Generative Lexicon Theory and Lexicography

Elisabetta Jezek
Dipartimento di Studi Umanistici
Sezione di Linguistica Teorica e Applicata
Università di Pavia
jezek@unipv.it

January 14, 2017

Abstract

The interplay between linguistic modelling and lexicographic practise has long been a debated issue. In this chapter we present an exercise aimed at showing that the interaction between the two is beneficial to both, provided that a solid methodology is in the background. As a case study, we focus on the encoding of argument structure for verbs in lexical resources, and use Generative Lexicon Theory as theoretical framework. We discuss the role of arguments in defining the meaning of verbs and identify which Generative Lexicon principles can be helpful for lexicographers in the task of compiling verb entries and disambiguating verb senses based on the nature of their arguments. Finally, we propose a refinement of the set of argument types proposed in the theory, as a result of corpus analysis in lexicographic setting.

1 Introdution

The interplay between theoretical modelling and lexicographic practise has long been a debated issue, seeing on the one side theoretical linguists regarding lexicography as a practice rather than a "real" science, and lexicographers claiming that a good dictionary is the ultimate test of any theory of lexical semantics. In this chapter, we propose an exercise which shows that the interplay between linguistic theory and lexicography is beneficial to both, provided that a solid methodology is in the background. As a case study, we focus on the encoding of the linguistic knowledge associated with verbs in dictionaries, and propose Generative Lexicon Theory (henceforth, GL, Pustejovsky 1005) as as a theoretical background. The Chapter is organized as follows. After a description of the theoretical framework (section

2), we review previous attempts of exploiting GL in lexicographic setting (section 3). We then introduce the main challenges associated with the representation of verbal entries in lexical resources (section 4), and identify and discuss the principles of GL that can be helpful in this task (section 5). Finally, we propose a refinement of the theory based on corpus-informed lexicographic work. Conclusions stress the interdependence of theoretical modelling and lexicographic practice (section 6).

2 Generative Lexicon Theory: an outline

Among the wide variety of linguistic theories at disposal today as an inspiration to lexicographic work, GL theory seems appealing, due to its focus on word polysemy and its concern in representing the variations that the meaning of words displays in actual use. Pustejovsky 1995 proposes that the linguistic knowledge associated with a lexical item may in the lexicon can be represented through four distinct but interconnected information structures, reported in (1):

(1) a. Qualia structure;
b. Lexical typing structure;
c. Event structure;
d. Argument structure.

Qualia Structure (QS) encodes the meaning of a word in terms of the relations it holds with the meaning of conceptually connected words. For example, the Formal (F) relation encodes the taxonomic information that qualifies the word, i.e. its semantic type (*animal*, *human*, *event*, *location* etc.). This provides a fairly general classification of the core meaning of the word. The Constitutive (C) relation encodes the *part_of* relation and its inverse, the relation *has_as_part*. The Telic (T), introduces the intended goal associated with the object, and the Agentive (A) specifies the factors involved in the objects origin. GL lexical representations are based on attribute value matrixes, according to which the four relations and their values can be represented as in (2):

(2)
$$\begin{bmatrix} \alpha \\ \text{QUALIA} = \begin{bmatrix} \text{F} = \textbf{what } \alpha \textbf{ is} \\ \text{C} = \textbf{what } \alpha \textbf{ is made of} \\ \text{T} = \textbf{function of } \alpha \\ \text{A} = \textbf{origin of } \alpha \end{bmatrix} \end{bmatrix}$$

When applied to a noun such as *house*, the resulting Qualia Structure is (3):

$$(3)\quad \begin{bmatrix} \textit{house} \\ \text{QUALIA} = \begin{bmatrix} \text{F} = \textbf{building} \\ \text{C} = \textbf{\{door, room, ...\}} \\ \text{T} = \textbf{live\_in} \\ \text{A} = \textbf{build} \end{bmatrix} \end{bmatrix}$$

which reads as follows: *house* has *building* as value for F; *door*, *room* etc. as values for C, *living_in* as value for T, and *build* as value for A. The Lexical Typing Structure locates a word within the type system of the language on the basis of the semantic type it denotes. For example, a *house* is a type of *building*, a *building* is a type of $artifact$ and so forth, along the taxonomic tree.

The Event Structure (ES) identifies the specific event type associated with words expressing events (predicative verbs such as *love*, *rain*, *run*, *eat* and event nouns such as *training*, *relation*, and *meal*). The primitive event types posited in GL are States (S), Processes (P) and Transitions (T). States are single events, which are evaluated relative to no other event (*love*, *know*). Processes are sequence of events identifying the same semantic expression (*run*, *push*)[1], while Transitions are sequence of events identifying semantic expressions, which are evaluated relating it to their opposition. They are typical of causative predicates (*open*, *build*), which are conventionally analyzed as involving an initial act or process followed by a resulting state (see (4), where P stands for predicative expression).

$$(4)\quad \begin{bmatrix} P \\ \text{ES} = \begin{bmatrix} \text{E1} = e_1 \\ \text{E2} = e_2 \end{bmatrix} \end{bmatrix}$$

For example, the Event Structure S for the verb *build* is:

$$(5)\quad \begin{bmatrix} \textit{build} \\ \text{ES} = \begin{bmatrix} \text{E1} = e_1 : \textit{process} \\ \text{E2} = e_2 : \textit{state} \end{bmatrix} \end{bmatrix}$$

Finally, Argument Structure (AS) encodes the participants in the event expressed by a predicate, which are grammatically selected as arguments.

[1] In this context, we abstract away from the changes which can be identified when looking at events dynamically and in more fine-grained details (see Pustejovsky 2013).

The standard inventory of GL argument types includes three basic types: *true arguments*, *default arguments* and *shadow arguments*. *True arguments* are obligatorily expressed in the syntax, *default Arguments* are part of the Qualia but may remain unexpressed under certain conditions, and *shadow arguments* are implicit in the predicate and cannot be expressed.

In a GL lexical entry, argument types are directly encoded in the representation of argument structure as illustrated in (6), where ARG is a true argument, D-ARG is a default argument, and S-ARG is a shadow argument.

$$(6)\quad \begin{bmatrix} \alpha \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \ldots \\ \text{ARG2} = \ldots \\ \text{D-ARG1} = \ldots \\ \text{S-ARG1} = \ldots \end{bmatrix} \end{bmatrix}$$

Importantly for the discussion that follows, arguments to a predicate in GL are *typed*, i.e. the verbs assigns a semantic preference to its argument, which is expressed as a semantic type.

For example, the verb *visit* in its sense of 'go to and spends some time in a place for tourism, business, or some other purpose' selects two true arguments: arg1 is of semantic type *human*, while arg2 is of semantic type *location*. This can be represented as in (7):

$$(7)\quad \begin{bmatrix} \textit{visit} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \textit{human} \\ \text{ARG2} = \textit{location} \end{bmatrix} \end{bmatrix}$$

The type of an argument is a value selected from an inventory of types in the language (Asher and Pustejovsky 2006, Pustejovsky 2011). In addition to the Montague types, e and t, GL assumes a richer subtyping over the entity domain than is typically assumed in type theory, for example, *physical*, *abstract*, *human*, *location*. As we will see, this information is particularly useful to assign senses to verbs in context and organize a verbal entry in a lexical resource accordingly.

Thematic role constraints on arguments, such as *Agent*, *Patient*, *Experiencer* etc. are not included in the basic GL argument structure representation. The view taken in standard GL is that argument selection is a typing mechanism, i.e. an operation through which the verb imposes a semantic type requirement on the argument, rather than a role.

Syntactic requirements are also not included in traditional GL verbal entries, as it is assumed that they are derived from semantic properties.

As regards compositionality, the theory foresees that the informational structures encoding lexical knowledge outlined above interact in composition according to three mechanisms that modulate the meaning of words in their context of use. These are:

(8) a. co-composition;
    b. coercion;
    c. subselection;

Of particular interest to our work is co-composition, which is frequently active in predicate-argument combination. Consider for instance the contexts in (9), where the Engl. verb *take* means something different depending on what is said to be taken:

(9) a. Take a tablet. (ingest)
    b. Take a train. (travel with)

According to GL, the difference in meaning between (9a) and (9b) is due to the application of the compositional mechanism of co-composition. When co-composition applies, the meaning of the verb is refined contextually by the information provided by the complement. Specifically, in the case of *take*, *tablet* and *train* provide the information that when they are "taken", a different type of action is performed; a tablet is ingested and a train is travelled with. In Pustejovskys terms, the intended goal associated with the complement (its T value) unifies with the verbal meaning, resulting in a novel interpretation of the verb in context. Under this view, the verbal meaning is built incrementally by combining words, and there is no need to assume a distinct sense of the verb for each use. This procedure is different from the classic composition principle, because the underlying operation is not a mere arithmetic sum; in co-composition both the predicate and the complement are active in building the resulting interpretation. We will come back on co-composition in 5.1, where we will address the task of distinguishing among verb senses in a lexicographic setting.

3 Previous attempts of exploiting GL in lexicography

GL has been already exploited in lexicographic projects, in particular those devoted to building lexical resources for computational applications. In this context, to the best of our knowledge, the component of GL which has been

used most is Qualia Structure. This is not surprising, as Qualia Structure provides a principled way to structure meaning definitions which take into account the subtle interplay between lexical information and pragmatic knowledge. The adequacy of Qualia roles for capturing key aspects of the meaning of words, especially as far as nouns are concerned, results clearly from a parsing of patterns in traditional dictionary definitions, which generally map quite easily on the dimensions expressed via Qualia roles. This is exemplified in the table below for the word *bottle*:[2]

PATTERN	**QUALE**	**DEFINITION**
is_a_kind_of/is_a	F	a bottle is a container
made_of/has_as_part	C	made of glass or plastic
used_for	T	used for storing beverages.

Perhaps the most significant contribution of GL to computational lexicography took place in the framework of the EU-sponsored SIMPLE project (Semantic Information for Multipurpose Plurilingual Lexicons), whose aim was to develop comprehensive semantic lexicons for 12 European languages. In this context, an extended version of the Qualia Structure was proposed, ie. Extended Qualia Structure (Lenci et al 2000), to capture subtle linguistic differences, while maintaining a systematic and consisting structuring of lexical representations (e.g. the relation "Concerns" is a subtype of the Constitutive qualia). Qualia Structure was also proposed as an organizing principle for the top ontology in EuroWordNet (Vossen 2001), as well as in the design of the Brandeis Semantic Ontology (Pustejovsky et al 2006). Turning to argument structure, the GL mechanism of semantic typing has been extensively used in the construction of PDEV (Pattern Dictionary of English Verbs, Hanks and Pustejovsky 2005), and T-PAS (Type predicate Argument Stuctures for Italian verbs, Jezek et al. 2014), to which we return in the next section. Finally, GL's event structure was developed into a subeventual lexical resource in Im (2013) that explores the principles of opposition structure and change in verbs.

4 Argument Structure in Lexicography

With notable exceptions, argument information associated with verbal entries (often referred to as *valence* (or *valency*) *information* and *complementation-*

[2]Less obvious is how Qualia Structure can contribute to the definition of verbs, a question we will not address in the current Chapter.

patterns) has traditionally played a minor role in dictionaries for native speakers, while it tends to be widespread in bilingual, multilingual and learner's dictionaries. This is because it is commonly believed that the main purpose of valence specification is to clarify how verbs work in syntactic environment, which is considered relevant for encoding but not for decoding purposes. In the following, we will argue that this is a gross underestimation of the explanatory power of valency with respect to verbal behavior, and support the view that a semantic interpretation of valency structure in lexicographic settings, as proposed in lexical models such as GL, offers benefits on several levels, primarily on drawing principled sense distinctions for verbs.

In computational lexicography, on the other hand, large-scale resources that provide relational information about predicates and their arguments have long proved to be indispensable tools for a wide range of NLP applications, where the participants of a certain event expressed by a predicate need to be detected. Consequently, many efforts have been done in compiling inventories of argument structures for NLP purposes and nowadays a variety of lexical resources are available for several languages.

Among the manually crafted resources, one of the most known is certainly FrameNet (Ruppenhofer et al. 2010), currently available for several languages, which provides a thorough analysis and corpus annotation of semantic frames, intended as "script-like conceptual structures that describes a particular type of situation or event along with its participants" (Ruppenhofer et al. 2010, p.5). VerbNet (Kipper-Schuler 2005) is instead a hierarchically organized verb lexicon currently available for the English language, based on Levin's verb classification (Levin 1993) and subsequent extensions; it is not corpus-based and tries to derive semantic descriptions from complementation patterns. Prop-Bank (Palmer 2005), focusing on English as well, is a treebank annotated with predicate-argument structures, focusing on semantic role annotation. WordNet (Fellbaum 1998), a database of semantic relations between words and their senses, also includes frames, although they are defined at a more coarse-grained level (i.e. for *give*, "Somebody —-s somebody something". FrameNet, VerbNet and WordNet are now integrated as a result of the SemLink project (Palmer 2009).

Another category of resources are those acquired automatically from large corpora through statistical analysis, such as VALEX for English (Korhonen et al., 2006) and LexScheme for French (Messiant et al., 2008). Both these resources consist of an inventory of subcategorization frames. The focus in this case is on syntactic realization rather than meaning.

A thorough examination of the available lexical resources (manually an-

notated, corpus-derived, compiled automatically etc.) shows that despite the growth in number, most resources carefully record the list of overt syntactic patterns, thematic roles, and/or semantic restrictions associated with argument positions, but they do not systematically address the core problem of explicitly linking the variations in syntactic behavior of verbs with their meaning variability in a systematic way. Nevertheless, it is widely acknowledged that the interplay between syntactic behavior and meaning variation is by no means univocal and may follow different patterns, summarized below (for a comprehensive discussion, see Levin Rappaport 2005).

(10) a. Same syntactic behavior, different sense:
"Peter *opened* the door" ('create an aperture')
"Peter *opened* a restaurant" ('start an activity').

b. Different syntactic behavior, same sense:
"Flightless birds *inhabit* the island" (tr)
"Flightless birds *live on* the island" (intr).

c. Different syntactic behavior, different sense:
"The process *has started* ('begin', 1 arg)"
"Prices *start* from 100 euro" ('have a value as starting point', 2 arg).

As shown in (10), syntactic behavior does not match verb sense variation in a straightforward way. Something more sophisticated is at play at the syntax-semantic interface, which is not captured by classification systems that conceive the syntactic and semantic dimension as separate modules, and prioritize syntactic over semantic description, as is commonly best practice in lexicography. Syntax-based systems of verb representation have long proved inadequate, causing redundancy in semantic encoding. This can be clarified with a straightforward example. Consider the verb *finish* in the two contexts below, where the object complement is realized with a direct object in (a) and an infinitive clause in (b).

(11) a. I just finished my first reading of the book.
b. I just finished reading the book.

Clearly, the meaning of the verb in the two sentences is the same, i.e. "bring an activity to an end". A resource which prioritizes subcategorization frames in verb classification and associates a distinct meaning to each frame will

then be forced to repeat the same semantic information for the two distinct syntactic realization.

In the following, we will examine how GL can be useful in filling this gap, by assuming semantics as the guiding principle in lexicographic description of verb argument structure.

5 Applying GL argument structure theory to building verbal entries

As referenced in the previous section, the key problem in introducing explicit argument structure information in lexical resources is to harmonize the encoding of syntactic (or subcategorization) frames with the distinctions drawn by the lexicographer among the verb senses. Given that the link between syntax and semantics is not biunivocal, one has to take a stance as to which information shall be presented first. Syntax-based approaches have proved useful but shown at the same time that they have limitations, such as the unnecessarily multiplication of senses, as in the example of *finish* above.

5.1 Syntax-Semantic Interface and the Type System

Let us then reverse the perspective and focus on semantics, to verify whether this perspective is more apt for lexicographic purposes. This allows us to observe that there appears to be an very efficient semantic principle guiding sense variation in verbs. This principle is based on the semantic type of the arguments, which, as we saw above, stands out as a prominent feature in GL argument structure theory. In simple terms, the principle stipulates that semantic distinctions among the different senses of a verb depend crucially on the semantic type of its arguments. For example, it is the semantic type preference of the direct object that distinguishes *toasting* a person ('celebrate') from *toasting* a piece of bread ('cook under radiant heat'). This distinction between the two senses of *toast* can only be expressed formally if the semantic type *human* is available in contrast with other semantic types, e.g. (here) $food$. See the corresponding GL representations in (12) and (13):

(12)
$$\begin{bmatrix} toast\textbf{.v} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = human1 \\ \text{ARG2} = human2 \end{bmatrix} \end{bmatrix}$$

(13)
$$\begin{bmatrix} toast\textbf{.v} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = human \\ \text{ARG2} = food \end{bmatrix} \end{bmatrix}$$

In order to exaplain how argument types affects verb meaning in context, we can recur to the GL principle of co-composition introduced in section 2. According to this principle, all constituents contribute functionally to the meaning of a complex expression; in this view, arguments not only compete the meaning of the verb but act functionally on it, co-determining the resulting meaning of the verb in context. In this view, this principle offers a systematic criterion and a solid methodology to draw sense distinction for verbs not only to scholars interested in compositional semantics but also to traditional and computational lexicographers.

GL mechanisms of semantic typing and co-composition have already been integrated with corpus-based lexicographic procedures, in particular the Corpus Pattern Analysis (CPA) methodology and the Theory of Norms and Exploitation supporting the CPA approach (Hanks 2013). The result has been used as a theoretical and methodological background in the construction of PDEV (Pattern Dictionary of English Verbs, Hanks and Pustejovsky 2005) and T-PAS (Type predicate Argument Structures for Italian verbs, Jezek et al. 2014), and to enhance verb sense disambiguation procedures in a computational setting (Rumshisky 2009).

PDEV and T-PAS are repositories of patterns for English and Italian verbs. Patterns consist of valencies plus the semantic types expected for such valencies, such as human, location, food, vehicle, and so forth. They are acquired from corpora through the inspection, for each valency slot, of the statistically relevant list of collocates that, according to empirical analysis, fill that argument position in text. An example of pattern for the English verb *attend* is given in (14), with the relevant semantic types (in double square brackets), the lexical set for the object position (in curly brackets), information about the syntactic role of arguments and an informal definition of the sense of the pattern, expressed in the form of a primary implicature that includes the typing constraints present in the pattern.

(14) Verb: *attend*
Pattern: [[Human]-subj] attend [[Activity]-obj]
Lexical set [[Activity]-obj]: {meeting, conference, funeral, ceremony, course, school, seminar, dinner, reception, workshop, wedding, concert, premiere, ...}
Primary implicature: [[Human]-subj] is present at [[Activity]-obj]

The theoretical underpinnings of this approach are that words in isolation are better characterized as having an abstract meaning potential rather

than a meaning as such, and that this potential may be exploited in different ways in actual usage. According to this view, unambiguous meaning is attached to linguistic units larger than words. Patterns are the minimal syntagmatic structures in which all words are unambiguous. Complements are regarded as arguments if they contribute to the way the verb is interpreted in the context of use; this provides an empirically grounded criterion to approach the traditional distinction between argument and adjunct, which is often questionable and hard to turn into robust generalizations.

To conclude, GL argument typing mechanism and the co-composition principle prove to be useful in providing lexicographers with a principled way to draw sense distinctions for verbs, generalizing from corpus data.

5.2 Argument Types and the Mandatory/Optional Contrast

Let us proceed with our survey by looking at another aspect of argument structure theory in GL, namely the proposed tripartite classification in true, default and shadow arguments. The goal is to verify whether this classification can be exploited in the encoding of argument structure information in lexical resources.

5.3 True Arguments and Pragmatics

According to GL, true arguments are those event participants that are selected by the predicate as parts of its meaning and that must be expressed syntactically, otherwise the predicate cannot be interpreted because its meaning remains incomplete.

To better understand the properties of true arguments, it is convenient to call attention to the fact that true arguments are part of the meaning of the verb in the sense that they encode the grammatically relevant participants but do not express inherent properties of the event itself, such as presence or absence of change and ontological type (motion, perception, communication), which are instead expressed by what we shall call the verbal *root* (drawing from Levin and Rappaport Hovav 2005); in other words, true arguments are *informative* with respect to these properties, i.e. they add information besides the actional and ontological information on the event type provided by the verbal root.

An examples of true argument is given in (15), where one can see there that the verb *lock* requires the thing being locked to be expressed.

(15) a. After she locked the front door, she went to bed.
b. *After she locked, she went to bed.

Nonetheless, several verbs allow one of their true arguments to remain unexpressed, under the condition that the referent of the argument may be retrieved or constructed from the immediate linguistic or situational or discourse context. Consider for example:

(16) John started at 8.30 and finished at 5pm.

In (16), what John started and finished must be identified in the context for the expression to be interpreted. That is, although the expression in (16) appears syntactically complete as it is, to understand what is being asserted, an interpreter must identify that thing. This kind of "unarticulated" true argument is currently non discussed in GL argument structure theory. We propose to call it a *pragmatically defaulted argument* (PD-Arg), a notion which shares some properties with, but does not correspond to, the notion of Definite Null Instantiation (DNI) introduced in Fillmore 1986. Additional examples of pragmatically defaulted arguments are given in (17):

(17) a. Marc left early and I followed (him).
b. When he found out (about ...), it was too late.
c. She tried (to ...) but did not succeed (in ...).

In (17) we don't know what to make of *follow*, *find out*, *try*, and *succeed*; these verbs force us to grab something from the context to saturate their meaning. Pragmatically defaulted true arguments are omissions of an instance salient in the situational context. Objectless verbs of this kind still denote a two-place relation, even though the second argument is not realized in surface syntax. Unless a particular filler is contextually assigned to the unrealized argument, the sentence remains semantically incomplete. With true arguments, the need of completion is not a contextual matter, but a context-independent property of the verb type. The verb cannot be interpreted unless all its true arguments are specified.

5.4 Defaulted Arguments

In GL theory, a default argument is defined as an event participant which is lexically selected as an argument by the predicate but unlike a true argument it may be left unexpressed, while still being present at the level of interpretation. The omissibility conditions of default arguments are scarcely investigated in the theory (with the exception of Pustejovsky, 2000). In the following, we will elaborate on default arguments in light of the notion of pragmatically defaulted true argument introduced in 5.3. To start with, let

us rename a default argument as intended in standard GL a *lexically defaulted argument* (LD-Arg), to distinguish it from pragmatically defaulted arguments. To understand how the two differ, consider the following examples of lexically defaulted arguments in (18a) and (18b):

(18) a. John ate at 5pm.
b. John read in the car.
c. *John listened in the car.

Notice that in contrast to the examples in (16) and (17), in (18), what John *ate* or *read* need not to be identified for the expressions in (18a) and (18b) to be interpreted. Similar examples are:

(19) a. John wrote to Mary.
b. John parked ten yards after the last house.

Again, in the contexts in (19) there is no need to retrieve the particular thing that John *wrote* and *parked*, as was the case of *start* and *finish* in (16). The unexpressed object is to be generically understood as the class of entities ($food$, $information$, $vehicle$) selected by the predicate. Note that only at this condition (i.e. the condition that the object is understood generically as a class, and the focus in on the action being performed), can the argument be lexically defaulted (i.e. left unexpressed). If a specific instance of the class needs to be mentioned (a pizza, a letter, the BMW etc.), lexical defaulting cannot apply, unless the specific instance can be reconstructed pragmatically, in which case, however, the correct interpretation of the omission is pragmatically and not lexically defaulted argument. To understand this, consider the following example:

(20) Mary prepared John's favorite dish and he ate too much (of it).

In this case, the targeted defaulted argument (between brakets) is not the whole class of $food$; rather, it is the specific instance mentioned in the previous linguistic context (John's favourite dish).

Frequently defaulted arguments are *locations*, *materials* or *substances* used in creating artifacts or objects used to perform actions. In (21), for examples, the defaulted argument specifying the end location of the plane after the flight in (21a) and the initial location of John in (21b) are left unexpressed:

(21) a. My plane arrived late. (where?)
b. John left before breakfast. (from where?)

Defaulted locations are typical of verbs of directed motion such as *arrive*, *leave*, *come*, *go* etc. With *arrive*, the location may also be a condition or state; in this case, however, it cannot be defaulted because in co-composition with an abstract argument the verb takes on a metaphorical meaning, as in (22) and (22):

(22) a. I had arrived at the same conclusion.
b. The tension arrives at maturity.

Note that certain verbs of directed motion do not allow the defaulting of the initial location, and select two true arguments lexically. For example in (23), it seems as though *abandon* has a requirement that the distinguished location argument be a true argument (23a) whereas *leave* does not have such a constraint (see (23b)).

(23) a. They were ordered to abandon the room.
b. *They were ordered to abandon.
b. They were ordered to leave (the room).

As referenced above, not only initial and end locations are typically defaulted, but also the material or substance being used in performing an action. A well-known example is the case of *build* (24):

(24) John built a house. (with what?)

Finally, another type of arguments which is frequently defaulted is the artifact being used in performing an action, as in (25):

(25) Mary cut the grass (with what?)

It has been claimed that *cut* selects for a default argument of the type *cutting_instrument*. The direct object (i.e. the sort of thing that one cuts) and not the verb, determines what value the default argument takes. For example, bread is cut with a knife, hair with a scissor, and so forth (cf. Searle, 1980).[3]

[3]Despite being arguments, lexically defaulted arguments encode a participant which class is highly predictable from the meaning of the verb. Predictability, however, does not seem to be a sufficient condition; for example, *garment* is predictable as object of *wear* but cannot be left out, as (26b) shows:

(26) a. This man does not wear a jogger suit for effect.
b. *This man does not wear for effect.

We will come back on predictability of argument types after discussing the last argument type, i.e. shadow argument, in section 5.5.

5.5 Shadow Arguments

In GL, a shadow argument is defined as an event participant which is semantically incorporated in the meaning of the verb (more precisely the verbal *root*, in our analysis). As such, it is obligatorily left out of the surface structure because it is redundant from the point of view of its contribution to the interpretation of the verb. An example of a verb with a shadow argument is *to phone*, which is conceptually the same as 'to call (somebody) using the phone' but which does not tolerate a prepositional phrase expressing the instrument being used in the action, as illustrated in (27):

(27) a. She phoned the office.
b. *She phoned the office on the phone.

In contrast to (27a), we see that in (27b), the meaning of the verb "shadows" the expression of the instrument through which the action of calling is performed. As opposed to a true argument, the inherent meaning of the verbal root and that of the participant are fused. When a specialization of the shadow argument is made, as in (28), and new information is added, the shadow is "lifted", and the expression of this information is possible. This is illustrated in (28).

(28) She phones the office on the mobile phone.

A mobile phone is a type of phone. An operation of subtyping is necessary to extract the shadow from the predicate meaning and present it as an "autonomous" event participant. The lexical representation of the verb *phone* is as follows:

(29)
$$\begin{bmatrix} \textit{phone}\textbf{.v} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \textit{human} \\ \text{ARG2} = \textit{human} \\ \text{S-ARG1} = \textit{phone} \end{bmatrix} \end{bmatrix}$$

As with *phone*, other verbs of instrument of communication such as *fax* ('send by fax'), *email* and *skype* shadow the participant representing the instrument through which the communication is performed:

(30) a. You can email your comments to the site administrator.
b. *You can email your comments to the site administrator by email.

This type of restriction appears to apply also with defaulted arguments, when they are meant to express the semantic class and not a specific instance, as in (31):

(31) a. *She used to drink a liquid.
b. *She was eating food.

Another class of verbs that licenses shadowed instruments is that of verbs describing motion using a vehicle, such as *to ship* and *to bike*. As with verbs of instruments of communication, these are denominal verbs, i.e. nouns that have come to be used as verbs, an issue on which we come back below:

(32) a. The goods were shipped the next morning.
b. *The goods were shipped by ship the next morning.

(33) a. We biked through the fields.
b.*We biked through the fields with the bike.

Additional examples of verbs shadowing the instrument are illustrated in (34). For each context, we give an informal paraphrase of the meaning of the verb, formulated in terms of decomposition. The paraphrase highlights the role played by the shadow argument in the event and shows that at the core of the meaning of a verb with a shadow argument there is generally a more basic predicate, such as *to cook* for *to microwave*, *to hit* or *to beat* for *to hammer*, *to fasten* for *to lace*, *to clean* for *to brush*, *to close* for *to lock*, and so forth.

(34) hammer metal (hit Obj with S-arg)
lace the shoes (fasten Obj with S-arg)
lock the door (close Obj with S-arg)
iron the dress (straighten Obj with S-arg)
brush your teeth (clean Obj with S-arg)

Note that a single predicate may have both a defaulted and a shadow argument. For example, the verb *lock* shadows the instrument being used to perform the closure, and licenses the *key* as a defaulted argument, i.e. by default, locks are operated by keys. Shadowing is conventionally thought to be confined to denominal verbs that allow for restrictive cognate prepositional phrases, as all of the above, and to verbs which license (optional) restrictive cognate objects, provided that they are specified in a restrictive sense. Shadowing within these two categories is widespead. In (35) we report examples of the latter class.

(35) a. The girl danced every dance.
b. She lived her life in Boston.
c. She slept a sound sleep.
d. I dreamed a dream tonight.
e. Think positive thoughts!

In the contexts in (35), the object further specifies the type of shadow which is otherwise not expressable.

As it happens, the process of shadowing, however, shows up in classes unrelated to (morphological) cognate object alternations. For example, shadowing occurs in contexts as in (36) below, and generally in all verbs shadowing body parts, such as such as *kick* (leg), *walk* (leg), *knock* (hand, finger), in particular perception verbs such as *smell*, (nose); *see*, *watch* and *look* (eyes); *listen*, *hear* (ears), *sneeze* (nose), *nod* (head), and so forth.

(36) * I smelt gas with my nose.

5.6 Argument types and Lexicography

After discussing the different argument types in GL, pointing out their properties, and introducing new criteria for the characterization of optional arguments, let us now focus how these argument types can be useful in lexicographic practise.

True arguments should always be specified in a verbal lexical entry - either implicitly in the definition, or explicitly with a valency formulae (such as [sogg-v], [sogg-v-arg], [sogg-v-arg-prep.arg], [sogg-v-compl.pred], adopted for example in the structuring of verbal entries in the Italian monolingual dictionary Sabatini Coletti, 2007)[4], as they contribute to complete and specify the specific meaning of the verb in their context.

Pragmatically defaulted arguments, on the other hand, need not be signalled in lexicographic context, as they are entirely context-dependent; linguistic expressions contained in dictionaries, as opposed to real text or utterances, usually do nor report this kind of context-dependent omissions.

Lexically defaulted argument should be specified in lexicographic resources, as they are licensed by certain verbs (*eat*, *write*, *park*), but not by others (*lock*, *wear*). Two strategies appear to compete with each other in lexicographic practice: *usage-based*, in which case the most frequent use is specified (e.g. for *fall*, only the goal location), and it logical maximal extension, in

[4]The formulaes read as follows: sogg = subject; v = verb; arg = argument; prep=preposition; compl.pred. = predicative complement

which case the largest possible extension of the complementation pattern is recorded, and bracketing is used to indicate optionality (e.g. for *fall*, both source and goal location). There seem not to be a clear underlying principle guiding one or the other strategy, and bracketing appears to be unsystematic. Focusing on the interpretation of the verb in context, and on the nature of the "missing" argument (instance vs. class), as discusssed above, can help the lexicographer in this respect. Moreover, meaning definitions should be carefully aligned with argument specification; in particular, bracketing should be used when the bracketed argument is included in the definition of the meaning of the verb for that particular case.

Finally, *shadow arguments* are part of the meaning of the verbal root and are therefore usually included in definitions in dictionaries (i.e. *to phone* is defined as 'to call by phone'), and excluded from subcategorization frames, when present, in lexical resources. This is consistent with the idea that they do not add up to the number of syntactic arguments.

5.7 On argument incorporation

The relation among the different argument types proposed in GL and the verb they are argument to, can be represented by assuming the notion of argument incorporation with respect to the verbal *root*. Under this view, there is a cline of argument incorporation, according to which some arguments are more incorporated in the verbal root and can be expressed only if they contribute new information, while others are less incorporated and easily projected in the syntax, as they add information with respect to the verbal root. This is schematized in (37), where it is proposed that shadow arguments are the most incorporated, and together with the verbal root constitute the inherent meaning of the predicate. Lexically Defaulted arguments come next along the continuum; they are highly predictable from the meaning of the verbal root and are expressed only when they are informative, i.e. when they point at a specific instance of the class of referents they denote, instead of at the class as a whole. Finally, true arguments are part of the meaning of the predicate, but are external to the verbal root; therefore they cannot be omitted unless the referent they denote can be reconstructed pragmatically.

(37) $[[[\mathrm{V}_{root}$ s-arg] ld-arg] t-arg]$\mathrm{V}_{predicate}$

From a lexicographic point of view, the notion of argument incorporation is a useful tool to identify the exact contribution of each participant to the

meaning of the verbal root and to establish which participant(s) should be included in its argument structure, and which should instead be regarded as additional information with respect to the core predication.

6 Concluding Observations

In this Chapter, we have focused on the interplay between lexicon theory and lexicographic practice. We have shown that GL provides a useful theoretical background for the lexicographic representation of the argumenthood information associated with verbal entries. In particular, we have provided evidence that the principles of semantic typing and co-composition stipulated in the theory are useful for the lexicographer to draw distinctions among verb senses. In addition, we have reviewed the argument types proposed in GL, and introduced a new distinction between pragmatically and lexically defaulted arguments. This novel analysis, we claim, can be helpful to guide lexicographers in identifying which participants contribute to complete the meaning of the verbal root - and should therefore be included in its argument structure-, and separate them from those which provide contextual information which does not determine the meaning that the verb is going to acquire in the context of use.

References

Asher, Nicholas and James Pustejovsky. 2006. "A Type Composition Logic for Generative Lexicon". In *Journal of Cognitive Science* 6:1-38.

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

Fillmore, Charles J. 1986. "Pragmatically controlled zero anaphora". In Vasiliki Nikiforidou et al. (eds.), *Proceedings of the XII Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA.

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

Hanks, Patrick and James Pustejovsky. 2005. "A Pattern Dictionary for Natural Language Processing". In *Revue Franaise de linguistique applique*, 10.2: 63-82.

Im, Seohyun. 2013. The generator of the event structure lexicon (GESL): automatic annotation of event structure for textual inference tasks". Ph.D. Dissertation, Brandeis University.

Jezek, Elisabetta, Magnini, Bernardo, Feltracco, Anna, Bianchini, Alessia, Octavian Popescu. 2014. "T-PAS: A resource of corpus-derived Types Predicate-Argument Structures for linguistic analysis and semantic processing". In *Proceedings of LREC*.

Kipper-Schuler, Karin. 2005. *VerbNet: A broad coverage, comprehensive verb lexicon*. Ph.D. Thesis. University of Pennsylvania, USA.

Korhonen Anna, Krymolowski Yuval and Ted Briscoe. 2006. "A Large Subcategorization Lexicon for Natural Language Processing Applications". In *Proceedings of LREC*. Genova, Italy.

Lenci Alessandro, Bel, Nuria, Busa, Federica, Calzolari, Nicoletta. Gola, Elisabetta, Monachini, Monica Ogonowski, Antoine, Peters, Ivonne, Peters, Wim and Nilda Ruimy. 2000. "SIMPLE: A general framework for the development of multilingual lexicons." In *International Journal of Lexicography* 13.4: 249-263.

Levin Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*, Chicago, The University of Chicago Press.

Levin Beth and Malka Rappaport Hovav. 2005. *Argument Realization*. Cambridge, Cambridge University Press.

Messiant, Cédric, Korhonen, Anna and Thierry Poibeau. 2008. "LexSchem: A large subcategorization lexicon for French verbs". In *Proceedings of LREC 2008*, Marrakech, Morocco.

Palmer Martha, Kingsbury P., and Daniel Gildea. 2005. "The Proposition Bank: An Annotated Corpus of Semantic Roles". In *Computational Linguistics* 31 (1), 71106.

Palmer, Martha. 2009. "Semlink: Linking propbank, verbnet and framenet." In *Proceedings of the Generative Lexicon Conference*, Pisa, Italy, 9-15.

Pustejovsky, James. 1995. *The Generative Lexicon*, MIT Press, Cambridge, MA.

Pustejovsky, James. 2000. "Lexical shadowing and argument closure". In Yael Ravin and Claudia Leacock (eds.) *Polysemy: Theoretical and Computational Approaches*, Oxford, Oxford University Press, 68-90.

Pustejovsky, James 2013. "Dymanic Event Structure and Habitat Theory". In *Proceedings of the 6th International Conference on Generative Approaches to the Lexicon* (GL2013), 1-20, Pisa, Italy.

Pustejovsky, James, Catherine Havasi, Jessica, Littman, Anna Rumshisky and Marc Verhagen. 2006. "Towards a generative lexical resource: The Brandeis Semantic Ontology". In *Proceedings of the Fifth LREC Conference*, Genoa, Italy.

Rumshisky, Anna. 2009. *Verbal Polysemy Resolution through Contextualized Clustering of Arguments*. Ph. D. thesis, Brandeis University.

Ruppenhofer Joseph, Ellsworth, Michael, Petruck Miriam R.L., Johnson, Christopher R. and Jan Scheffczyk. 2010. *FrameNet II: Extended theory and practice*. International Computer Science Institute, University of Berkeley. (Manuscript, Version of September 14, 2010).

Searle, John 1980. "The background of meaning". In Searle, John, Kiefer, Ferenc and Manfred Bierwisch (eds.) *Speech Act Theory and Pragmatics*. Dordrecht, Reidel, 221-232.

Vossen, Piek. 2001. "Tuning Document-Based Hierarchies with Generative Principles". In *Proceedings of The First International Workshop on Generative Approaches to the Lexicon* (GL2001), Geneve, Switzerland.