

The Representation of Complex Telic Predicates in Wordnets: the Case of Lexical-Conceptual Structure Deficitary Verbs

Palmira Marrafa

University of Lisbon, Faculdade de Letras de Lisboa – DLGR
Alameda da Universidade, 1600-214, Lisboa
Palmira.Marrafa@netcabo.pt

Abstract. This paper has a twofold aim: (i) to point out that telicity is both a lexical and a compositional semantic feature; (ii) to propose a straightforward solution to represent lexical telicity in wordnets-like computational lexica. The approach presented here subsumes the basic idea that lexicon is not a repository of idiosyncrasies. It is rather organized following a few general (universal or parametrical) constraints. In this context, despite the fact that the paper is mainly concerned with Portuguese, cross-linguistic generalizations can be captured, on the basis of a contrastive examination of data. The analysis focus on the behavior of complex telic predicates, in particular those which are deficitary with regard to their lexical-conceptual structure. In order to represent appropriately such predicates in wordnets, the specification of information regarding semantic restrictions, within the corresponding synsets, is proposed as well as a telic state relation.

1 Introduction

Telicity is mostly considered a compositional property of meaning. This paper attempts to make evident it is also a lexical feature and, as a consequence, it has to be represented in the lexicon. A concrete proposal to encode telic information of complex predicates in wordnets is provided.

This proposal emerges from the need of representing the predicates referred to in the Portuguese WordNet (WordNet.PT), which is being developed in the EuroWordNet framework.

From an empirical point of view, the work presented here mainly deals with complex telic predicates, in particular with those which involve lexical-conceptual structure (LCS) deficitary verbs, in the sense defined in previous work (cf. [4] and [5]).

The paper is divided in three main sections: the first one briefly describes the EuroWordNet model; the second one discusses the lexical-conceptual structure (in the sense of [7]) of complex predicates on the basis of a semantics of events, arguing for the lexical nature of telicity, and adduces evidences supporting the idea that some verbs define a deficitary lexical-conceptual structure; finally, the third main section presents an integrated proposal to encode LCS deficitary verbs and their troponyms in wordnets.

2 Wordnets: the EuroWordNet Framework

EuroWordNet is a multilingual database with individual wordnets for several European languages related by an Inter-Lingual-Index (ILI), as sketched in Figure 1.

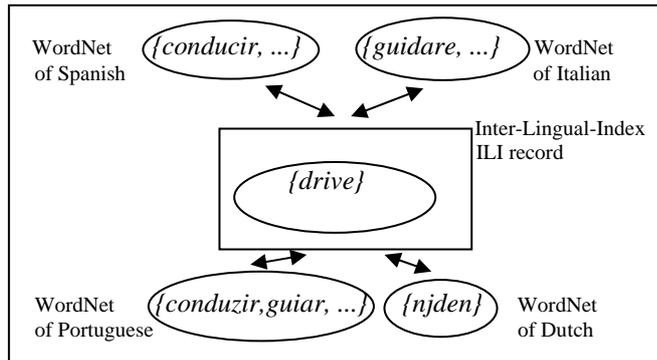


Fig. 1. EuroWordNet general architecture (adapted from [10])

Although initially conceived in the context of a European project, the EuroWordNet model is language independent. Therefore it is extendable to all languages of the world.

The individual wordnets are fundamentally structured along the basic lines of the Princeton WordNet ([1],[2]and[6]).

A wordnet is a conceptual-semantic network, in which the basic units are concepts, represented by sets of synonyms (synsets). A synset contains the set of lexicalizations for a given concept. For instance, the Portuguese expressions *bica*, *café expresso*, *cimbalino* are included in the same synset, since all of them are lexicalizations for the same concept (lexicalized in English by *espresso*).

The meaning of a lexical unit is derived from its relations with the other members of the same synset (lexical relations) and with other synsets (lexical-conceptual relations), as illustrated in Figure 2.

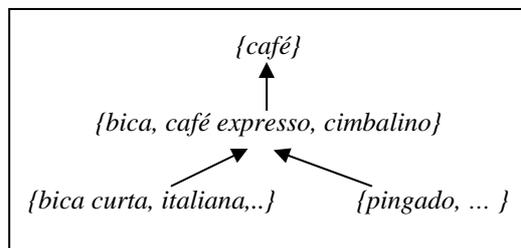


Fig. 2. Examples of relations in WordNet.PT

The meaning of *bica* is partially¹ derived from the synonymy relation with the other expressions inside the same synset and from the conceptual relations with the synset *{café}*, which represents a more general concept, and with the synsets *{bica curta, italiana,...}* and *{pingado,...}*, which represent more specific concepts.

Meaning emerges from the structure of the network. In a certain sense, it is constructed.

Though the conceptual-semantic relations are not the same for all lexical categories, as pointed out by Fellbaum [1], hierarchical relations are the major structuring relations. As well as nouns are mainly arranged by the hyperonymy/hyponymy relation, illustrated above, verbs are primarily organized by troponymy, a manner-of relation which also builds hierarchical structures, as exemplified below:

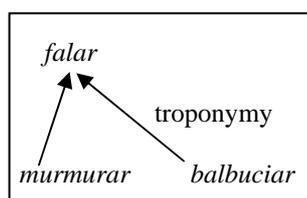


Fig. 3. Examples of troponymy relations

The verbs *murmurar* (“murmur”) and *balbuciar* (“babble”) are troponyms of *falar* (“talk”), specifying aspects related with volume and fluency of the talker, respectively.

The whole-part relation, or meronymy, is another major relation. For instance, *cafeine* is linked to *coffee* by meronymy.

A similar relation is specified for verbs, namely the sub-event relation. To give an example, *pay* is linked to *buy* by the sub-event relation.

The database also includes a set of relations which cover several aspects of semantic entailment. They are used to encode information on the participants typically involved in a given event.

3 Telic Complex Predicates

3.1 Lexical Conceptual Structure

This paper specifically deals with the so-called resultative constructions, like illustrated below:

- (1) He painted the wall yellow.
- (2) He washed the clothes white.

Both *yellow* and *white*, are resultative expressions. Sentence (1) entails that the wall became yellow as a result of painting. Similarly, sentence (2) entails that the clothes became white as a result of washing.

¹ Figure I does not describe exhaustively the relations specified for the synsets considered here.

Expressing the result of the event denoted by verb, the resultative expression integrates the predicate, as extensively discussed in [4] and [5]. In other words, the verb plus the resultative constitute a complex predicate.

As referred to by Stephen Wechsler (cf. [11]), “[i]f there is any aspect of resultatives that is completely uncontroversial, it is that they are telic: they describe events with a definite endpoint”.

Despite this general assumption, there is a major controversy on whether or not the telic aspect of such constructions is an inherent feature of the meaning of the corresponding verbs.

The compositional hypothesis, defended by Verkuyl [9], has been argued for in recent works (see, for instance [3] and [8]) on the basis of contrasts like the following:

- (3) a. John painted his house in one year / *for one year.
- b. John painted houses *in one year/for one year.

At a first glance, these examples suggest that (3)a. is telic and (3)b. is atelic and, consequently, that telicity depends on the nature of the internal argument, more precisely, on its quantifying system. Hence, telicity is a compositional feature of VP and not a lexical feature of V.

However, the relevant opposition seems to be transition *vs* process (or accomplishment *vs* activity, in other terms) and not telic *vs* atelic aspect.

As defended in [4], though the global event in (3)b. is a process, its main sub-events are not atomic events, but transitions. Let us compare the structure of the global event of (3)a. and (3)b., represented by (4)a. and (4)b., respectively.

- (4) a. $[_T [_P e_1 \dots e_n] e_m]$
 T, Transition; P, Process; e, atomic event
 $e_m > e_n$
- b. $[_P [_{T1} [_P e^1_1 \dots e_n] e_{m1}] \dots [_{Tt} [_P e^t_1 \dots e_k] e_{m2}] \dots]$
 $e_{m1} > e_n, e_{m2} > e_k$

Similarly to e_m in a., e_{m1} and e_{m2} , in b., are telic states. This suggests that, although telicity is a compositional feature regarding the whole sentence, it is also an intrinsic feature of the verb.

By default, verbs like *paint* or *wash* are associated to the following LCS:

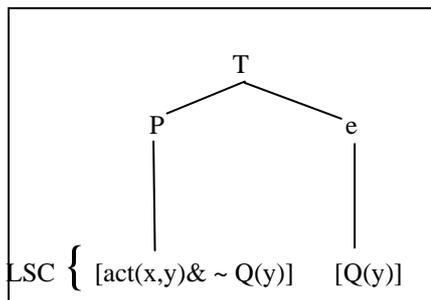


Fig. 4. Telic verbs LCS

Instantiating the variables with the data of sentences (1) and (2) we obtain (i) and (ii), respectively:

- (i) [act(he,wall)& ~ painted_yellow(wall)],
[painted_yellow(wall)]
- (ii) [act(he,clothes)& ~ washed_white(clothes)],
[washed_white(clothes)]

The absence of the resultative (*yellow* and *white* in these cases) does not have any impact on the LCS, as shown below:

- (iii) [act(he,wall)& ~ painted(wall)],
[painted(wall)]
- (iv) [act(he,clothes)& ~ washed(clothes)],
[washed(clothes)]

However, in the case of verbs like the Portuguese verb *tornar* (“make”), discussed below, it seems impossible to assign a value to Q independently of the resultative.

3.2 LCS Deficitary Predicates

Let us examine the following data:

- (5) a. Ele tornou a Maria feliz.
“He made Mary happy”
- b. [act(ele,Maria)& ~ feliz(Maria)],
[feliz(Maria)]
- c. [act(ele,Maria)& ~ tornada_feliz(Maria)],
[tornada_feliz(Maria)]

The LCS of (5)a. seems to be (5)b. and not (5)c.. More concretely, Q , the telic state, is instantiated just with the resultative.

Additionally, the absence of the resultative induces ungrammaticality, as shown:

- (5)d. *Ele tornou a Maria.
“He made Mary”

Along the same lines of [4] and [5], verbs like *tornar* are defended here to be LCS deficitary, in the following sense:

Informal def.:

$$\forall v((v \text{ a verb}, \exists \varepsilon, \varepsilon \text{ the LCS of } v, \exists \pi, \pi \text{ the set of content properties of } \varepsilon, \pi = \emptyset) \Rightarrow \text{LCS_deficitary}(v))$$

Since $\pi = \emptyset$, the LCS can not bear an appropriate interpretation. In these circumstances, the ill-formedness of sentences like (5)d. is previewed. A syntactic structure that projects an anomalous LCS does not satisfy the requirement of full interpretation in Logical Form. Hence, it is ruled out.

In this particular case, the resultative not only expresses a lexical feature but it also fills the gap of the LCS of the verb.

These facts render evident that the representation of complex telic predicates in the lexicon, in particular the representation of those which are LCS deficitary, has to include information regarding the resultative, i.e., the telic expression.

4 Representation in WordNet

4.1 Synset Specifications

As referred to in the first section, concepts are represented in the network by synsets. Each synset includes the lexicalizations for a given concept. Therefore, synsets are supposed to include only lexicalized information.

As the analysis presented here has rendered evident, we have to extend synsets to another kind of information to represent the predicates at issue in an appropriate way.

It would not be adequate to overtly include in the synset all the expressions that can integrate the predicate, among other reasons, because they seem to constitute an open set.

A simple and plausibly solution is proposed below:

```
{tornar |SEM|REST <|TELOS|REST|<REL state|>||>||}
```

As observed, the telic expression of the predicate is represented by a feature structure description that partially specifies the semantic restrictions (SEM|REST) imposed by the verb.

The list of those restrictions includes the attribute TELOS, which stands for the entailed result, whose value includes the specification of a state, more precisely, a relation (REL) whose value is a state.

The pair *REL state* accounts for the fact that the state affects (expresses a relation with) an argument. We can even be more specific and include information to identify the concerned argument, but that is somewhat marginal to the main goals of this paper.

4.2 Telic State Relation

Verbs like *entristecer* (“sadden”) and *alegrar* (“make happy”) denote events that involve a change of state as well, but incorporate the expression that denotes the final state.

In order to capture the relation of the incorporated expression both with the corresponding verb and with the superordinate of that verb, a new relation – more precisely, the telic state relation – has to be included in the set of the internal relations of wordnets, since the existing sub-event relation is not specific enough to account for the facts discussed here. The sub-event relation stands for lexical entailment involving temporal proper inclusion. It has nothing to do with the geometry of the event.

On the contrary, the telic state relation regards the atomic sub-event (or state, in other words), which is the ending point of the global event and affects the theme.

In the case of verbs like *tornar*, that sub-event is implicit – but underspecified – in the meaning of the verb, as referred to above.

The troponyms of this class of predicates, on the other hand, incorporate the telic state, as (6) makes evident:

- (6) a. Ele entristeceu a Maria.
“He saddened Mary”
- b. *Ele entristeceu a Maria triste.
 *He saddened Mary sad”

The representation proposed in Figure 4 accounts very straightforwardly for the facts discussed.

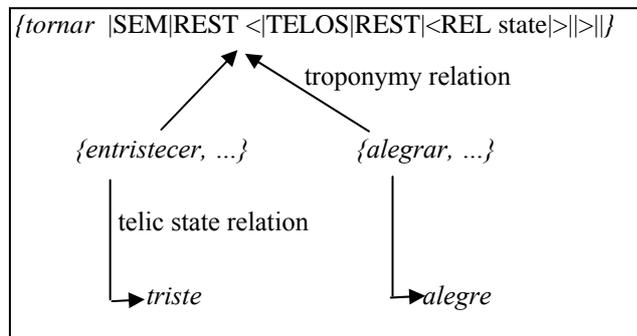


Fig. 5. Subnet for *tornar*

This representation both captures the troponymy relation with the semi-underspecified superordinate synset and relates the TELOS value of superordinate with the telic state incorporated of the troponym.

5 Conclusion

The proposal presented in this paper has a strong empirical motivation and enhance the expressive power of wordnets.

Feature structures are high flexible modelling structures and allow for the specification of information, be it semantic or syntactic, in a very principled way.

The new relation proposed allows for a more integrate and fine grained representation of the facts at issue.

Enriching wordnets in the sense proposed here will open new possibilities for the application of this powerful basic resource in the wide and challenging domain of language and information technologies.

Acknowledgment I thank Instituto Camões for having funded the WordNet.PT project first phase and Fundação Calouste Gulbenkian for funding current basic research which permits to continue to develop this project.

References

1. Fellbaum, C., "A Semantic Network of English: The mother of All WordNets", in P. Vossen (ed.) *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, Dordrecht: Kluwer Academic Publishers, 1998a, 209-220.
2. Fellbaum (ed.) *WordNet: An Electronic Database*, MA: The MIT Press, 1998b.
3. Krifka, M., "Thematic Relations as Links between Nominal Reference and Temporal Constitution", in I. Sag and A. Szabolcsi (eds.), *Lexical Matters*, Stanford, CSLI, 1992.
4. Marrafa, P., *Predicação Secundária e Predicados Complexos – Modelização e Análise*, PhD. dissertation, Lisbon, University of Lisbon, 1993.
5. Marrafa, P., "Predicados Télicos Descontínuos: análise e computação", in P. Marrafa e A. Mota (orgs.), *Linguística Computacional: investigação fundamental e aplicações*, Lisboa, APL/Edições Colibri, 1999, 173-189.
6. Miller G., R. Beckwith, C. Fellbaum, D. Gross, K. J. Miller, "Introduction to WordNet: An On-line Lexical Database", *International Journal Of Lexicography*, 3(4), 1990, pp. 235-244.
7. Pustejovsky, J., "The Syntax of Event Structure", *Cognition*, 41, 1991, 47-81.
8. Schmitt, C., *Aspect of the Syntax of Noun Phrases*, University of Maryland, PhD. Dissertation, 1996.
9. Verkuyl, H., *On the Compositional Nature of Aspects*, Dordrecht, D. Reidel, 1972.
10. Vossen, P., *EuroWordNet: General Document*, University of Amsterdam, 1999.
11. Wechsler, S., "An Analysis of English Resultatives Under the Event-Argument Homomorphism Model of Telicity", *Proceedings of the 3rd Workshop on Text Structure*, University of Texas, 2001.