Encoding Event and Argument Structures in Wordnets

Raquel Amaro, Sara Mendes, and Palmira Marrafa

CLG, Center of Linguistics of the University of Lisbon,
Av. Prof. Gama Pinto, 2, 1649-003 Lisbon, Portugal
{ramaro,sara.mendes}@clul.ul.pt,palmira.marrafa@netcabo.pt
http://www.clul.ul.pt/clg

Abstract. *In this paper we propose the codification of argument and event structures in wordnets, providing information on selection properties, semantic incorporation phenomena and internal properties of events, in what we claim to be an affordable procedure. We propose an explicit expression of argument structure, including default and shadow arguments, through three new relations and a new order feature. As synsets in wordnets are associated to a given POS, information on the selection properties of lexical items is added. We show that the systematic encoding of event structure information, through five new features at synset level, besides providing the grounds for describing the order of arguments, enriches the descriptive power of these resources. In doing so, we crucially contribute to making wordnets rich and structured repositories of lexical semantic information, that allow for the extraction of argument and event structures of lexical items, thus enhancing their usability in NLP systems.*

Key words: relational lexica, wordnet, argument structure, event structure

1 Introduction

Wordnets are electronic relational databases structured as networks of relations between synsets (sets of synonymous word forms of the same POS), and focusing on conceptual and semantic relations such as synonymy, antonymy, hyperonymy, meronymy, and so on. The original model corresponds to the Princeton WordNet ([1], [2]), a lexical-conceptual database for English containing nouns, verbs, adjectives, and adverbs.

The use of WordNet as a lexical base in NLP applications has made its shortcomings apparent in the perspective of application developers and has led to the need for finer-grained lexical descriptions allowing computational systems to deal automatically with various complex linguistic phenomena in a general and systematic way. For these reasons, and in order to improve the usability of this resource for a variety of applications, the association of semantic and some syntactic information to the WordNet model has been object of research since its appearance (see [3], [4], [5] or [6], among others).

In this paper we propose the encoding of argument and event structures in wordnets, providing information on selection properties, semantic incorporation phenomena and internal properties of events, in what we claim to be an affordable procedure. Aiming at building a computational relational lexicon that models both the semantic and syntactic properties of lexical items, our proposal consists in explicitly expressing argument and event structures in WordNet.PT ([7]), an electronic relational database for Portuguese, developed following the EuroWordNet framework, henceforth EWN ([8]), including default and shadow arguments (as defined in the Generative Lexicon model, henceforth GL ([9], [10])), through a small set of new relations and features which entirely preserve the architechture of the model.

1.1 Related Work

The association of semantic and some syntactic information to the WordNet model has been the subject of much research work developed since its appearance (see for instance [3], [4] or [5]). More recently, and with regard to cross-POS relations concerning selectional properties, several approaches have been taken. [6], for instance, propose the integration of selectional properties in WordNet through the extension of existing word-to-class statistical models to class-to-class preferences that allow computational systems to learn the selectional preferences for classes of verbs, and associate statistic information on the selectional preferences of each sense of a given word (synset). [11], in a similar approach, focus on the identification and integration of phrasets (a type of synset containing multi-word units) to model combinatory idiosyncrasies of lexical units; when the introduction of a phraset is not justified, syntagmatic relations between verbs and their arguments are stated. [12] propose the introduction of instances of verb subjects and direct objects, extracted from linguistically analyzed and annotated corpora. On a different approach, and aiming at enhancing the density of the network, [13] propose the introduction of a new type of relation, based on the concept of evocation, to connect synsets which evoke or bring other synsets to mind.

One of the major results of research on WordNet enhancement is EWN ([8]), a resource reflecting research both on lexical semantics and on the usability of wordnets in NLP applications. Such research resulted in the definition and implementation of a wider variety of lexical-conceptual relations than the set used in Princeton WordNet, focusing on more comprehensive lexical-conceptual relations and cross-POS relations. Specifically, the EWN model describes selectional properties through role relations, as the ones illustrated below, in the sense that these establish a relation between event-denoting synsets and synsets denoting the participants typically involved in them.

1. a. ROLE AGENT relation: *an entity denoted by N1 is the one/that who/which does the event denoted by V1/N2, typically in intentional way*
   b. ROLE PATIENT relation: *an entity denoted by N1 is the one/that who/which undergoes the event denoted by V1/N2*

Role relations are based on thematic role assignment, thus somehow assuming the function of syntactic mapping: agents are commonly assumed to be syntactically realized in subject position, patients in object position, and so forth, as shown by the EWN tests presented above. These relations, however, are defined to connect nodes that lexicalize a given thematic function of an event, or whose thematic function is necessarily or typically associated to the concept denoted, and do not explicitly express selectional properties. Also, as most relations in wordnets, role relations are designed to represent semantic and conceptual properties of the concepts lexicalized in a given language, which do not necessarily reflect selectional properties with precision.

2. a. {*teacher*}N ROLE AGENT {*teach*}V/{*teach*}V INVOLVED AGENT {*teacher*}N, **but** *teach* selects an animated entity as subject.

Focusing on the EWN model, [14] and [15] argue that wordnets, being concept-based computational lexica, should include information on event and argument structures for allowing computational grammars to cope with a number of different lexical semantics phenomena. Specifically, these authors propose new cross-POS relations linking adjectives and nouns, but also the representation of telicity of LCS deficitary verbs, directly related to the structure of these events, through the inclusion of a telic sub-event relation. Also, the relevance of verb argument structure for determining co-hyponymy compatibility is shown and the need to integrate prepositions in wordnets motivated.

Following this approach, in this paper we propose the encoding of argument and event structures in wordnets, through the definition of a small set of new relations and features, providing information on selectional and event internal properties of lexical items, as well as order constraints on their syntactic mapping, without compromising the WordNet model.

2 Encoding Argument Structure

As described above, the EWN model encodes some selection properties of lexical items through *role* relations, that connect nodes lexicalizing a specific thematic function of a given event or whose thematic function is necessarily or typically associated to that event. However, and in spite of being based on thematic role assignment, these relations do not necessarily encode the argument structure of lexical items nor explicitly express selection properties.

In GL ([9], [10]), the argument structure is a level of representation in which the number and type of arguments of a lexical item is stated, including the definition of the semantic properties of its logical arguments, but also syntactic mapping information. The integration of this information in wordnets results in an increase of relevant information available on the semantic and syntactic properties of lexical items. The information defined in the argument structure significantly complements the lexical-semantic information represented through *role* relations and contributes to a more accurate and complete description of the data, enhancing the usability of wordnets as a resource for meaning computation purposes.

Our proposal for encoding argument structure consists in explicitly expressing it through three new relations and a new order feature, overcoming the shortcomings of establishing correspondences between thematic role assignment and syntactic mapping, and providing some information on subcategorization properties. These three relations reflect the three types of arguments considered in GL – true arguments, shadow arguments and default arguments – and are informally defined as follows:

3. SELECTS/ IS SELECTED BY relation (for true arguments)
   $\{$synset$\}_1$ SELECTS $\{$synset$\}_2$ and $\{$synset$\}_2$ IS SELECTED BY $\{$synset$\}_1$ iff $\exists\, x$: $x \in \{$synset$\}_1$ and $\exists\, y$: $y \in \{$synset$\}_2$, and the syntactic realization of x requires the syntactic realization of y, or of z, if $\exists\, z$: $z \in \{$synset$\}_3$ hyponym of $\{$synset$\}_2$.
   Example: $\{$*gallop*$\}_V$ SELECTS 1 $\{$*equine*$\}_N$
4. INCORPORATES/IS INCORPORATED IN relation (for shadow arguments)
   $\{$synset$\}_1$ INCORPORATES $\{$synset$\}_2$ and $\{$synset$\}_2$ IS INCORPORATED IN $\{$synset$\}_1$ iff:
   i) the concept denoted by the $\{$synset$\}_1$ entails the specific concept lexicalized by the $\{$synset$\}_2$;
   ii) $\exists\, x$: $x \in \{$synset$\}_1$ and $\exists\, y$: $y \in \{$synset$\}_2$, and the co-occurrence of x and y is only licensed by subtyping or specification processes; and
   iii) in case of conjoint incorporations, ii) only applies to the element with reference potential.
   Example: $\{$*poison*$\}_V$ INCORPORATES 3 $\{$*with*$\}_{Prep}$ conj1:1
   INCORPORATES 3 $\{$*poison*$\}_N$ conj2:1
5. SELECTS BY DEFAULT/IS SELECTED BY DEFAULT BY relation (for default arguments)
   $\{$synset$\}_1$ SELECTS BY DEFAULT $\{$synset$\}_2$ and $\{$synset$\}_2$ IS SELECTED BY DEFAULT $\{$synset$\}_1$ iff:
   i) the concept denoted by the $\{$synset$\}_1$ entails the underspecified concept denoted by the $\{$synset$\}_2$;
   ii) $\exists\, x$: $x \in \{$synset$\}_1$ and $\exists\, y$: $y \in \{$synset$\}_2$ and the co-occurrence of x and y is only licensed by subtyping or specification processes; and
   iii) in case of conjoint default selections, ii) only applies to the element with reference potential.
   Example: $\{$*build*$\}_V$ SELECTS BY DEFAULT 3 $\{$*of*$\}_{Prep}$ conj1:1
   SELECTS BY DEFAULT 3 $\{$*material*$\}_N$ conj2:1

In order to index the arguments established by SELECTS, INCORPORATES and SELECTS BY DEFAULT relations to a given order, it is necessary to implement an order feature, expressed by numerical tags. Arguments are integrated in a list $\langle 1, 2, \ldots, n\rangle$, ordered from the less oblique to the more oblique[1].

[1] Order here refers to the so-called basic order of constituents. The list of arguments provides the basic unmarked position of arguments, not aiming at accounting for other possible syntactic positions.

The combination of the SELECTS relation with the numerical tags of the order feature allows for extracting the order in which arguments are syntactically realized, depending on their position in the list defined at event structure level. The relation tagged with 1 indicates the argument that is realized in the less oblique position (subject position, in the case of verbs, object position, in the case of nouns), the relation tagged with 2 indicates the argument that is realized in object position, and so on. The INCORPORATES and SELECTS BY DEFAULT relations, although referring to arguments generally not syntactically expressed, are also tagged to assure their correct syntactic position in subtyping or specification contexts, as showed bellow:

6. a. #[The man]$_1$ poisoned [the cattle]$_2$ [with poison]$_3$.
   b.[The man]$_1$ poisoned [the cattle]$_2$ [with arsenic]$_3$.
7. a. #[The man]$_1$ built [the house]$_2$ [of material]$_3$.
   b. [The man]$_1$ built [the house]$_2$ [of wood]$_3$.

Although the integration of argument information in wordnets adds some effort to the encoding task, it does not amount to a true surplus of work since argument structure properties are directly related to the concepts denoted and are frequently required to disambiguate senses and/or to determine the degree of sense differentiation. Thus, the additional work required can amount to significant gains in coherence and sustained sense differentiation options, which make it worth it.

3 Integrating Event Structure

Event structure is the level of representation that regards the internal properties of an event associated to a lexical item. In GL, this level of representation refers to four internal properties of events: their subevents list (E_1= $\mathbf{e_1}$,. . . , E_n= $\mathbf{e_n}$); their Aktionsart type; temporal and order restrictions of their subevents; and their head subevent.

Event structure is the most internal level of representation of event denoting lexical items in GL, in the sense that it comprises semantic properties that are not necessarily (or even not at all) related to external elements. For these reasons, and in contrast with argument structure information, event structure cannot be integrated in wordnets via lexical-conceptual relations established between existing synsets, since the properties it defines are hardly ever lexicalized and thus are not reflected in the nodes in the network.

Given these specificities, we propose to encode event structure as additional information at the synset level, through the use of features that mirror the aforementioned attributes. Also, we claim the need for introducing a new feature that enables the statement of the list of arguments of a given event.

The features presented in Table 1 mirror the attributes used in the GL model. The attribute *event type* can have one of three possible values, corresponding to the three event types, as defined in [16]: **state** (atomic event, not evaluated with regard to any other), **process** (sequence of identical events (complex or not)) and **transition** (event evaluated regarding another event, composed of a process that culminates in a final state, different of the initial one).

ATTRIBUTE	VALUES
event type	*state* *process* *transition*
arguments	$\langle$ *1, 2, ..., n* $\rangle$
subevents	*e1(2,3), ..., en(1,2)*
restrictions	$< \alpha$, $\circ\alpha$, $< \circ\alpha$
head	*e1...n*

Table 1. Event structure features

The value of the attribute *arguments* consists in the list of arguments selected by the event denoted by the synset, ordered from the less oblique to the most oblique one. The natural numbers making up the elements of this list correspond to the *order* feature values (cf. section 2) associated to argument structure relations, allowing the indexation of the selected nodes to a given position in the list.

The *subevents* attribute allows for listing the subevents that compose transition denoting events (according to the established typology of events), with information on the arguments of each subevent. For instance, a transition type event, such as the one denoted by *build*, has as subevents an event argument that corresponds to the process (of building) that leads to a final state (e_1(1,2,3)) and a second event argument that corresponds to this final state (of being built) (e_2(2,3)).

The *restrictions* attribute allows for expressing the three possible temporal ordering relations of subevents established in the GL model: exhaustive ordered part of ($< \alpha$), exhaustive overlap part of ($\circ\alpha$), and exhaustive ordered overlap ($< \circ\alpha$).

Finally, the *head* feature determines the head subevent of a given event-denoting lexical item, this way accounting for Aktionsart properties (achievement vs. accomplishment type events), as well as for events lexically underspecified with regard to event headedness, that typically enter causative/inchoative alternation constructions (see [10], for a detailed discussion).

The features presented in Table 1 allow the expression of event structure without any loss of information. Note, however, that lexicalized subevents are also stated through lexical-conceptual relations at the network level. Thus, verbs that have conceptually individuated and lexicalized subevents, such as {*breathe*} or {*sadden*}, are respectively characterized through HAS SUB-EVENT and HAS TELIC SUBEVENT relations, as follows:

8. {*breathe*}HAS SUBEVENT {*inhale*};{*breathe*} HAS SUBEVENT {*exhale*}
9. {*sadden*}V HAS TELIC SUBEVENT {*sad*}ADJ

Along the lines of what has been argued in the previous section, we consider adding information on event structure to wordnets to be also an affordable effort in wordnet development, involving only filling in the values of new features associated to event denoting synsets.

Although features convey additional information that is not expressed through lexical-conceptual relations, our motivation is that the systematic statement of event structure information, besides providing the grounds for argument order description and consequent syntactic mapping, enriches the descriptive power of these resources, crucially contributing to making wordnets rich and structured repositories of lexical semantic information, thus allowing the extraction of argument and event structures of lexical items, and hence fine-grained and rich lexical entries.

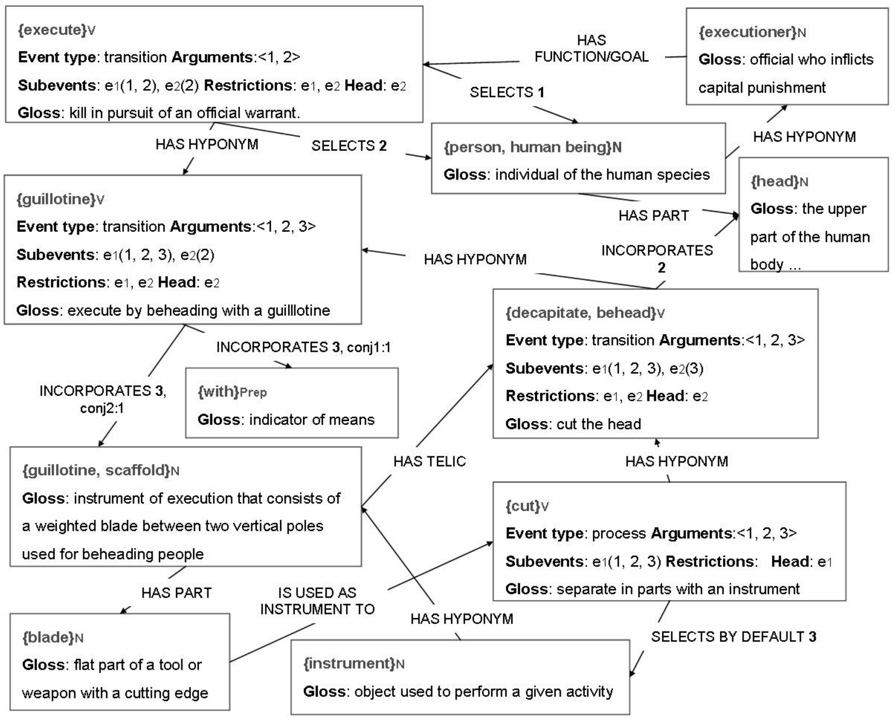

Figure 1. Wordnet fragment with event and argument information

4 Final Remarks

The enhancement strategies proposed in this paper contribute to a significant enrichment of wordnets with semantic and syntactic information that, along with the available lexical structures, the lexical-conceptual relations that connect them in relational models of the lexicon and the percolation of information intrinsic to this model of the lexicon, render wordnets more complete and usable resources for a great variety of NLP tasks and applications. We furthermore show that the different modeling structures of GL and WordNet models are truly complementary and concur to a more accurate representation of the mental lexicon.

References

1. Miller, G. A.: WordNet: an online Lexical Database. Special Issue of International Journal of Lexicography. 4, vol. 3 (1990)
2. Fellbaum, C.: A Semantic Network of English: the Mother of all WordNets. In: Vossen, P. (ed.) EuroWordNet: a Multiligual Database with Lexical Semantic Networks, pp. 137–148. Kluwer Academic Publishers, Dordrecht (1998)
3. Kohl, K., Jones D., Berwick R., Nomura, N.: Representing Verb Alternations in WordNet. In: Fellbaum, C. (ed.) WordNet. An Electronic Lexical Database, pp. 153–178. The MIT Press, Cambridge, MA (1998)
4. Leacock, C., Chodorow, M.: Combining Local Context and WordNet Similarity for Word Sense Identification. In: Fellbaum, C. (ed.) WordNet. An Electronic Lexical Database, pp. 265–284. The MIT Press, Cambridge, MA (1998)
5. Harabagiu, S., Moldovan, D.: Knowledge Processing on an Extended WordNet. In: Fellbaum, C. (ed.) WordNet. An Electronic Lexical Database, pp. 353–378. The MIT Press, Cambridge, MA (1998)
6. Agirre, E., Martinez, D.: Integrating selectional preferences in WordNet. In: Proceedings of the 1st International WordNet Conference. Mysore (2002)
7. Marrafa, P.: The Portuguese WordNet: General Architecture and Semantic Internal Relations. DELTA, Brasil (2002)
8. Vossen, P. (ed.): EuroWordNet A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht, Netherlands (1998)
9. Pustejovsky, J.: The Syntax of Event Structure. Cognition 41, pp. 47–81 (1991)
10. Pustejovsky, J.: The Generative Lexicon. The MIT Press, Cambridge, MA, (1995)
11. Bentivogli, L., Pianta E.: Extending WordNet with Syntagmatic Information. In: Sojka, P., Pala, K., Smrz, P., Fellbaum, C., Vossen, P. (eds.) Proceedings of the 2nd International WordNet Conference, pp.47–53. Brno, Czech Republic (2004)
12. Lemnitzer, L., Wunsch, H., Gupta, P.: Enriching GermaNet with Verb-noun Relations - a Case Study of Lexical Acquisition. In: Proceedings of LREC 2008, pp. 156–160. Marrakech, Morocco (2008)
13. Boyd-Graber, J., Fellbaum, C., Osherson, D., Schapire R.: Adding Sense, Weighted Connections to WordNet. In: Proceedings of the 3rd International WordNet Conference. Brno, Czech Republic (2006)
14. Marrafa, P.: The representation of Telic Complex Predicates in Wordnets: the case of lexical-conceptual structure deficitary verbs. In: Research on Computing Science. vol. 12 (2005)
15. Amaro, R., Chaves, R. P., Marrafa. P., Mendes, S.: Enriching wordnets with new Relations and with event and argument structures. In: Proceedings of CICLing 2006 Conferences on Computational Linguistics and Intelligent Text Processing, pp. 28–40. Mexico City, Mexico (2006)
16. Marrafa, P.: Predicação Secundária e Predicados Complexos em Português. Análise e Modelização. PhD dissertation. University of Lisbon, Lisbon (1993)