<i>è</i> '(he/she/it) is'	12.01490336790594	51	3315
21	R	<i>con</i> 'with'	11.00840295640959	37	2185
22	R	<i>a</i> 'to'	10.461699377011513	54	4306
23	R	<i>non</i> 'not'	10.228448796895266	32	1892

**Table 71.** List of collocates of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian) – Z(10.0), 3L-3R, NC20-C20.

Rank	Position	Collocate	Z-value	Frequency (collocation)	Frequency (corpus)
1	L	<i>sviluppo</i> 'development'	336.045933792608	714	1337
2	L	<i>lo</i> 'the'	121.53012992072979	219	946

3	L	<i>dello</i> 'of (the)'	100.70877150801307	100	290
4	L	<i>obiettivi</i> 'goals'	96.38812725584586	108	368
5	L	<i>per</i> 'for'	78.8446575696533	339	5011
6	L	<i>uno</i> 'a'	76.33045459074246	88	387
7	R	<i>e</i> 'and'	70.84716013197199	437	9772
8	L	<i>di</i> 'of'	65.88427388633067	472	12745
9	L	<i>17</i> '17'	59.76251347862947	41	138
10	L	<i>sullo</i> 'on (the)'	57.984757742556944	27	64
11	L	<i>equo</i> 'equitable'	57.87293646534183	21	39
12	R	<i>2030</i> '2030'	48.991277975304826	50	300
13	L	<i>allo</i> 'to (the)'	48.1591887849856	41	210
14	L	<i>mobilità</i> 'mobility'	43.389540573555806	25	97
15	R	<i>il</i> 'the'	40.009698668864196	210	6617
16	L	<i>fondazione</i> 'foundation'	39.08704596328378	33	205
17	L	<i>un</i> 'a'	34.85951987995088	129	3414
18	L	<i>italiana</i> 'Italian'	31.48666955693559	23	153
19	R	<i>in</i> 'in'	31.055757724910386	137	4602
20	R	<i>che</i> 'that'	29.89700679306982	142	5211
21	L	<i>modello</i> 'model'	29.599901741780027	20	131
22	R	<i>Nazioni</i> 'Nations'	29.583779245378533	22	158
23	L	<i>gli</i> 'the'	29.425094543346106	61	1134
24	R	<i>Unite</i> 'United'	28.914868970505317	20	137
25	R	<i>la</i> 'the'	28.35347108707025	143	5727
26	R	<i>a</i>	27.351686522512765	118	4306

		'to'			
27	R	<i>territorio</i>	24.47471645223664	22	226
		'territory'			
28	R	<i>della</i>	24.129200136535406	79	2558
		'of (the)'			
29	L	<i>futuro</i>	23.575645642260003	23	264
		'future'			
30	R	<i>sociale</i>	22.36391372952022	23	291
		'social'			
31	R	<i>è</i>	21.639766682294514	83	3315
		'(he/she/it) is'			
32	R	<i>ha</i>	21.40150164400806	53	1507
		'(he/she/it) has'			
33	R	<i>una</i>	21.35589146041549	64	2129
		'a'			
34	L	<i>più</i>	20.683432492742124	53	1594
		'more'			
35	R	<i>le</i>	20.499442848780337	73	2871
		'the'			
36	R	<i>del</i>	20.350451246368685	84	3714
		'of (the)'			
37	R	<i>si</i>	19.62396679988834	58	2041
		'-'			
38	R	<i>i</i>	18.69333081850072	68	2915
		'the'			
39	L	<i>città</i>	18.29683884629485	20	321
		'city, cities'			
40	R	<i>delle</i>	18.063051070771984	48	1658
		'of (the)'			
41	R	<i>nel</i>	16.95611844031462	44	1568
		'in (the)'			
42	R	<i>con</i>	16.936026592607583	53	2185
		'with'			
43	M	<i>sia</i>	16.792647037248866	20	374
		'(he/she/it) is; both'			
44	R	<i>anche</i>	16.62108648684937	38	1244
		'also'			
45	M	<i>dalla</i>	16.463399948659863	22	463
		'from (the)'			
46	R	<i>su</i>	15.858932013646012	25	626
		'on'			
47	R	<i>non</i>	15.802280196258101	46	1892
		'not'			
48	L	<i>sul</i>	15.442763719995773	22	517
		'on (the)'			
49	R	<i>questo</i>	14.558419627544058	23	620

		'this'			
50	R	<i>alla</i>	14.519772953413874	32	1133
		'to (the)'			
51	R	<i>al</i>	14.51090510170216	42	1834
		'to (the)'			
52	L	<i>dei</i>	14.505602761875862	42	1835
		'of (the)'			
53	R	<i>tra</i>	14.438058170439596	27	842
		'between'			
54	R	<i>da</i>	14.296075577552937	41	1798
		'from'			
55	R	<i>ai</i>	12.909768708575786	20	586
		'to (the)'			
56	R	<i>ma</i>	12.81865704487783	28	1085
		'but'			
57	R	<i>sono</i>	12.535871379019147	32	1419
		'(they) are'			
58	L	<i>degli</i>	12.13100628668531	21	708
		'of (the)'			
59	R	<i>come</i>	10.473743740671214	24	1122
		'like; how'			

**Table 72.** List of collocates of *SOSTENIBILE* 'sustainable' in the SusCorp (Italian) – Z(10.0), 5L-5R, NC20-C20.