Elisabetta Jezek

Semantics first! Rethinking the role of syntax in verb classification

(doi: 10.1418/113926)

Lingue e linguaggio (ISSN 1720-9331)
Fascicolo 1, gennaio-giugno 2024

Ente di afferenza:
Università degli Studi di Pavia (Unipv)

SEMANTICS FIRST! RETHINKING THE ROLE OF SYNTAX IN VERB CLASSIFICATION

ELISABETTA JEZEK

ABSTRACT: In this paper, we address two major weaknesses of syntactically-oriented studies that have been conducted since the 1990s in order to pin down the semantic properties of verbs and group them into classes. These limitations include the challenge of clearly distinguishing arguments from adjuncts, as well as the difficulty of systematically accounting for the optionality in the syntactic realization of arguments. We argue that these weaknesses may be overcome by adopting a semantic-oriented approach in which semantic analysis takes priority and encompasses *semantic flexibility* of word meanings. Specifically, we propose that arguments are complements that play a role in influencing the meaning that the verb acquires in context, whereas adjuncts do not have an impact on the verb semantics. We also propose that whether an argument is syntactically optional or mandatory in a sentence is determined by semantic factors, specifically how much the argument is semantically incorporated into the verb it depends on. Our analysis is based on Italian data but the conclusions are generally applicable for other languages as well.

KEYWORDS: verb meaning, semantic flexibility, co-composition, argument structure, verb class.

1. INTRODUCTION[1]

Syntax-first approaches are an established model in theoretical linguistics for studying verbs at the syntax-semantics interface. These approaches generally start by identifying subcategorization frames and their alternations for each verb and then analyze them to pin down correlations between the syntactic and semantic properties (Levin 1993; Dang *et al.* 1998; Kipper *et al.* 2008). Semantic properties may include thematic roles and selectional preferences of the arguments, as well as decompositions of the verb semantics (Brown *et al.* 2022).

Despite the results that have been achieved, these approaches have shown not to be without problems. Two significant problems regard the inability to

[1] I would like to thank two reviewers for their insightful comments on an earlier version of the paper.

clearly distinguish arguments from adjuncts and the difficulty in accounting for the optionality in the syntactic realization of specific arguments.[2] In this contribution, we propose an approach to studying verbs at the syntax-semantic interface that offers a solution to these problems. The core of the approach is that the analysis of verbs and their arguments starts from their meaning instead of their syntactic behaviour and integrates syntactic analysis only as a secondary step and as a corollary to semantic analysis. Moreover, the approach is based on the principle of *semantic flexibility*, according to which words in combination influence each other meanings.

The paper is structured as follows. After reviewing the primary constituents of the meaning of verbs (§ 2), we critically evaluate the pros and cons of syntactically-oriented studies that have been systematically pursued from the '90 to derive the semantic properties of verbs (§ 3). First, we highlight the strengths of this approach. Second, we address the two significant weaknesses reported above (§ 4) and show that these weaknesses may be overcome by adopting a semantics-oriented approach from the start (§ 5). Specifically, we propose that arguments are complements that influence the meaning that the verb acquires in context, whereas adjuncts are inactive on the verb meaning, and that syntactic optionality (vs. mandatoriness) is licenced by semantic factors, including the degree to which an argument is semantically incorporated in the verb on which it depends.

2. THE MEANING OF VERBS

The meaning of verbs includes at least three dimensions, which, together, express different aspects of the event that the verbs report.[3] These are: the temporal-aspectual structure of the event, the participants of the event (the arguments or valency of the verb), and the ontological type of the event. Together, these components constitute the three main dimensions of analysis that have been pursued in linguistic studies – jointly or individually – to identify verb classes. We review them in the following subsections.[4]

[2] See Mereu (2020) for an overview.

[3] In the following, we use the term *event* to refer to what Emmon Bach (1986, 6) called *eventualities*, i.e., static and dynamic situations. Also, we focus on predicative verbs and do not include copulas or constructions in which the verb is not acting as predicate.

[4] The goal of our paper is to offer a fresh perspective on verb classification and to think anew about the nature and role of verbal arguments; we do not suggest formal means to model the aspects of verb meaning reviewed here.

2.1 The temporal-aspectual structure of the event

The event expressed by the verb is placed in time, and if it is endowed with duration, it generally has an internal structure that can be broken down into phases or intervals. The analysis of these phases is generally implemented through linguistic tests such as the adverbial modification introduced by the expressions 'in *x* time' and 'for *x* time'. The applicability of the first would reveal that the event expressed by the verb includes a telic or culminative component; that of the second would exclude it. Based on this test, it is possible to distinguish verbs expressing *states* (*possess*, *remain*), *processes* (*work*, *sleep*, *walk*; also consisting of repeated acts such as *semelfactives sob*, *blink*, *cough*, *knock*), all admitting 'for x time' but not 'in *x* time'; punctual or culminating events (*find*, *arrive*, *burst*) that accept 'in *x* time' but not 'for *x* time'; incremental or gradual events (degree achievements, incremental theme- or multi-scalar verbs: *increase*, *cool down*), which admit both adverbials. The event receives a specific temporal encoding when it becomes the denotation of a verb (i.e., its meaning). In principle, there may be a mismatch between an event "as it is observed in the world" and the way it is encoded in the meaning of verbs, although this is not the norm.[5]

2.2 The event's participants

The event expressed by the verb typically involves participants, such as animate or inanimate, concrete or abstract entities (including places) that hold a relation of *participation* with the event itself and play different roles: the agent, the experiencer, the patient, the receiver, the origin, the destination, the instrument, and so on. The linguistic construal of the event makes some of these participants grammatically relevant: these are the participants called *arguments* or *actants*, which, contrary to other complements (the so-called *adjuncts* or *circumstantials*, cf. Tesnière 1959), must be mandatorily expressed in the syntax because they complete the meaning of the verb. Depending on the number of arguments, it is possible to distinguish zero-argument (*snow*, *flash*), one-argument (*sleep*), two-argument (*abolish*, *inhabit*, *dwell*), up to three-argument (*give*, *dedicate*) verbs. The existence of verbs with four arguments (*translate*) is controversial (Levin & Rappaport Hovav 2005). When adopting the notion of argument, an interesting question arises: which aspect(s) of the event may become an argument? Typically, participants or properties of participants in the event

[5] Several studies have pointed out that verb meanings represent *construals* of events rather than the events themselves (see, among others, Levin & Rappaport Hovav 2005: 19). For the inherent temporal properties of the event expressed by verbs, often referred to as *Aktionsart* or *Actionality*, see the taxonomies of Vendler (1967) and Dowty (1979).

(including other events) may become arguments (such as the intentionality of the agent, the causes of the event, instruments, and psychological states). Moreover, some participants may be *incorporated* in the verb semantics, whereas others may be *backgrounded*, an issue to which we return later. Besides bearing a role, which is generally assigned to them by the verb, the arguments display semantic properties; in particular, they belong to a semantic class (Physical object, Human, Institution, Animal, Artifact, Abstract entity, etc.). This class represents the *selectional restriction* (Chomsky 1965) or *preference* (Wilks 1975) imposed by the verb: for example, the It. verb *abitare* 'inhabit, live in/at' only admits Animates in Subject position, preferably Humans and less typically Animals, and rules out Institutions, as the example in (1b) shows:

(1) a. *Luca abita a Milano*
'Luca lives in Milan'

b. **L'ONU abita a New York*
'The UN lives in New York'

Finally, arguments require a specific syntactic realization (Subject, Direct Object, Prepositional Complement, etc.) which together constitutes the subcategorization frame for the verb, for example, for *abitare* the subcategorization frame of (1) is <Subj> abita <PP>.

2.3 The ontological type

The ontological type of the event allows one to identify verb classes such as motion verbs (*go*, *walk*), manner verbs (*slip*), verbs of perception (*feel*, *hear*), verbs of cognition (*understand*, *grasp*), and so forth. Several attempts have been made to define a list of verb classes on semantics grounds. For example, in the WordNet Project (Fellbaum 1998), a list of 15 top-level distinctions in the verbal domain has been defined in order to provide the collaborators working on the project with a broad initial classification for lexicon items. The proposed classes, called *supersenses*, are listed in Table 1, where the first column reports the name of the class and the second column a set of English examples and/or descriptions.[6]

The proposed classes in Table 1 raise a number of issues from a linguistic perspective: what is the use of the top-level distinctions? What do they tell

[6] Although ontological or formal-semantic precision is not one of the main characteristics of WordNet – as noted by one reviewer – to our knowledge this is one of the early and few attempts to identify broad classes of verbs primarily on semantic/cognitive grounds. Another similar attempt is FrameNet, which provides more detailed groupings based on the notion of frame (Fillmore & Baker 2010). For classifications based on the analysis of the syntax-semantics interface, see § 3.

Class	EXAMPLES
BODY	grooming, dressing, bodily care
CHANGE	size, temperature change, intensifying
COGNITION	thinking, judging, analysing, doubting
COMMUNICATION	telling, asking, ordering, singing (animal sounds)
COMPETITION	fighting, athletic activities
CONSUMPTION	eating, drinking
CONTACT	touching, hitting, tying, digging

Class	EXAMPLES
CREATION	sewing, baking, painting, performing (cooking)
EMOTION	feeling
MOTION	walking, flying, swimming
PERCEPTION	seeing, hearing, feeling
POSSESSION	buying, selling, owning
SOCIAL	political and social activities and events
STATIVE	being, having, spatial relations
WEATHER	raining, snowing, thundering

TABLE 1: TOP-LEVEL DISTINCTIONS FOR VERBS IN WORDNET.

us about the syntactic behaviour of verbs, the inferences they give rise to, the participants they involve, the manner in which the event takes place, and the purpose of the event? These questions together lead to a broader question: on what grounds is it possible to group verbs in semantic classes? The operation of identifying semantic classes, that is, groupings of verbs with similar meanings and organising them in a hierarchy that distinguishes, for example, between types of motion verbs, types of perception verbs, and so on, is intuitively simple but technically complex (see the overview in Jezek 2016: 128). In general, the difficulties encountered in classifying verbs in semantic terms can be said to derive primarily from the fact that the meaning of verbs consists of a bundle of features with different semantic prominence. Consider, for instance, the case of verbs describing a change of position, like *sit*: although the action of sitting involves the motion of the person performing the action (a motion we may characterize as "internal"), this is not the prominent feature in the meaning of the verb, and it would appear odd to classify *sit* as a verb of motion on a par with *enter* and *exit* (it is, in fact, categorized as a 'verb of assuming a position' in Levin 1993). But how is semantic prominence to be defined? Several scholars contend that prominent features are those that are transparent in the syntax, i.e., those that influence the syntactic behaviour (Levin & Rappaport Hovav 2005: chapter 1); nevertheless, linking an observed syntactic behaviour to the appropriate semantic component is a notoriously arduous task. Another problem is that verbs encode only some aspects of the event they denote while presupposing others. The distinction between denoted

and presupposed information is by no means easy to make, although there are clear cases. *Arrive*, for example, presupposes motion but encodes, in fact, the result/effect of such motion, consisting in the fact that the person or thing arrived is located in a place which differs from the one it was located in before the arriving event took place. On this ground, it is reasonable to question whether *arrive* should be considered a verb of change of location rather than a verb of motion and which criteria can distinguish between the two.

3. TRADITIONAL APPROACHES: FROM SYNTAX TO SEMANTICS

3.1 Syntactic alternations

The traditional classification of verbs is syntax-oriented and rests on the distinction between transitive verbs and intransitive verbs, i.e., about the ability of a verb to be accompanied or not by a direct object. A more refined classification is based on the observation of the syntactic alternations allowed by the verbs, i.e., of the range of syntactic configurations that each verb can present. This may or may not include the number of arguments (cfr. § 2.2). For example, for Italian, some verbs only allow transitive uses (*abolire* 'abolish', *indossare* 'wear', *affittare* 'rent'), others only allow intransitive uses (*camminare* 'walk', *russare* 'snore', *arrivare* 'arrive', *cadere* 'fall'), and others allow both (*affondare* 'sink', *guarire* 'heal').[7] Looking into those alternations, and following the influential work of Levin (1993), Jezek (2003) proposed to classify Italian verbs based on the presence or absence of four different syntactic realizations: transitive use (TR), intransitive use with auxiliary *have* (UNERGATIVE), intransitive use with auxiliary *be* (UNMARKED UNACCUSATIVE), intransitive use with pronominal marker *si* (and *be* as an auxiliary, MARKED UNACCUSATIVES).[8] The result is a list of 15 classes, which exhausts all the possible combinations of these four realizations, as shown in Table 2, which provides an example of a verb for each class.

To illustrate, consider the verb *abolire* 'abolish' in (2), which presents only a transitive use and therefore falls into class 1, and the verb *russare* 'snore' in (3) that exhibits only an unergative use, and falls into class 2.

[7] Examples for these classes are provided below.

[8] The distinction between unergatives and unaccusatives was firstly introduced in syntactic studies by Perlmutter 1978. The existence of two classes of unaccusatives (here called marked and unmarked) was established only later by studies focused on languages which exhibit variation in the realization of the pronominal marker in unaccusative uses (see Jezek 2003, for an overview).

CLASS	VERB	TR	UNERG	UNMARKED UNACC	MARKED UNACC
1	abolire 'abolish'	x	-	-	-
2	russare 'snore'	-	x	-	-
3	cadere 'fall'	-	-	x	-
4	pentirsi 'repent'	-	-	-	x
5	squillare 'ring'	-	x	x	-
6	approfittare 'take advantage'	-	x	-	x
7	ammuffire 'get mouldy'	-	-	x	x
8	sedimentare 'settle'	-	x	x	x
9	mangiare 'eat'	x	x	-	-
10	affondare 'sink'	x	-	x	-
11	alzare 'raise'	x	-	-	x
12	ingiallire 'yellow'	x	-	x	x
13	continuare 'continue'	x	x	x	-
14	chiudere 'close'	x	x	-	x
15	bruciare 'burn'	x	x	x	x

TABLE 2: ITALIAN VERB CLASSES AND SYNTACTIC ALTERNATIONS (JEZEK 2003).

(2) *abolire*

TR
Il governo ha abolito la tassa
the Government has abolished the tax
'The Government abolished the tax'

UNERG
**Il governo ha abolito*
the Government has abolished
*'The Government has abolished'

UNMARKED UNACC
Il governo è abolito
the Government is abolished
*'The Government is abolished'

MARKED UNACC
**Il governo si è abolito*
the Government PRON is abolished
*'The Government is abolished'

(3)	*russare*	TR	**Il raffreddore ha russato Gianni* the cold has snored Gianni *'The cold snored Gianni'
		UNERG	*Gianni ha russato* Gianni has snored 'Gianni snored'
		UNACC	**Gianni è russato* Gianni is snored
		MARKED UNACC	**Gianni si è russato* Gianni PRON is snored

The case of *bruciare* 'to burn' is more complex, as it admits all four uses, as shown in (4) (class 15):

(4)	*bruciare*	TR	*L'incendio ha bruciato la casa* the fire has burned the house 'The fire has burned the house'
		UNERG	*La casa ha bruciato per ore* the house has burned for hours 'The house has burned for hours'
		UNMARKED UNACC	*La casa è bruciata* the house is burned *'The house is burned'
		MARKED UNACC	**La casa si è bruciata* the house PRON is burned *'The house is burned'[9]

The classification performed based on a large sample of Italian verbs (cf. Jezek 2003 for details on the dataset) shows that the semantic properties primarily represented in the syntax of the verbs (particularly, the syntactic alternations that the verbs exhibit) are those related to the *event type*, that is, the aspectual or actional properties (process vs. state vs. change of state, cf. § 2.1), *agency*, and *causation*. For example, TR verbs (that is, verbs of class 1 that only admit

[9] It is well known that the pronominal marker *si* may occur in Italian in several constructions with different semantic values along the parameter of voice (reflexive, middle and passive respectively). In the current study, we consider only intransitive constructions, where *si* marks spontaneity/lack of causalness, event completion, or both. We exclude reflexive constructions in which *si* signals coreference with the subject (*Luca si è vestito* 'Luca got dressed') and passive constructions (*Si vedono le stelle* 'the stars are seen', *Si abolirono le tasse* 'taxes were abolished'). For an overview of pronominal uses in Italian see Jezek (2003: 130-141). On the passive reinterpretation of the *si*+V construction in Italian, see Cennamo (2001).

a transitive use) tend to be lexically agentive, i.e., they rarely occur with unintentional subjects (*abolire* 'abolish').[10] Verbs that alternate between TR and UNERGATIVE (class 9), on the other hand, mainly denote processes (*mangiare* 'eat', *bere* 'drink') rather than states or happenings. Finally, verbs that alternate between TR and UNACCUSATIVE (UNMARKED and MARKED) (classes 10 to 12) predominantly describe changes of state (*ingiallire* 'become yellow').

3.2 Lexical decomposition and syntax-semantics linking

Several works have attempted to predict the syntactic realization of arguments on semantic grounds, on the basis of decompositional analyses of verb meaning. A good part of Levin & Rappaport Hovav's (2005) book is devoted to linking theories that build on decomposed representations built around semantic primitives such as ACT, CAUSE, and BECOME in (5).

(5) a. *dry*: [[x ACT] CAUSE [y BECOME <DRY>]]
b. *jog*: [x ACT <JOG>]

The books and articles of Robert Van Valin (see Van Valin 2005) are also particularly relevant in this respect. Van Valin (2012) clarifies that a fundamental issue dividing theories of the syntax-semantics interface is whether the semantic representation of sentences is projected from the lexical representation of the verb which determines to a large extent the syntactic structure (projectionist view, Levin & Rappaport Hovav, 2005 and Van Valin, 2005) or whether it is constructed based on the NPs or PPs co-occurring with the verb (constructionist approach); the latter approach covers the frameworks which posit the existence of constructional meanings (Goldberg 1995; Michaelis & Ruppenhofer 2001) as well as those which do not and derive the meaning of sentences from the verb plus cooccurring elements (Jackendoff 1997; Pustejovsky 1995). Note that the semantics-first approach we advocate in this paper is not projectionist. The projectionist approach is not semantics-first by definition, because it originates in the syntax; it aims at spelling out why the syntax of a verb (and the sentences it projects) is what is it on semantic grounds and at pinning down semantic properties from syntactic ones, rather than accounting for the variability in the compositional semantic behaviour of verbs and their arguments. By contrast, the approach we argue for is in line with the constructionist one; specifically, we exploit and expand previous studies conducted by 'enriched compositionalists'

[10] Note that most verbs that are generally classified as agentive allow for non-agentive uses in specific contexts. For example the verb *ferire* 'injure' can be used in a unvoluntary sense. This is not the case of TR verbs such as *abolire* 'abolish'.

such as R. Jackendoff and J. Pustejovsky (cf. section 5.1) to account for the two problems introduced in § 1 and examined in § 4 below.

4. PROBLEMS WITH SYNTAX-FIRST APPROACHES

As stated in § 1 and § 2, the syntactic alternation and decompositional models represent refined approaches to classifying verbs with respect to the model based exclusively on transitivity/intransitivity. However, there are at least two connected weaknesses. The first is the well-known difficulty in drawing a distinction between arguments and adjuncts concerning a specific verb. The second is the optionality of some arguments, which questions the very notion of argument, which is grounded on the idea that arguments are complements that require mandatory syntactic realization. The first of these problems is exemplified in (6), where the adverbial 'in front of the hotel' has an uncertain status with respect to the verb *arrive* (argument, thus mandatory, or adjunct, thus not mandatory), while the second is exemplified in (7), in which the direct object of the verb *drink*, expressed in (7a), is absent in (7b):[11]

(6) *Gianni rilesse quel che aveva scritto fin quando il taxi non arrivò davanti all'albergo*
'Gianni reread what he had written until the taxi arrived in front of the hotel'

(7) a. *I ragazzi bevono birra al pub*
'The boys are drinking beer in the pub'
b. *Il vecchio si riposò, bevve e fu assalito da uno strano pensiero*
'The old man rested, drank, and was haunted by a strange thought'

In the following sections, we propose two solutions to these two problems. Our proposals underscore the crucial role of semantics in determining the distinction between arguments and adjuncts and in licensing syntactic optionality.

5. SEMANTICS FIRST

In contrast with the common practice that tackles the problems highlighted in § 4 starting from syntactic analysis, we argue that those problems are better understood if the analysis is performed starting from the semantic level, particularly from the following assumptions: i.) the meanings of the arguments play an active role in the semantic composition with the verb and influence the

[11] The examples in the paper are derived and adapted from the ItTenTen 2020 corpus (about 13 billion words), queried through the Sketch Engine online platform (`https://app.sketchengine.eu`).

verb meaning, whereas those of the adjuncts do not; ii.) arguments may be incorporated at different degrees into the meaning of the verb. According to our proposal, the syntactic level is secondary and dependent on the semantic one. We first review our semantic approach, and then apply it to the problems raised in § 4.

5.1 Enriched composition in semantics

Our proposal departs from the traditional semantic account, according to which polysemy is a list of pre-defined senses stored in each lexical entry, and embraces the contextual approach, in which words have a core meaning that is flexible and varies from occurrence to occurrence as a function of the interaction with the other words they combine with, and of the situation of utterance (Recanati 2012). Accordingly, our approach also departs from the principle of compositionality outlined in formal semantics (Partee 1995: 153), which claims that the meaning of a complex linguistic expression is a function of the meanings of its component parts and the way they are put together. In contrast to these views, we claim that the meanings of words not only add up but also influence each other when they combine: from a theoretical point of view, we move from *composition* seen as an additive operation to the view of *enriched composition* (Jackendoff 1997) and *co-composition* (Pustejovsky 2012), understood as operations of mutual semantic adjustments between the meaning of the combining words. Under this line of reasoning, in the combination of a verb with its argument, the meaning of the verb is co-dependent on the meaning of its arguments, and the variation in the semantics of arguments causes a variation in the verbal meaning.

An example applied to the verb *prendere* 'take' is provided in Figure 1, where we graphically represent the co-composition of *prendere* with two different Direct object arguments, i.e. *pillola* 'pill' on the left side and *treno* 'train' on the right side, both satisfying the verb's selectional restriction (Physical Object) but, crucially, resulting in two different meanings for the verb.[12] The example is based on the assumption that the meanings of nouns, represented by the oval shapes in the Figure, include information the typical actions performed with the denoted entity in their sublexical information.[13] When the verb selects

[12] In the Figure, we focus on the co-composition with the Direct object and, for reasons of space, do not discuss the one with the Subject. Note that *prendere una pillola* 'take a pill' and *prendere il treno* 'take a train' are not light verb constructions (LVC) as LVC require an NP denoting an event that can function as a predicate (as in *prendere una decisione* 'take a decision'). Neither *pill* nor *train* are event nominals, and in both examples, *prendere* acts as a full predicate.

[13] This assumption has been formalised in different ways in semantic studies; see, in particular, the notion of *quale* proposed in Pustejovsky (1995).

its arguments (the Direct object typed as a Physical Entity in our case), these actions (*ingest* for pill and *travel with* for train) are activated and fused with the meaning of the verb. This operation results in two meanings for the verb *prendere*, that are dependent on the meaning of the Direct object argument.

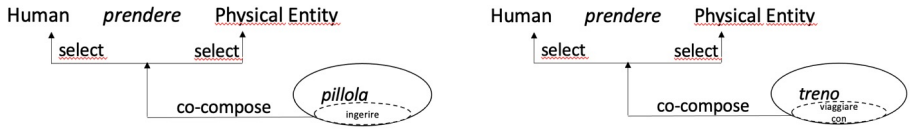

FIGURE 1: CO-COMPOSITION OF *prendere* + DIRECT OBJECT.

Co-composition is a unifying principle that may apply to different syntactic roles (Subjects, Direct objects, etc.) and to a wide range of semantic adjustment phenomena. These include the shift from literal to figurative meanings of verbs, as in (8), where *divorare* 'devour' means 'eat eagerly' in combination with *pasta* 'pasta' and 'read eagerly' in combination with *romanzo* 'novel'.[14]

(8) a. *divorare la pasta*
'devour the pasta' (= eat eagerly)

b. divorare un romanzo
'devour a novel' (= read eagerly)

In (8), the noun *romanzo* 'novel' does not satisfy the selectional restriction of the Direct object of *divorare* (i.e., it is not a type of Food). Nevertheless, *romanzo* (and similar documents: *libro* 'book', *i classici* 'the classics', etc.) chooses *divorare* as a metaphorical 'vehicle' to indicate a typical action performed with it (*leggere* 'read'), and activates a reinterpretation of *divorare* based on its sublexical information. Interestingly, the basic meaning of *divorare* ('consume quickly and eagerly') is maintained but applied to a different type of entity, which, in turn, adjusts it according to its own semantics. In Lakoff & Johnson's terms (2008), one could talk of a mapping from the source domain Food to the target domain Document (i.e., the Document is seen in terms of the Food). In Figure 2, we offer a schematization of this process. On the left side, the resulting meaning of *divorare* is 'eat eagely', on the right side, it is 'read eagerly'.

The choice of the 'vehicle' of the metaphor, that is, of the term (*divorare* in our case) whose literal meaning provides the lens (*screen* in the terminology of Black 1954) through which the action associated with the noun is interpreted, is cognitively motivated a posteriori, but idiosyncratic, in the sense that it is not

[14] In (8) we provide a synonym for each sense that the verb acquires in combination with a different argument.

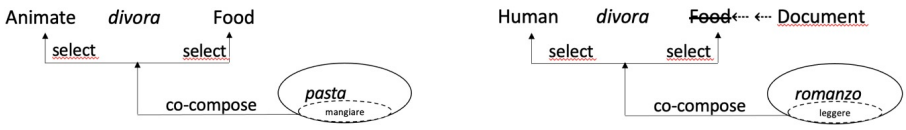

FIGURE 2: CO-COMPOSITION OF *divorare* + DIRECT OBJECT.

predictable. The differences between languages demonstrate this: compare e.g., it. *lanciare un grido* vs. fr. *pousser un cri* 'launch a cry'. These differences must not be surprising as metaphors can rest on different similarities. Beyond the word choices, these similarities are interpretable from a cognitive perspective. For example a non-native speaker who knows the it. word *divorare* in its literal meaning would be able to understand the expression *divorare un romanzo*, even if in his/her language, the same concept is expressed using a different metaphor. According to our proposal, once the 'vehicle' of the metaphor has been chosen (*divorare* 'devour' for *romanzo*, *lanciare* 'launch' for *grido* 'cry', *soffocare* 'suffocate' for *rivolta* 'revolt', etc.), the modulation of the sense of the 'vehicle' in the context is not idiosyncratic but takes place in a predictable way in accordance with the semantics of the noun and the encyclopedic knowledge associated with it.

In essence, if we adopt the view proposed in this section, the perspective of analysis of verbs and their arguments is reversed. Instead of claiming that the number and type of arguments vary according to the meaning of the verb, as in the traditional account, we claim that the verb's meaning varies according to the semantics of the argument(s). In the next sections, we adopt this theoretical background and apply it to the analysis of the two issues raised in § 4.

5.2 The argument-adjunct distinction

In § 5.1, we departed from the traditional view that a word in the semantic composition behaves either as an active functor or as a passive argument. Instead we adopted the idea that arguments can both complete and influence the meaning of the verb they combine with. This move has important consequences on our account of the argument-adjunct distinction. Specifically, we can assume that the ability of arguments to co-compose with the verb does not apply to adjuncts, which are instead inactive on verbal semantics and never play a role in determining the sense that the verb assumes in its context of use: if they do, they are arguments.

According to this line of reasoning, the verb *partire* 'leave' has a single argument in the example in (9a) (the Subject *la macchina* "the car") not for syntactic reasons but because that single argument is sufficient for the verb

to take on the meaning of 'start (off)'. As it happens, that single argument is essential to ensure that the verb acquires this meaning and not another possible meaning it could acquire (e.g. 'go away, move away from a place'). In (9a), the adverbial expressions are instead adjuncts, as they neither complete nor modify the meaning of the verb, which is determined solely by the Subject argument. In example (9b), instead, the verb *partire* takes two arguments, as long as both the Subject (*dolore* 'pain') and the expression indicating the source (*dal braccio* "from the arm" / *al livello del collo* "at the neck level") are necessary to complete the meaning that the verb takes on in the context ('originate from'). In other words, according to our account, the unacceptability of the expression *"the pain starts" is the result of a semantic inconsistency and not a syntactic one. Alternatively, the syntactic inconsistency arises from semantic factors:

(9) a. *La macchina non parte più / mai al primo colpo*. (1 arg)
'The car no longer starts / never starts the first time'

b. *Il dolore parte dal braccio / a livello del collo*. (2 arg)
'The pain starts from the arm / at the neck level'

The assumption that the distinction between arguments and adjuncts originates in the semantic co-composition of the verb and its complements and that only the arguments may act as functors in this process is fully supported by the empirical observation that the valence of the verb often varies according to its meaning. When one attempts to draw the distinction between arguments and adjuncts, what must be observed, therefore, is not the verb but the meaning it acquires in the specific context of use.

We applied the approach described above to identify verb senses and to tell apart arguments from adjuncts in the construction of the T-PAS resource, an inventory of argument structures for Italian verbs focused on the semantics of arguments.[15] In T-PAS, for each verb meaning, a specific Typed Predicate-Argument Structure (informally called pattern) is provided. An example of a pattern for the verb *guidare* 'drive' in its 'operate' sense is [Human] guida [Road Vehicle]. Arguments are defined in terms of semantic classes notated between square brackets, called *semantic types*. Patterns are corpus-derived, i.e., they are acquired through manual clustering and annotation of corpus instances, following the Corpus Pattern Analysis (CPA) methodology (Hanks 2013). Several studies on T-PAS data have shown that verb-argument co-composition is a useful principle for discriminating verb senses and distinguishing arguments

[15] The T-PAS project (Jezek *et al.* 2014) was developed at the Department of Humanities of the University of Pavia, with the technical support of Lexical Computing Ltd. The resource can be freely accessed and downloaded at `https://tpas.unipv.it`.

from adjuncts.[16] Also, T-PAS data allows the definition of semantic classes for verbs based on the semantics of the arguments. This is possible thanks to a function available in the public access of the resource, called "Browse Semantic Types in Patterns", through which the user can search for verbs according to the semantic type(s) of their argument(s). The function allows one to search either for a single syntactic function (i.e., Direct Object: [Food]) or for a combination of functions, i.e., Subject [Human] & Direct Object: [Food]. A list of four classes is reported in Table 3: vehicle_verbs are verbs that take Vehicles as Direct objects, and so forth.

CLASS	EXAMPLES
VEHICLE_VERBS	*assaltare* 'assault', *caricare* 'load', *dirottare* 'hijack', *guidare* 'drive', *parcheggiare* 'park', *prendere* 'take', *rifornire* 'refuel', *tamponare* 'rear-end', *truccare* 'trick'
HAIR_VERBS	*annodare* 'knot', *pettinare* 'comb', *sistemare* 'arrange', *spazzolare* 'brush', *spettinare* 'mess up', *tagliare* 'cut', *tingere* 'dye', *tirare* 'pull'
MONEY_VERBS	*accettare* 'accept', *accreditare* 'credit', *attirare* 'attract', *cedere* 'yield', *congelare* 'freeze', *depositare* 'deposit', *distribuire* 'distribute', *emettere* 'issue', *guadagnare* 'earn', *incassare* 'cash in', *pagare* 'pay', *prestare* 'lend', *scommettere* 'bet', *vincere* 'win'
PLANTS_VERBS	*abbattere* 'cut down', *coltivare* 'cultivate', *curare* 'cure', *falciare* 'mow', *macinare* 'grind', *piantare* 'plant', *raccogliere* 'harvest', *rispettare* 'respect', *rovinare* ruin', *tagliare* 'cut', *trapiantare* 'transplant', *travasare* 'repot'.

TABLE 3: VERB CLASSES BASED ON THE SEMANTICS OF THEIR DIRECT OBJECTS.

Note that the verbs in the classes may vary considerably in their semantic spectrum, and, consequently, their distribution. For example, in the case of hair_verbs, *tirare* 'pull' has a broad semantic spectrum, and one might arguably claim that it is unusual or far less frequent with an argument like *capelli* 'hair' than with other arguments. This is confirmed by corpus analysis, according to which the typicality score of the combination of *capelli* with *tirare*, calculated through logDice in the Sketch Engine platform[17] is lower (7.4) than that of

[16] Currently, T-PAS contains 1160 analyzed verbs, 5529 patterns and ca. 250,000 annotated corpus instances. The references to the studies performed on TPAS data are listed at `https://tpas.unipv.it/publications/`.

[17] The typicality score in the Sketch Engine platform is based on *logDice*, a metric that indicates how strong the collocation of two words is. The higher the score, the stronger the collocation. A low score means that the words in the collocation also frequently combine with many other words.

capelli with the narrow spectrum verb *pettinare* 'comb' (9.0). At the same time, *tirare* is more typical for *capelli* than *annodare* 'tie' (4.6), and less typical for *capelli* than *tagliare* 'cut' (9.4). We conclude that each class has members that are more or less prototypical and that the typicality score can help define the internal organization of the classes.

It has been noted that the classes in Table 3 show similarities with the *frames* (Fillmore 1982/2006) contained in the FrameNet repository.[18] This is correct, although the methodologies adopted in the two resources (FrameNet and TPAS) to identify frames and patterns differ radically. The frames listed in FrameNet are stipulated with the help of corpus data annotation but they are not induced by manual clustering of corpus concordances based on the semantics of arguments, as are the patterns in T-PAS. The study of the similarities and differences between cognitive frames and semantic patterns represents an interesting direction of future research that could valuably contribute to our understanding of how semantic classes for verbs are organized and can be identified.

5.3 Argument optionality and argument types

As referenced in § 2.2, next to influencing the meaning of verbs, arguments complete it, providing an essential semantic component of the linguistically encoded event, namely the participants. This is why they are syntactically mandatory, as in the case of the direct object of the verb *lucidare* 'polish' in (10) (see the agrammaticality in (10b)):

(10) a. *Luca ha lucidato le scarpe*
'Luca polished the shoes'

b. **Luca ha lucidato*
*'Luca polished'

It has been noted, however, that "sometimes an argument is obligatorily left out of the surface structure because it is subsumed as a part of the meaning of the predicate" (Fillmore 1969: 119). According to Pustejovsky (1995), this particular type of argument (shadow argument) can be expressed only by making it more specific. For example, the argument *con il telefono* 'with the phone' can co-occur with the verb *telefonare* 'phone' only if the type of phone is specified, as in (11b), and not in (11a):

(11) a. **Luca ha telefonato a Luisa con il telefono*
*'Luca phoned Luisa with the phone'

[18] The FrameNet resource can be accessed at `https://framenet.icsi.berkeley.edu`. I am thankful to the participants of the Workshop on *Argument Structure*, held at Roma Tre University in September 2022, for pointing this out.

b. *Luca ha telefonato a Luisa con il suo telefono portatile*
'Luca phoned Luisa with his mobile phone'

Interestingly, Jezek (2018) notes that shadow arguments do not occur only with denominal verbs such as *telefonare* 'phone' but also with verbs involving body parts, such as *camminare* 'walk' (feet) and *vedere* 'see' (eyes):

(12) a. **Luca stava camminando a piedi*
*'Luca was walking by foot'
b. *Luca stava camminando a piedi scalzi*
'Luca was walking barefoot'

(13) a. **Luca ha visto una stella cadente con gli occhi*
*'Luca saw a shooting star with his eyes'
b. *Luca ha visto una stella cadente con i propri occhi, a occhio nudo*
'Luca saw a shooting star with his own eyes, with the naked eye'

It was also noted that there are arguments, known as hidden arguments, that cannot be expressed in the verb's syntax at all. An example is provided by the arguments of verbs of implicit creation (Bisetto & Melloni 2005; Jezek 2014), that is, verbs that denote the creation of something through the representation of a source (Jezek & Pustejovsky 2019), such as *fotografare* 'photograph', *copiare* 'copy' and *dipingere* 'paint'. Implicit creation verbs denote the coming into being of a new entity as a result of the event itself, which does not appear superficially as the Direct object of the verb; instead, the Direct object expresses the source of the representation (*representation source theme* in Dowty's (1979) terminology). This is the case of the Direct object in (14): the created entity (the photograph) is not expressed nor expressible as an argument of the verb in the syntax:

(14) *Luisa ha fotografato la targa*
'Luisa photographed the license plate' (= created entity → photograph)

The resulting entity, although not expressed, can be referred to indirectly via anaphoric expressions. In (15), for example, the clitic pronoun establishes an anaphoric relation with the unexpressed created entity, the resulting photograph:

(15) *Luisa ha fotografato la targa e l'ha inviata a Luca*[19]
'Luisa photographed the licence plate and sent it to Luca'

[19] It should be noted that in a sentence such as *Luisa ha fotografato il tramonto e l'ha mandato a Luca* 'Luisa photographed the sunset and send it to Luca' the clitic shows agreement with the noun *tramonto* (masculine) and not with the noun *fotografia* (feminine). This may be

The anaphoric test supports the hypothesis that, with these verbs, the participant that expresses the result of the action is an argument, but it is hidden and not expressible in the syntax of the verb.

Finally, it is well known that some arguments show optionality in the syntax, that is, they can be dropped or omitted, as noted in § 4. In Jezek (2018), it is argued that syntactic optionality is licensed by either pragmatic or semantic factors. A pragmatic omission occurs when an argument remains unexpressed because the referent is known to the interlocutors. In this case, the saturation of the valency of the verb takes place through the retrieval of this information from the situational context. This is the case of the examples in (16), in which the dots in the brackets highlight the information that must be reconstructed for the expressions to be semantically complete and interpretable:

(16) *Gianni iniziò (. . .) alle 8.30 e finì (. . .) alle 15*
'Gianni started (...) at 8.30 and finished (...) at 3pm'

The omissions in (16) have a pragmatic and/or textual nature (Definite Null Instantiations, Fillmore 1986). The verbs *iniziare* and *finire* still denote a relationship between two participants, even if the second argument is not realized in the syntax, and force the listener to find something in context to saturate their meaning. If an element is not contextually assigned to the omitted argument, the sentence remains semantically incomplete. A different type of omission is at play when syntactic optionality is licensed by semantic factors. In this case, Jezek (2018) argues that an argument may be dropped when it can be predicted based on the verb semantics. Specifically, the implicit information that needs to be filled in to saturate the verb is the category of things that the verb typically selects, for example, the FOOD class in the case of *eating* in (17a) and the DOCUMENT class in the case of *reading* in (17b):

(17) a. *Luisa ha mangiato alle 18.00*
'Luisa ate at 6pm'

b. *Luca legge spesso in treno*
'Luca often reads in the train'

In other words, the target of the omission in (17a) and (17b) is the particular class of objects intended by the verb for that argument, and not a specific referent of the class, as in the case of the pragmatic omission in (16). Only

accounted for by assuming that the clitic refers metonymically to the information (the sunset) transferred through the photograph. In any case, the transfer cannot take place in the absence of a resulting object on which the information is recorded: this holds also for ex. (15). See Jezek & Pustejovsky (2019) for further details.

under the condition that the omitted object is understood generically as a class and the focus is on the action being performed, can the omission be interpreted as semantic. If the target of the omission is a specific referent of the class (for example, a pizza for FOOD; a letter for DOCUMENT), then the omission should be interpreted as pragmatic.

5.4 On the incorporation of arguments

After reviewing the previous semantic accounts of the variability in the syntactic behaviour of different types of arguments, we propose a comprehensive account in which the relation between the verb and its arguments can be understood and represented as a relation of greater or lesser incorporation. The notion of incorporation has been used primarily in syntactic studies (Mithun & Corbett 1999). Following a suggestion that dates back to Gruber (1965), Jezek (2018) applied it to semantics. According to this proposal, the arguments show varying degrees of semantic incorporation into the meaning of the verb they depend on, which, in turn, determines their syntactic behaviour. This concept is represented in (18), where the verbal root is meant not as the root in morphological terms but as the root in semantic terms (Levin & Rappaport Hovav 2005), coinciding with what we have called the ontological type of verbs (motion, perception, cognition, etc.) in § 2.

(18) $[[[[V_{root}\ arg_1]\ arg_2]\ arg_3]]$

The most incorporated arguments (arg1) are the arguments that reside in the root. They can be extracted and expressed in the syntax only if they add new information to the information already introduced by the root, that is, if they are 'informative' (as in (11b)). Partially incorporated arguments (arg2), on the other hand, are less built-in in the root and are, therefore, more easily projected into the syntax, but they can be omitted because they are still predictable from the verbal meaning. This is the case of arguments that can be omitted based on semantic factors (see (17)). Finally, arguments that are not incorporated to any degree (arg3) cannot be omitted, except for pragmatic reasons (that is, unless the referent they introduce can be reconstructed pragmatically; see the examples in (16)). According to our proposal, only arg3 are outside the verbal root: this justifies their syntactic mandatoriness in all contexts, as in (10). Whether the type of argument discussed in relation to creation verbs (cf. (14)) shall be accounted for in terms of incorporation, remains a matter of debate. The representation in (18) serves as a mnemonic illustration of the different degrees of argument incorporation and has a theoretical status. It is possible to integrate this idea into theoretical frameworks, as long as they allow for different types

of arguments. For instance, based on the proposed argument type distinctions in Generative Lexicon (Pustejovsky 1995), Jezek (2018) suggested that arg1 represent *shadow arguments*, arg3 represent *true arguments*, and arg2 represent *true arguments* that may be semantically *defaulted*.

6. CONCLUDING OBSERVATIONS

After reviewing the three main components of verb meaning, in this contribution we have initially focused on the utility of syntactic approaches to classifying verbs and their arguments at the syntax-semantic interface, especially if they consider the alternation patterns in which each verb occurs. Secondly, we have highlighted that these approaches leave two critical issues unsolved, namely the argument-adjunct distinction and the syntactic optionality of arguments. We have proposed a solution based on semantic principles. We have shown that adopting a semantic perspective as a starting point in analyzing the relationship between the verb and its arguments, and in particular assuming that arguments play an active role in the semantic composition with the verb whereas adjuncts are inactive on the verb meaning, can contribute to account for phenomena that cannot be explained by the traditional syntax-first approach. Moreover, we have shown that by assuming that arguments may be semantically incorporated to different degrees in the verbal meaning, one can successfully account for their syntactic optionality, when it is not due to pragmatic factors.

REFERENCES

Bach, E. (1986). The Algebra of Events. *Linguistics and Philosophy*, 9. 5–16.

Bisetto, A. & C. Melloni (2005). Result Nominals: A Lexical-Semantic Investigation. In *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5),* September 15-18, 2005, Fréjus, France, a cura di Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli e Sergio Scalise. Online, 393–412.

Black, M. (1954). Metaphor. In *Proceedings of the Aristotelian Society* (New Series), volume 55 (1954/55). Oxford, UK: Oxford University Press, 273–294.

Brown, S.W., J. Bonn, G. Kazeminejad, A. Zaenen, J. Pustejovsky & M. Palmer (2022). Semantic representations for NLP using Verbnet and the generative lexicon. *Frontiers in Artificial Intelligence*, 5.

Cennamo, M. (2001). Classi verbali e cambiamento sintattico: la reinterpretazione passiva del costrutto riflessivo. In G. Salvi (ed.) *Semantica e lessicologia storiche,* Atti del XXXII Congresso Internazionale della Società di Linguistica Italiana. Roma: Bulzoni.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

Dang, H.T., K. Kipper, M. Palmer & J. Rosenzweig (1998). Investigating regular sense extensions based on intersective Levin classes. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics,* Volume 1 (Montreal, QC). 293–299.

Dowty, D.R. (1979). *Word Meaning and Montague Grammar*. Dordrecht: Reidel.

Fellbaum, C. (ed.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: The MIT Press.

Fillmore, C.J. (1969). Types of lexical information. In F. Kiefer (ed.) *Studies in Syntax and Semantics*, 109–137. Dordrecht / Boston: Reidel.

Fillmore, C.J. (1982/2006). Frame Semantics. *Cognitive Linguistics: Basic Readings*, 34. 373–400.

Fillmore, C.J. (1986). Pragmatically Controlled Zero Anaphora. In V. Nikiforidou (ed.) *Proceedings of the XII Annual Meeting of the Berkeley Linguistics Society*. Berkeley, Cal., Berkeley Linguistic Society, 95–107.

Fillmore, C.J. & C.F. Baker (2010). A Frames Approach to Semantic Analysis. In B. Heine & H. Narrog (eds.) *The Oxford Handbook of Linguistic Analysis*. Oxford, UK/New York, NY: Oxford University Press.

Goldberg, A. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: The University of Chicago Press.

Gruber, J.S. (1965). *Studies in lexical relations.* Ph.D. thesis, Cambridge, MA: Massachusetts Institute of Technology.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge MA: The MIT Press.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: The MIT Press.

Jezek, E. (2003). *Classi di verbi tra semantica e sintassi*. Pisa: ETS Edizioni.

Jezek, E. (2014). Classes of Creation Verbs. In S. R. & F. Masini (eds.) *Word Classes: Nature, Typology and Representations*, 37–50. Amsterdam & Philadelphia: John Benjamins.

Jezek, E. (2016). *The Lexicon: An Introduction*. Oxford, UK: Oxford University Press.

Jezek, E. (2018). Partecipanti impliciti nella struttura argomentale dei verbi. In S. Dallabrida & P. Cordin (eds.) *La grammatica delle valenze*, 55–71. Firenze: Cesati.

Jezek, E. & J. Pustejovsky (2019). Dynamic Interpretation of Predicate-Argument Structure. *Lingue e Linguaggio*, 18(2). 179–207.

Jezek, E., B. Magnini, A. Feltracco, A. Bianchini & O. Popescu (2014). T-PAS; A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland. European Language Resources Association (ELRA), 890–895. `https://tpas.unipv.it/`.

Kipper, K., A. Korhonen, N. Ryant & M. Palmer (2008). A large-scale classification of English verbs. *Language Resources and Evaluation*, 42. 21–40.

Lakoff, G. & M. Johnson (2008). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: The University of Chicago Press.
Levin, B. & M. Rappaport Hovav (2005). *Argument Realization*. Cambridge: Cambridge University Press.
Mereu, L. (2020). *Semantica della frase*. Roma: Carocci.
Michaelis, L.A. & J. Ruppenhofer (2001). *Beyond Alternations. A Constructional Model of the German Applicative*. Stanford: CSLI Publications.
Mithun, M. & G. Corbett (1999). The Effect of Noun Incorporation on Argument Structure. In L. Mereu (ed.) *Boundaries of Morphology and Syntax*, 49–71. Amsterdam-Philadelphia: Benjamins.
Partee, B. (1995). Lexical semantics and compositionality. In D. Osherson (ed.) *Invitation to Cognitive Science* (2nd ed.) in Gleitman, L. and Liberman, M. (eds), Part I: *Language*, 311–360. Cambridge, MA: The MIT Press.
Perlmutter, D.M. (1978). Impersonal passives and the Unaccusative hypothesis. In *Berkeley Linguistic Society*, 157–189. 4, Berkeley, CA.
Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.
Pustejovsky, J. (2012). Co-compositionality in grammar. In M. Werning, W. Hinzen & E. Machery (eds.) *The Oxford Handbook of Compositionality*, 371–384. Oxford: Oxford University Press.
Recanati, F. (2012). Compositionality, flexibility and context-dependence. In M. Werning, W. Hinzen & E. Machery (eds.) *The Oxford Handbook of Compositionality*, 175–91. Oxford: Oxford University Press.
Tesnière, L. (1959). *Éléments de syntaxe structural*. Paris: Klincksieck.
Van Valin, R.D. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
Van Valin, R.D. (2012). Lexical representation, Co-composition, and Linking Syntax and Semantics. In J. Pustejovsky, P. Bouillon, H. Isahara, K. Kanzaki & C. Lee (eds.) *Advances in Generative Lexicon Theory*, 67–107. Cham: Springer.
Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.
Wilks, Y. (1975). A preferential, pattern-seeking, semantics for natural language inference. *Artificial Intelligence*, 6(1). 53–74.

Elisabetta Jezek
University of Pavia – Department of Humanities
Piazza del Lino 2, 27100 Pavia
Italy
e-mail: elisabetta.jezek@unipv.it
https://orcid.org/0000-0003-2518-5200