

Extending WordNets to Implicit Information

Palmira Marrafa

Universidade de Lisboa, Faculdade de Letras and CLUL
CLG – Computation of Lexical and Grammatical Knowledge Research Group
Av. Prof. Gama Pinto, 2
1649-003 Lisboa, Portugal
Palmira.Marrafa@netcabo.pt

Abstract

WordNets mostly deal with lexicalized expressions and lexical-semantic relations among them. Concepts are represented by sets of synonyms (synsets), which constitute the edges of the network. Each synset includes the lexicalized expressions that correspond to a given concept. This paper adduces evidences which support the claim that some concepts, expressed by a subtype of complex telic predicates, are semi-lexicalized, in the sense that the lexicalized expressions corresponding to them do not express, let us say, the whole concept. Since concepts are the basic units of WordNets, they have to be fully represented. Accordingly, a representation which includes the non-lexicalized information is defended. Besides, a new internal semantic relation is proposed, in order to capture appropriately the relationship between the semantics of these predicates and the semantics of their troponyms.

Keywords: Complex predicates, telicity, lexical-conceptual structure, semantic relations

1 Introduction

The proposal presented here is motivated by the research developed to include complex telic predicates in the Portuguese WordNet (WordNet.PT).

The predicates at issue are telic predicates, since they denote an event that entails a change of state of the theme. However, it is not uncontroversial that telicity is a lexical propriety. As a matter of fact, it is rather mostly considered a compositional property of verb phrases.

The proposal defended here argues for that telicity can be both a lexical and a compositional semantic feature. In section 1, which discusses the lexical-conceptual structure (LCS) of telic predicates, evidences that support this analysis are adduced. Section 2 attempts to make evident that verbs like the Portuguese verb *tornar* (“make”) define a deficitary lexical-conceptual structure which, let us say, is compensated by the telic (resultative, in other words) expression they select. Accordingly, an integrated proposal to represent this class of predicates in WordNets constitutes the third, and last, main section of this paper.

2 Complex telic predicates

The proposal presented here specifically deals with the so-called resultative constructions, like the following:

- (1) He painted the wall yellow.

The situation described in (1) entails that the wall became yellow as a result of painting. As extensively discussed in Marrafa (1993), in this type of constructions the verb plus the resultative behave like a complex predicate.

As referred to by Wechsler (2001), “[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 (1972), has been argued for in recent works (see, for instance Schmitt (1996)) on the basis of contrasts like the following:

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

At a first glance, these examples suggest that (2) is telic and (3) is atelic and, consequently, that telicity depends on the nature of the internal argument. Hence, telicity would be a compositional feature of VP and not a lexical feature of V.

However, the relevant opposition seems to be transition *vs* process (in the sense of Pustejovsky (1991)) and not telic *vs* atelic aspect.

As defended in Marrafa (1999) and Marrafa (2003), though the global event in (3) is a process, its main sub-events are not atomic events, but transitions. Let us compare the structure of the global event of (2) and (3), represented by (i) and (ii), respectively:

- (i) $[_T [P e_1 \dots e_n] e_m]$
T, Transition; P, Process; e, atomic event; $e_m > e_n$

- (ii) $[P [_{T1} [P e^1_1 \dots e_n] e_{m1}] \dots [_{Tl} [P e^l_1 \dots e_k] e_{m2}] \dots]$,
 $e_{m1} > e_n, e_{m2} > e_k$

Similarly to e_m in (i), e_{m1} and e_{m2} , in (ii), 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* are associated to the following LCS:

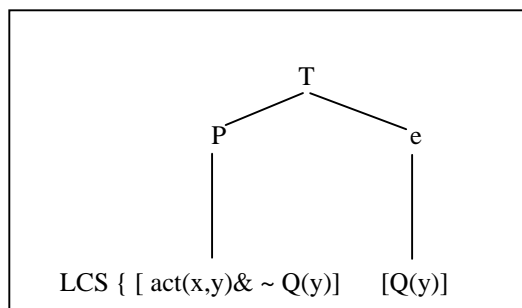


Figure 1: Complex telic predicates LCS

Considering the data of the first example above, we obtain:

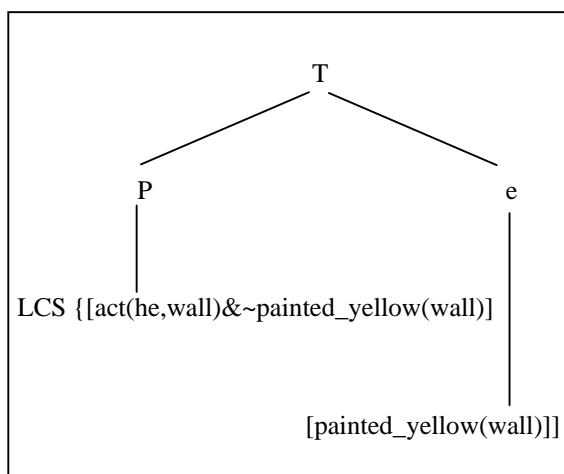


Figure 2: *paint* LCS (resultative expressed)

The absence of the resultative (*yellow*) does not have any impact on the LCS, as we can see:

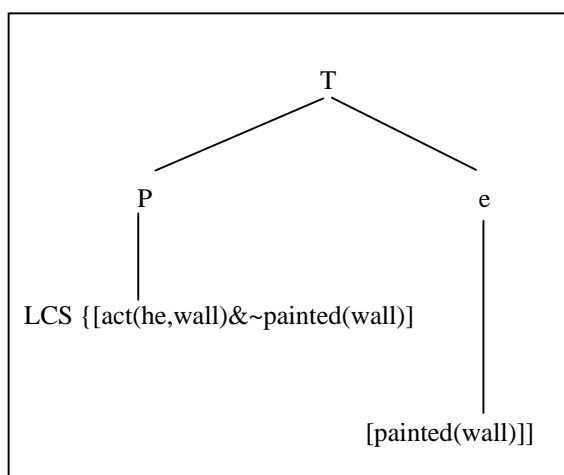


Figure 3: *paint* LCS (resultative non expressed)

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

2.1 LCS deficitary predicates

Let us now consider the following sentence:

- (4) Ele tornou a Maria feliz.
 (“He made Mary happy”)

The LCS associated to it seems to be (iii) and not (iv):

- (iii) [[act(ele,Maria)&~feliz(Maria)],
 [feliz(Maria)]]
 (“[[act(he,Mary)&~happy(Mary)],
 [happy(Mary)]]”)
- (iv) [[act(ele,Maria)&~tornada_feliz(Maria)],
 [tornada_feliz(Maria)]]
 (“[[act(he,Mary)&~made_happy(Mary)],
 [made_happy(Mary)]]”).

This means that *Q* is instantiated just with the resultative. The absence of the resultative induces ungrammaticality, as expected:

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

Along the same lines of Marrafa (1993) and Marrafa (1999), verbs like *tornar* are defended here to be LCS deficitary, in the sense they do not include in their denotation the set of content properties of the telic state of their LCS, as stated below:

Informal def.:
 $\forall v((\text{verb}(v), \exists \epsilon, \text{LCS_of_}v(\epsilon), \exists e, \text{telic_state}(e), e \subset \epsilon, \exists \pi, \text{set_of_semantic_features_of}(\pi, e), \pi = \emptyset) \Rightarrow \text{LCS_deficitary}(v))$

Since that set is empty, the LCS cannot bear an appropriate interpretation. A syntactic structure that projects an anomalous LCS is, then, previewed to be ruled out (it does not satisfy the requirement of full interpretation).

In this case, the resultative fills the gap of the LCS of the verb (cf. the contrast between (4) and (5)).

Therefore, these facts show that the representation of the predicates at issue has to include information regarding the telic expression.

3 WordNet specifications

WordNet.PT is being developed in the EuroWordNet framework (Vossen, 1999), which in turn follows the same basic lines of Princeton WordNet (Fellbaum, 1998) with regard to the individual WordNets.

The basic units of WordNets are concepts, which are represented by sets of synonyms (synsets). A synset contains the set of lexicalizations for a given concept.

The meaning of a lexical unit is derived from its relations with the other members of the same synset and with other synsets.

According to the analysis presented here, 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. Taking the synset of *tornar* as an example, a simple and plausibly solution is proposed here, in an abbreviated Attribute-Value Matrix (AVM):

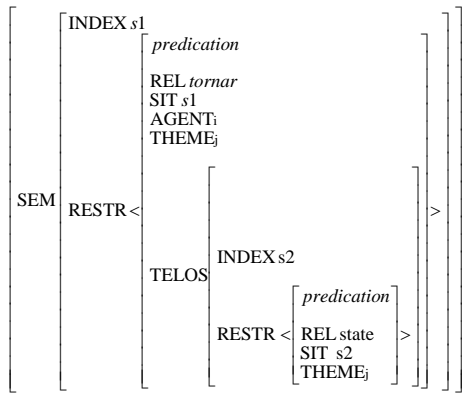


Figure 4: *tornar* feature description

The telic expression of the predicate is represented by a feature structure description that partially specifies the semantic restrictions (RESTR) imposed by the verb. The value of RESTR is a list of conditions that have to be satisfied by the situation *s1*, which corresponds to the main predication. Those restrictions include the attribute TELOS, which stands for the entailed result. The value of TELOS includes a list of restrictions (the value of the lowest RESTR) regarding the embedded situation, *s2*, which includes a state (cf. REL *state*), the telic state, that affects the THEME.

Verbs like *entristecer* (“make sad”) 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 information specified for the superordinate of that verb, a new relation – the telic state relation – has to be included in the set of the internal relations of WordNets.

The existing sub-event relation is not specific enough to account for the facts discussed. It stands for lexical entailment involving temporal proper inclusion. It does not account for the geometry of the event.

On the contrary, the telic state relation regards the atomic sub-event that is the ending point of the global event.

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. Let us examine some examples:

- (6) a. Ela entristeceu o João.
 (“She made_sad John”)

- b. Ela tornou o João triste.
 (“She made John sad”)
- c. *Ela entristeceu o João triste.
 (“She made_sad John sad”)

As shown, *entristecer* can be paraphrased by *tornar triste* and cannot co-occur with *triste*.

Linking *triste* to *entristecer* by the telic state relation, we can also capture the relation of the TELOS value of the superordinate with the telic state incorporated in the troponym, as represented below:

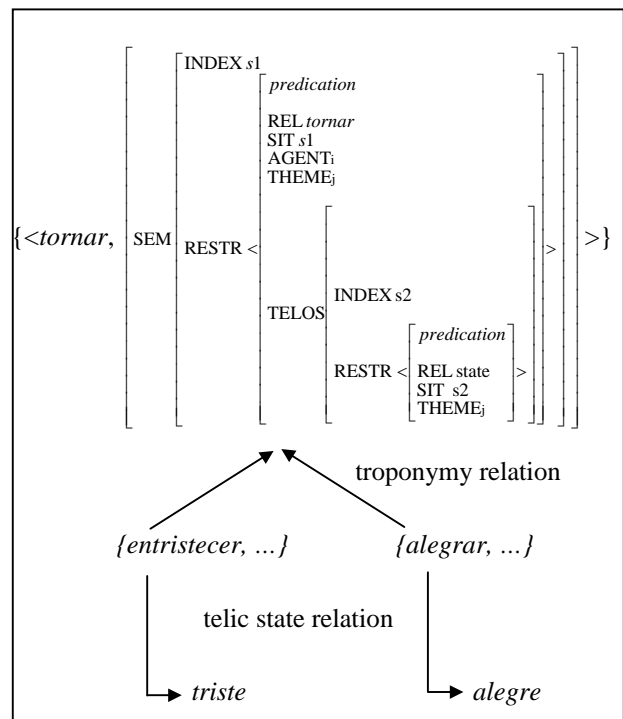


Figure 5: *tornar* troponyms and telic state relation

This representation accounts for the relevant analysis argued for here: (i) the semantics of verbs like *tornar* involves an underspecified telic feature, which should be represented in the lexicon, in general, and in WordNets, in particular; (ii) the troponyms of such verbs are assigned a specific value for that feature, which can be captured by relating them to the adjectives that can excoreporate the corresponding information, or, in other words, can instantiate the underspecified telic feature of their superordinate.

Conclusion

The proposal presented in this paper has a strong empirical motivation.

As argued for, a verb like *tornar* is just the lexicalized part of a concept, which also involves non-lexicalized information.

Since concepts are the basic units of WordNets, their descriptive adequacy cannot be preserved if concepts are not fully represented.

Therefore, to represent the concepts associated to the verbs at issue implies to encode non-lexicalized information, as defended.

Additionally, the specification of this kind of information associated to the encoding of the telic state relation, also proposed here, makes it possible to build more fine grained subnets for the resultative predicates. In a certain sense, the global proposal presented here contributes to the enhancement of the expressive power of WordNets.

The next step of this work is to implement this approach in WordNet.PT.

Implementation has to take into account the possible use of WordNet.PT in computational linguistics and language engineering applications as well as its on-line direct exploitation.

Although this seems to be a non-trivial task, it is strongly justified by linguistic evidences.

References

- Fellbaum (ed.) (1998). WordNet: An Electronic Database. MA: The MIT Press.
- Marrafa, P. (1993). Predicação Secundária e Predicados Complexos – Modelização e Análise. PhD. Dissertation. Lisbon: University of Lisbon.
- Marrafa, P. (1999). Predicados Télicos Descontínuos: análise e computação. In P. Marrafa e A. Mota (orgs.), *Linguística Computacional: investigação fundamental e aplicações*. Lisbon: APL/Edições Colibri, 1999, 173-189.
- Marrafa, P. (2003). The Representation of Complex Telic Predicates in WordNets: The Case of Lexical-Conceptual Structure Deficitary Verbs”. In Proc. of Convergences ‘03’, International Conference on the Convergence of Knowledge, Culture, Language and Information Technologies, Alexandria, Egypt: Library of Alexandria.
- Pustejovsky, J. (1991). The Syntax of Event Structure. *Cognition*, 41, 47-81.
- Schmitt, C. (1996). Aspect of the Syntax of Noun Phrases. PhD. Dissertation. University of Maryland.
- Verkuyl, H. (1972). On the Compositional Nature of Aspects. Dordrecht, D. Reidel.
- Vossen, P (1999). EuroWordNet: General Document. University of Amsterdam.
- Wechsler, S. (2001). An Analysis of English Resultatives Under the Event-Argument Homomorphism Model of Telicity. In Proceedings of the 3rd Workshop on Text Structure. University of Texas.