Lexical-Conceptual Relations as Qualia Role Encoders

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 show how wordnets can be used for building computational lexica that support generative processes accounting for phenomena such as the creation of meaning in context. We propose the integration of qualia information in wordnets through the association of lexical-conceptual relations to qualia roles, in what is a simple and low cost procedure, as it makes use of information already encoded in wordnets. This association between lexical-conceptual relations and qualia aspects allows us to describe the qualia structure of lexical items in a consistent way, without any loss of information and with the advantage of identifying the semantic predicates that can be values of qualia roles.

Key words: wordnets, qualia information, lexical-conceptual relations, generative lexica

1 Introduction

Wordnets are electronic databases developed along the general lines of the so-called Princeton WordNet, [1,2] containing nouns, verbs, adjectives, and adverbs. This database is structured as a network of relations between *synsets* (sets of synonymous word forms). Although originally developed as an experiment on the organization of the mental lexicon, WordNet has been widely used as a lexical base in NLP applications.

In this context, the need for finer-grained lexical specifications allowing computational systems to deal automatically with various complex linguistic phenomena in a general and systematic way has been pointed out by many different authors. The association of syntactic and semantic information to the WordNet model, in particular, has been object of research since its appearance (see for instance [3,4,5] or [6], among others).

One of the major achievements of research on WordNet enhancement is EuroWordNet [7], a resource reflecting research both on lexical semantics and on wordnets as lexical resources for NLP applications. Such research resulted in the definition and implementation of a wider variety of lexical-conceptual relations than the set used in the Princeton WordNet, focusing on more comprehensive lexical-conceptual relations and on cross-POS relations.

In this paper we pursue these lines of investigation, building on research developed under the scope of WordNet.PT [8] (henceforth, WN.PT) to model the semantic and syntactic properties of lexical items, while providing the relevant informational structures describing the nature of lexical meaning. Although we set off from Portuguese data,

Petr Sojka, Aleš Horák, Ivan Kopeček and Karel Pala (Eds.): TSD 2010, LNAI 6231, pp. 26–33, 2010.

we claim that our approach is language independent and widely extendable to other languages.

The relevance of such work lies on the fact that lexical items are the basic building blocks in language, being involved in a large diversity of syntactic and semantic phenomena. Being so, in order to account for this behavior, as well as for the polymorphic properties and creative use of lexical items, the need for defining lexical entries as complex informational structures becomes apparent (see [9] and [10], among others). The Generative Lexicon [11,9] (henceforth, GL) is one of the most illustrative examples of this perspective on lexical modeling. In this framework, the Lexicon is a complex system constituting a crucial part of the grammar of natural languages. Lexical units are represented as informational structures, organized according to a finite set of rules which allow for accounting for the creation of meaning in context and for describing the relation between syntax and semantics concerning several relevant linguistic phenomena.

The different levels of representation considered in GL set the grounds for information sharing between lexical items, whereas the generative mechanisms described in this framework assure a coherent and recursive codification of the information. In this paper we focus on the first part of the problem, i.e. on contributing to an accurate description of lexical items so that such representations can serve as a suitable input to the aforementioned generative mechanisms.

Concurrently, computational relational lexica such as wordnets straightforwardly provide us with the grounds for a lexical inheritance device allowing us to encode information sharing between lexical units. Hence, this paper combines these features of wordnets and GL, this way contributing to richer lexica through a parsimonious strategy of lexical modeling.

2 GL Structures and Wordnets

Having motivated the integration of additional information in the WordNet model for increasing both its accuracy and usability as a lexical base for NLP applications, in this section we present our proposal for using information in wordnets as encoders of GL structures, more specifically for using the relations encoded in WN.PT to obtain qualia information.

2.1 Qualia Structure

Qualia structure is the level of representation in which the semantic content of a lexical item is encoded, through the properties and events that define it. There are four basic qualia roles, determining the lexical-semantic structure of lexical items:

- CONSTITUTIVE: expressing the relation between an object and its constituent parts;
- FORMAL: representing the features which distinguish an object within a larger domain;
- TELIC: stating the purpose and function of an object;
- AGENTIVE: enumerating the factors involved in the origin or 'bringing about' of an object.

In terms of the linguistic characterization of lexical items, qualia structure establishes the set of semantic restrictions introduced by a word in context. Although these semantic restrictions can sometimes undergo semantic change – in metaphoric contexts for instance –, they explain ill/well-formation contrasts between certain structures as well as available readings. Let us briefly go through some examples to illustrate how qualia values have impact on the well-formedness of sentences.

1. a. Anne scaled the fish.
   b. *Anne scaled the chicken.
   c. *Anne plucked the fish.
   d. Anne plucked the chicken.
2. a. John started the house.
   b. = John started to build the house.
3. a. John started the book.
   b. = John started to write the book.
   c. = John started to read the book.

In (1) above, the ill-formation of (1)b and (1)c can be explained by the different values for the **constitutive role** of *chicken* and *fish*: *chicken* have *feathers* and *fish* have *scales*. Such values, in combination with the semantic restrictions introduced by the main verbs in (1) – both *to scale* and *to pluck* have shadow arguments incorporated in their semantics: *scale* and *feather*, respectively –, justify the syntactic contrasts between (1)a and (1)b, and between (1)c and (1)d.

Besides accounting for grammaticality contrasts, as illustrated in (1), qualia roles can also explain available readings. Differently from the previous example, in (2) and (3) there are no grammaticality contrasts. In this case the challenge consists in explaining how the readings in (2)b, (3)b and (3)c are derived. Qualia structure enables the association of particular properties and activities to nouns, which in turn provides the verb selecting the NP in which these nouns occur with the information for contextualising its sense. *house*, in (2), is associated to a **building event** in the **agentive role** of its qualia structure and this is how this event is made available to the reading presented in (2)b. In (3), we have two different possible readings of (3)a, in (3)b and (3)c, because *book* is associated to two events in its qualia structure: a **writing event** in the agentive role and a **reading event** in the telic role, thus making both events available.

Based on data such as these, [9] defines a set of generative mechanisms which take the information identified above to derive available readings. However, not all lexical items associated to events in their qualia structure display the linguistic behavior illustrated in (3). If we take a word like *pen*, a writing implement, we see that not all the readings in (3) are available for this lexical item, although *pen*, just like *book*, is associated to a **constructing event** in the **agentive role**, and to a **writing event** in the **telic role**. A sentence like *John is starting the pen* has a single available reading: John is starting to construct (by assembling parts or carving, for instance) a *pen*. The reading in which he is starting to write with a pen is never obtained. This clearly poses a problem to [9]'s generalization presented above. Data like this indicate that further specifications are called for, for instance in the form of qualia subtyping, or some other strategy allowing generative mechanisms to account for these contrasts. But this is a question that is clearly outside the scope of the work depicted in this paper, which we are currently addressing independently. Despite counter examples like the one discussed above, the relevance of qualia information in the characterization of the semantic contribution of lexical items is

nonetheless undeniable, the impact it has on their linguistic behavior being what still calls for further analysis and specifications.

2.2 Qualia Information in Wordnets

Qualia information reflects the semantic content of the lexical item described. Being so, specifying such information in relational models of the lexicon is entirely consistent with the nature of such models. As noted in the literature (see [12,13], among others) WordNet structuring relation – *hyperonymy/hyponymy* –, for instance, refers to the **formal quale** of a lexical item, *meronymy* relations to the **constitutive quale**, *cause* relations to the **agentive quale**, and so on.

Qualia information has also been noted to be expressed indirectly in wordnets at other levels. [13], for instance, in the context of research on metaphorical uses of lexical items, puts forth a proposal consisting on automatically extracting **agentive** and **telic** information from the glosses in WordNet.

As to the EWN model, it reflects to some extent the notion of qualia information in its structure, as GL qualia structure was used for defining and building its top-ontology [14]. This top-ontology reflects a taxonomic approach to qualia structure, as it does not allow for simultaneously accounting for multiple aspects of the meaning of a specific lexical item (see [11]).

Pursuing this taxonomic approach to qualia information in EWN, [12] puts forth a proposal for distinctively modeling taxonomic hyponyms (those associated to 'canonical' *hyperonymy* relations and corresponding to the **formal quale** of the lexical item) and orthogonal hyponyms (split according to the other three qualia dimensions) in Danish wordnet. This strategy results in the distinction of *constitutive hyponyms* (*idiot* 'idiot' or *geni* 'genius'), *telic hyponyms* (*garvestof* 'tanning agent' or *flugtbil* 'getaway car') and *agentive hyponyms* (*fodgĂŚnger* 'pedestrian' or *cyklist* 'cyclist') (examples taken from [12]: 5). However, there are some stebacks to this taxonomic approach to qualia information.

In fact, although [12] accounts for phenomena such as compatible co-hyponymy (see [15,16]), it continues to lack the ability to mirror the fact that lexical items can be, and often are, simultaneously characterized by multiple qualia.

Reflecting different goals and strategies, the approaches discussed above reinforce the acknowledgement of a general need for being able to obtain more complete informational structures from lexical resources, while showing at the same time that relational models of the lexicon like wordnets convey, more or less directly, a significant amount of relevant information regarding the internal semantic content of lexical items.

2.3 Qualia Information in WordNet.PT

In previous sections we showed how wordnets are particularly well placed to satisfy such a need: their network of relations, in combination with additional information also encoded in these resources, make informationally rich lexical entries available; at the same time, the hierarchy of relations provides a natural inheritance structure, straightforwardly allowing for information sharing between lexical entries, and thus for

an economical lexical encoding. We motivated the relevance of using qualia structure to characterize lexical items, and went through some proposals for integrating this level of representation in wordnets, many of which with promising results. However, most of the proposals available in the literature are partial and often oriented towards solving particular problems of a specific application, hence not addressing the question of lexical representation in a systematic and global perspective. Doing so is our goal.

Lexical-conceptual relations reflect intrinsic or prototypical properties that characterize a given concept. Having this in mind, we analyzed the relations available in WN.PT to determine whether these can be matched to express qualia properties, and which properties are these. By doing so, we established a correspondence between the lexical-conceptual relations considered in WN.PT and qualia roles, besides the general acknowledged correspondence between *meronymy* and **constitutive** properties, and *hyponymy* and **formal** properties.

In the table presented below we show how the relations in wordnets developed in the EWN model can be used to systematically express all qualia roles. Achieving this requires only the definition and codification of two new relations in wordnets. As aforementioned the *is hyponym of* and *has as part* relations express the generic cases of **formal** and **constitutive** properties, respectively. Being so, it is only necessary to define relations to express the generic case of **agentive** and **telic** properties. Thus, we introduce the relations *results/originates from* and *has as function/goal* to fill in this gap. Their respective counterparts – *results in/originates* and *is function/goal of* –, however, reflect **formal** properties.

4. RESULTS/ORIGINATES FROM/RESULTS IN/ORIGINATES relation

   {synset}1 RESULTS/ORIGINATES FROM {synset}2 and {synset}2 RESULTS IN/ ORIGINATES {synset}1

5. HAS AS FUNCTION/GOAL/IS FUNCTION/GOAL OF relation

   {synset}1 HAS AS FUNCTION/GOAL {synset}2 and {synset}2 IS FUNCTION/ GOAL OF {synset}1

The Table 1 on pages 31–32 summarizes our analysis, grouping WN.PT relations, in the second column, that correspond to values of the same qualia role. Given the quality of the information stated through each specific relation, the two directions of a same relation can refer to different qualia aspects, as can be observed above. Also these relations[1] constitute values of qualia roles if, and only if, they are canonical relations defining the meaning of a lexical item, i.e. if not marked by the reversed feature (indicator of inverse relation by default).

[1] As some of the labels used in WN.PT were changed to be more transparent for non-specialist users, we present their corresponding relations in EWN. Regarding WN.PT-specific relations, their relevance has been argued for in [17,18].

WN.PT	WN/EWN	EXAMPLE
Formal role		
is hyponym (subtype) of	*has hyperonym*	{animal}N→ {living being}N
has as a characteristic	–	{carnivorous}Adj→ {shark}N
is related to	–	{marine}Adj→ {sea}N
has manner	*has manner*	{sway}V→ {smoothly}Adv
is part of	*has holonym*	{brick}N→ {wall}N
is individuated part of	*has holo part*	{motor}N→ {machine}N
is portion of	*has holo portion*	{drop}N→ {liquid}N
is member of	*has holo member*	{member}N→ {club}N
is substance/material of	*has holo made of*	{fiber}N→ {tissue}N
is sublocation of	*has holo location*	{palm}N→ {hand}N
is subevent of	*is subevent of*	{inhale}V→ {breathe}V
is telic subevent of	–	{suspicious}Adj→ {suspect}V
characterizes with regard to	*is value of*	{big}Adj→ {size}N
sets attribute value to	–	{tall}Adj→ {plus}Adv
co relates with	*co_role*	{roof tile}N→ {roof lattice}N
causes	*causes*	{kill}V→ {die}V
has as result	*involved result*	{build}V→ {building}N
has as location	*involved location*	{swim}V→ {liquid}N
has as source location	*involved source direction*	{unbox}V→ {box}N
has as goal location	*involved target direction*	{emigrate}V→ {foreign country}N
results in/originates	–	{architecture}N→ {blueprint}N
is function/goal of	–	{learn}V→ {pupil}N
Telic role		
has as function/goal	–	{pupil}N→ {learn}V
has telic subevent	–	{sadden}V→ {sad}Adj
is the instrument used for	*role instrument*	{pen}N→ {write}V
is the location for	*role location*	{pool}N→ {swim}V
is the source location of	*role source direction*	{start line}N→ {race}N
is the goal location of	*role target direction*	{prison}N→ {incarcerate}V
is transformed in	*co_patient result*	{tadpole}N→ {frog}N
is used to obtain	*co_instrument result*	{coffee grinder}N→ {coffee powder}N
acts to obtain	*co_agent result*	{lumberjack}N→ {wood}N
relates as agent with the object	*co_agent patient*	{lumberjack}N→ {tree}N
relates as object with the agent	*co_patient agent*	{brick}N→ {mason}N
uses as instrument	*co_agent instrument*	{mason}N→ {trowel}N
is used as instrument by	*co_instrument agent*	{trowel}N→ {mason}N
relates as object with the instrument	*co_patient instrument*	{mortar}N→ {trowel}N
relates as instrument with the object	*co_instrument patient*	{trowel}N→ {mortar}N

WN.PT	WN/EWN	EXAMPLE
Agentive role		
results/originates from	–	{blueprint}N→ {architecture}N
is caused by	*is caused by*	{find}V→ {search}V
is the result of	*role result*	{roastbeef}N→ {roast}V
results from the transformation of	*co_result patient*	{frog}N→ {tadpole}N
results from the use of	*co_result instrument*	{coffee}N→ {coffee pot}N
results from the action of	*co_result agent*	{bread}N→ {baker}N
Constitutive role		
has as part	*has meronym*	{wall}N→ {brick}N
has as individuated part	*has mero part*	{car}N→ {wheel}N
has as portion	*has mero portion*	{liquid}N→ {drop}N
has as member	*has mero member*	{club}N→ {member}N
has as substance/material	*has mero madeof*	{coffee}N→ {caffeine}N
has as sublocation	*has mero location*	{hand}N→ {palm}N
has subevent	*has subevent*	{breathe}V→ {inhale}V
has telic subevent	–	{sadden}V→ {sad}Adj

Table 1. WN.PT relations as qualia roles encoders

As made apparent in the table presented above, we claim that the integration of qualia roles in wordnets is a straightforward and low cost procedure since lexical-conceptual relations in wordnets already encode intrinsic or prototypical properties that characterize the concept lexicalized by each synset. This way, associating the relevant relations to a given qualia aspect allows us to encode the qualia information in a coherent and consistent way, without any loss of information. This approach also has the advantage of determining which semantic predicates can be values of qualia roles.

3 Final Remarks

In this paper we show how the specification of fine-grained lexical descriptions in computational relational lexica provides the grounds for accounting for several lexical semantic phenomena. Our strategy contributes to enhancing wordnets usability as computational lexica supporting generative processes to account for phenomena such as the creation of meaning in context.

In our proposal, the association of qualia roles to synsets in wordnets is achieved very straightforwardly. This procedure, which hardly adds any costs to the effort in wordnets development, as it only involves the definition of two additional relations, allows for determining the semantic predicates that can be values of qualia roles in a coherent and consistent way. This strategy provides us with the mechanisms for describing and encoding crucial semantic properties of lexical items, providing the relevant information at the lexical level.

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, 353–378. The MIT Press, Cambridge, MA (1998)
6. Agirre, E., Martinez, D.: Integrating Selectional Preferences in WordNet. In: Proceedings of the First International WordNet Conference. Mysore (2002)
7. Vossen, P. (ed.): EuroWordNet – A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht, Netherlands (1998)
8. Marrafa, P.: The Portuguese WordNet: General Architecture and Semantic Internal Relations. DELTA, Brasil (2002)
9. Pustejovsky, J.: The Generative Lexicon. The MIT Press, Cambridge, MA (1995)
10. Jackendoff, R.: Semantic Structures. The MIT Press, Cambridge, MA (1990)
11. Pustejovsky, J.: The Syntax of Event Structure. Cognition 41, pp. 47–81 (1991)
12. Pederson, B., Sørensen, N.: Towards Sounder Taxonomies in WordNets. In Ontolex 2006, pp. 9–16. Genova, Italy (2006)
13. Veale, T.: Qualia Extraction from WordNet. In: Proceedings of the 3rd International Workshop on Creative Systems, The 2003 International Joint Conference on Artificial Intelligence. Acapulco, Mexico (2003)
14. Vossen, P. (ed.): EuroWordNet – A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht, Netherlands (1998)
15. Mendes, S., Chaves, R. P.: Enriching WordNet with Qualia Information. In: Proceedings of the Workshop on WordNets and Other Lexical Resources at NAACL 2001 Conference, pp. 108–112. Pittsburgh (2001)
16. 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)
17. 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)
18. Mendes, S.: Syntax and Semantics of Adjectives in Portuguese: Analysis and Modeling. Ph.D. dissertation, University of Lisbon, Lisbon (2009).