Integrating Prepositions in Wordnets: Relations, Glosses and Visual Description

Raquel Amaro
NOVA CLUNL – Faculty of Social Sciences and Humanities – Universidade NOVA de Lisboa
& CLUL – Faculty of Arts – Universidade de Lisboa
Avenida de Berna 26-C 1069-061 Lisbon, Alameda da Universidade 1600-214 Lisbon
raquelamaro@fcsh.unl.pt

Abstract

Lexicology and lexicon models are necessarily concerned with content words, being grammatical and functional categories often set aside. Currently, however, lexicographers work for real needs and, in a NLP perspective, the nature of computational lexicons reflects a necessary match between what we know about the mental lexicon and what we need to encode about the set of words of a given language. Prepositions, in many languages, combine these two prototypes of words – lexical and functional –, as they can have full meaning or serve solely as structural aids. Based on the analysis of Portuguese prepositions related to the expression of movement, this paper describes how the integration of prepositions in wordnets is possible and quite easy, requiring mainly the linguistic adaptation of the tests and conditions that mediate the establishment of the relations of synonymy, antonymy, hyperonymy and cause. In what concerns lexicographic strategies, the integration of prepositions shows the difficulty of establishing equivalences between the concepts denoted by prepositions in different languages, as well as the difficulty of using glosses in natural language to describe their meaning. The use of visual information may obviate this issue, while posing issues on implementation.

Keywords: prepositions, wordnets, visual information

1. Introduction

One of the first appointed differences between the theoretical study of the lexicon – lexicology – and the crafts of making lexical resources – lexicography – is the set of words that is considered relevant for the first and that has necessarily to be described for the second (Crystal, 1995). Lexicology and models of the mental lexicon are essentially concerned with so-called content words, being grammatical or functional categories often set aside (Klein, 2001). Currently, however, lexicography works considering real users' needs, and often focuses its strategies for NLP purposes (Gouws, 2004). The nature of modern computational lexicons can thus be described as the perfect or necessary match between what we know and figure about the mental lexicon, considering conceptual and semantic properties, and what we need to encode about *all* the words of a given language in order to make a lexicon useful, whether this information is functional or not.

Prepositions, in many languages, perfectly combine these two prototypes of words, as they aggregate items that undoubtedly have meaning and items that serve solely as structure markers (Hernández-Pastor & Periñán-Pascual, 2016). In fact, accounting for prepositions in wordnets is listed as one of the existing challenges for the development and application of wordnets (Workshop Challenges for Wordnets, Bond & Piasecki, 2017) and prepositions are included at least in Bulgarian WordNet[1] (Dimitrova et al., 2014), although not establishing many (if any) relations with other nodes in the net (test, for instance, след in http://dcl.bas.bg/bulnet/).

Based on the analysis of Portuguese prepositions related to the expression of movement, this paper further explores the integration of prepositions in wordnets, showing how they can be modeled, which relations serve to encode their meaning and/or function, and how glosses and crosslinguistic equivalences can be inadequate to provide a clear grasp of the concept prepositions denote.

In the next sections, we review different approaches to prepositions, as well as further explore the motivations for integrating them in wordnets (section 2); we present our proposal for modeling prepositional concepts in wordnets, considering semantically full prepositions and argument-marking prepositions (section 3); we discuss the issues concerning semantic description, crosslinguistic equivalence, glosses and visual information (section 4); and lastly we present our final remarks (section 5).

2. Prepositions

Prepositions are fairly common in natural languages, and their treatment is of high impact in NLP tasks (Hernández-Pastor & Periñán-Pascual, 2016).

The analysis of prepositions has many times been considered under the scope of the relation between prepositions and the nouns they co-occur with (*on Thursday*, *in the morning*), or the verbs that select them (*dream of*, *care about*) (Veerspoor, 1997), directly related to cases where the semantic contribution of prepositions to the meaning of the phrase or sentence seems or is void. In fact, this aspect of the combination of prepositions with other lexical items is what usually makes them difficult to be computationally processed and in many cases disambiguated (*Ele sonhou com a irmã*. = he dreamt of his sister; *Ele morou com a irmã*. = he lived with his sister).

However, many prepositions display a constant semantic content, which is crucial for the determination of the meaning of prepositional phrases and sentences (*since February* vs. *until February*; *at home* vs. *from home*) (Bannard & Baldwin, 2003).

In what concerns their semantic description, research on prepositions has taken three main directions:

i) large-scale symbolic accounts of preposition semantics (Dorr, 1997's 497 senses of English transitive and intransitive prepositions formalized in a lexical conceptual semantics framework; Canesson & Saint-Dizier, 2002's description of French prepositions in PrepNet; Jensen & Nilsson, 2003's description of prepositions through a finite set of universal binary role relations; Srikumar & Roth, 2013's set of relations established by prepositions);

ii) prepositional phrase disambiguation (O'Hara & Wiebe, 2003's account of prepositional phrases tokens according to case-roles, or McShane et al., 2005's

[1] http://dcl.bas.bg/en/resursi/wordnet/

ontological semantic analyzer for disambiguating homonym prepositions); and

iii) distributional accounts of preposition semantics (such as Bannard & Baldwin, 2003's work on particles and transitive prepositions for a valence-conditioned classification of English prepositions).

In many of these cases, as well as in more conventional approaches such as traditional normative grammars, semantically full prepositions are commonly organized according to notions such as purpose, goal, location, temporality, cause, etc., across languages such as French (Saint-Dizier, 2008), English (Jensen & Nilsson, 2003) or Portuguese (Cunha & Cintra, 1984). According to these works, the semantic value of prepositions can be compared to those of other POS.

2.1 Related work

Several researchers have studied prepositions and the ontological organization of prepositions, adopting a similar approach to that of WordNet, given that prepositions are described according to their conceptual properties.

PrepNet (Saint-Dizier, 2005, 2008) is such an example. PrepNet is a database for prepositions structured in two levels: the abstract notion level (conceptual level, language independent) and the language realization level (which deals with the realizations for various languages). Abstract notions are organized in a first stage that characterizes the semantic family of the notions (localization, manner, quantity, company, etc.), a second stage that accounts for the different facets of each semantic family (source, destination, or via, for instance), and a third stage that captures the modalities of a given facet (such as basic manner, manner by comparison, manner with a reference point, etc.). The language representation level includes syntactic frames and semantic and domain restrictions.

PrepNet approach to the representation of the meaning of prepositions can be used as the base for integrating prepositions in wordnets, since the abstract notion can help in the establishment of prepositional higher nodes in wordnets as well as in the establishment of the sets of hyponyms. However, we observed that the facets and modalities expressed by prepositions are not necessarily the same in every language.

As mentioned before, accounting for prepositions in wordnets is listed as one of the existing challenges for the development and application of wordnets (Workshop Challenges for Wordnets, Bond & Piasecki, 2017). Nevertheless, in current days, not many of these lexical resources include prepositions. In fact, a survey of the information displayed on the presentation pages of each of the wordnets included in the Global WordNet Association list of wordnets in the world[2] show us that, from the 124 resources listed, 30 do not state the POS considered (from which we assume they encode the same POS treated in Princeton WordNet: nouns, verbs, adjectives and adverbs); 42 state they do not consider prepositions; 28 do not present webpages for the resources; 21 do not have functional webpages or are in maintenance and 2 do not provide information in English. Only the Bulgarian WordNet[3] (Dimitrova et al., 2014) states the inclusion of prepositions (as well as other functional words such as conjunctions), although these seem to be somewhat loose in the net, according to the observation of the nodes for some prepositions in http://dcl.bas.bg/bulnet/.

Following Amaro (2009), the motivation for integrating prepositions in wordnets comprises two sets of reasons:

i) theoretical (semantic) reasons: prepositions denote notions such as cause, location, temporality, etc., as demonstrated by several earlier and current studies;

ii) practical (functional) reasons: even semantically empty prepositions, which are idiomatic, add information useful for NLP purposes, contributing to the usability and relevance of wordnets.

The following sections illustrate further these aspects.

2.2 Dataset

The set of prepositions considered in this paper was compiled from prepositions commonly used in the expression of movement in Portuguese (Amaro, 2009), such as *de* (≈ from), *a* (≈ to), *até* (≈ until/to), *para* (≈ to, in the direction of, towards), *por* (≈ through), *em* (≈ in), *sobre* (≈ on top of, over), *entre* (≈ between), etc.

We also considered multiword expressions such as *acima de* (≈ above), *atrás de* (≈ behind), *ao lado de* (≈ next to, close to), *por baixo de* (≈ under), *em direção a* (≈ to, towards, in the direction of), and so on, since these fixed expressions behave like prepositions (see Cunha & Cintra, 1984; Baldwin et al., 2009). These correspond to multiword expressions that refer to prepositional meaning or have a prepositional function and are expressions that

i) do not undergo inflection, internal modification or word order variation, i.e. "words with spaces" (Sag et al., 2002):

(1) a. Ele colocou o livro mesmo ao lado da jarra.
he placed the book exactly at.the side of the vase (≈ next to)

b. *Ele colocou o livro mesmo aos lados da jarra.
he placed the book exactly at.the sides of the vase

c. *Ele colocou o livro ao lado mesmo da jarra.
he placed the book at.the exactly side of the vase

d. *Ele colocou o livro ao lado esquerdo da jarra.
he placed the book at.the left side of the vase

e. *Ele colocou o livro mesmo do lado à jarra.
he placed the book exactly of.the side at.the vase

ii) can often be replaced by simple prepositions, as illustrated in (2):

(2) a. The mouse ran in the direction of/to the table.

b. The man stood quiet in front of/before the judges.

3. Modeling prepositions in WordNet

3.1 Semantically full prepositions

Diverging from the approaches for modeling the semantics of prepositions in a deeper fashion and with specific sets of relations (Saint-Dizier, 2008; Srikumar & Roth, 2013, Schneider et al., 2015), we demonstrate that it is possible to model prepositions with full meaning (i.e., prepositions whose semantic content is crucial for the determination of the meaning of phrases, such as *before noon* vs. *after noon*) through relations already available in WordNet model, namely synonymy, antonymy, hyperonymy/hyponymy and cause/is caused by.

[2] http://globalwordnet.org/wordnets-in-the-world/

[3] http://dcl.bas.bg/en/resursi/wordnet/

These relations correspond to the ones defined in Fellbaum (1998) and Vossen (2002) and require only the adaptation of the tests and definitions to the specificity of this POS. Specifically, prepositions require a complement (usually a Noun Phrase) and cannot be linguistically tested without considering the entire Prepositional Phrase. Although based on the studied prepositions for Portuguese, the definitions and tests presented here are expected to serve for any language. For that reason, whenever possible, English examples will be used to illustrate the tests.
The adapted definitions and tests are presented below[4].

(3) Synonymy relation

Definition:
P1 is synonym of P2 in C iff
if P1 then P2 and if P2 then P1

Test:
if the mouse is *under* the table then the mouse is *underneath* the table, and if the mouse is *underneath* the table then the book is *under* the table: **True**

under	is synonym of	*underneath*
underneath	is synonym of	*under*

--> {under, underneath}$_{Prep}$

Synonymy relations between prepositions, in Portuguese at least, are not very productive, even considering the synonymy notion bound to a given context. However, they still exist, in particular between atomic and multiword prepositions.
Prepositional synsets can also be related to each other by antonymy.

(4) Antonymy relation

Definition:
P1 is antonym of P2 iff
i) P1 and P2 are co-hyponyms;
ii) P1+NPi/VPi is the opposite of P2+NPi/VPi and P2+NPi/VPi is the opposite of P1+NPi/VPi

Test 1:
i) *under* and *on top of* are both hyponyms of *in, at*: **True**
ii) *under* the table is the opposite of *on top of* the table and *on top of* the table is the opposite of *under* the table: **True**

Test 2 (negation):
if P1+NPi/VPi then not P2+NPi/VPi and
if P2+NPi/VPi then not P1+NPi/VPi
if the cat is *under* the table, then the cat is not *on top of* the table: **True**
if the cat is *on top of* the table, then the cat is not *under* the table: **True**

under	is antonym of	*on top of*
on top of	is antonym of	*under*

[4] The notations used in the definitions and tests correspond to: P = Preposition; NP = Noun Phrase; VP = Verb Phrase; AdjP = Adjectival Phrase; { } = synset/node in the net; / = or. The index *i* assures that the complements considered for P1 and P2 are the same.

Antonymy relations between prepositions, as it happens with adjectives (cf. Mendes, 2009), are quite relevant for further modeling prepositional concepts given that they allow to express opposite facets of several notions such as opposite locations with regard to a given ground object (ex.: *under* vs. *on top of*; *to inside of* vs. *to outside of* (see Figures 1 and 2)), opposite directions (ex.: *upwards* vs. *downwards*, *to* vs. *from*), opposite temporal relations (*after* vs. *before*), etc.

(5) Hyponymy/hyperonymy relation

Definition:
P2 is hyponym of P1 and P1 is hyperonym of P2 iff
i) *P2 is P1+NPi/VPi/AdjPi, but*
ii) *P1 is not P2+NPi/VPi/AdjPi*

Test 1:
under is *in*+the space below, but *in* is not *under*+the space below: **True**

{*under*}$_{Prep}$	is hyponym of	{*in, at*}$_{Prep}$
{*in, at*}$_{Prep}$	is hyperonym of	{*under*}$_{Prep}$

Test 2 (conditions for replacement and anaphora):
P2 is hyponym of P1; and
i) the complement of P1 denotes a reference that is equal or includes the reference denoted by the complement of P2;
ii) if P2 then P1, but if P2 then not P1
iii) P1 can be used as anaphoric element for P2.

under is hyponym of *in*: **True**
the room includes the table: **True**
If the mouse is *under* the table, then the mouse is *in* the room: **True**
If the mouse is *in* the room, then the mouse is *under* the table: **False**
The mouse is under the table. So, while it was in the room, nobody entered.
#The mouse was in the room. So, while it was under the table, nobody entered.

The testing for hyponymy/hyperonymy relations requires considering the inclusion relations established between the prepositional complements, following the described in Vossen (2002: 21) for hyponymy relations between nouns.
The definitions and tests proposed here show the feasibility of modeling prepositional concepts in wordnets, with some level of meaning description. Figures 1, 2, 3 and 4 present examples of hyponymy nets for Portuguese prepositions related to the expression of movement and spatial relations.
The study of Portuguese prepositions related to the expression of movement also allowed us to observe that, although seeming quite similar to prepositions indicating location, and almost seeming compositionally built, prepositional expressions denoting goal and source locations (Figures 2 and 3) do not result from the combination of prepositions denoting location, in Figure 1.
First, if these expressions were regular and compositional, the occurrence of not allowed combinations would be minimal and accidental. However, on the contrary, it is not possible to express a source or goal location using the

prepositions *de* or *para* + *em* (the top nodes of the three subtrees presented):

(6) *Ele foi de em a escola para em a rua.
he went from in the school to in the street

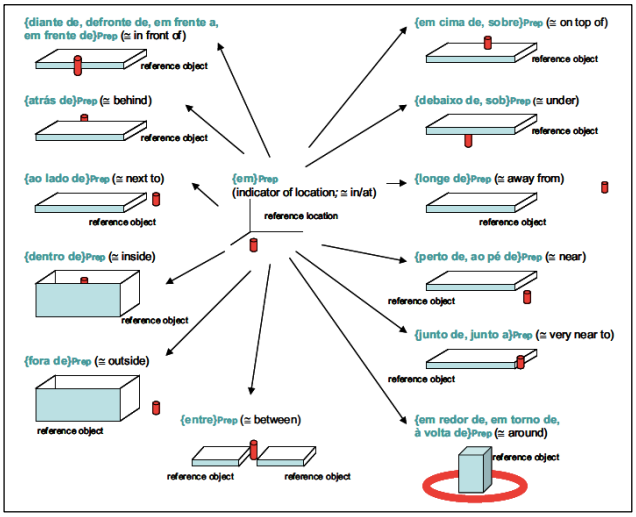

Figure 1: Hyponymy network of prepositional synsets denoting indicators of location

A closer view also reveals that several combinations of elements from the subnets presented are not possible:

(7) a. *para/de em cima de (≈ to/from on top of)

b. *para/de em baixo de (≈ to/from on under of)

c. *para/de em frente a (≈ to/from in front of)

d. para trás de/*de trás de/*em trás de (≈ to behind/from behind/in behind)

e. para debaixo de/de debaixo de/*em debaixo de (≈ to under/from under/under)

f. em torno de/*para torno de/*de torno de (≈ in around of/to around of/from around of)

However, intuitively, the concepts of location, source location and goal location seem to be strongly related. This is the case given that moving to a final location (goal) causes being in that location, and, on the contrary, moving from a given location (source) causes not being in that location. Being so, it is possible to link these concepts in wordnets through cause relations.
The definition and testing of cause/is caused by relation between prepositional nodes is presented below, as well as their application to the synsets {to}$_{Prep}$ (indicator of goal) and {from}$_{Prep}$ (indicator of source) and {in, at}$_{Prep}$ (indicator of location), for explanatory purposes.

(8) Cause/is caused by relation

Definition:
P1 causes P2 iff
P1+Ni causes/has as consequence P2+Ni, but not the converse.

Test:
a. (He moved) *to* the street causes/has as consequence (he is) *in* the street but (he is) *in* the street does not cause/have as consequence (he moved) *to* the street

{to}$_{Prep}$ causes {in, at}$_{Prep}$
{in, at}$_{Prep}$ is caused by {to}$_{Prep}$ (non-factive)

b. (He moved) *from* the street causes/has as consequence (he is) not *in* the street but (he is) not *in* the street does not cause/have as consequence (he moved) *from* the street

{from}$_{Prep}$ causes {in, at}$_{Prep}$ (negative)
{in, at}$_{Prep}$ is caused by {to}$_{Prep}$(negative) (non-factive)

In order to test the cause relation between a prepositional synset indicator of source location and another prepositional synset indicator of location, in (8)b, it is necessary to include negation, since the consequent state of moving from a given location amounts to not being in that location.
In WordNet, the negation label is used to explicitly express that a given relation does not hold. It is used to block unwanted implications, as non-inherited relations (Vossen 2002:16). The case presented here does not correspond exactly to the same situation, given that there is no prototypical relation to be inherited. The negation label is only used here for explanatory purposes.[5]

3.2 Argument-marking prepositions

One of the main reasons leading to the little attention dedicated to prepositions when it comes to their semantic content is directly related to semantically empty prepositions, that is, prepositions serving only functional or grammatical purposes. This set can be further divided in i) functional prepositions, i.e. prepositions that regularly indicate syntactic functions that do not depend on selection restrictions of specific lexical items (as, for instance, the preposition *a* in Portuguese, which regularly and invariably marks the indirect object of ditransitive verbs); and ii) argument-marking prepositions, i.e. prepositions whose only function is to mediate between a given predicate and its arguments (Sag & Wasow 1999: 157), as illustrated below for Portuguese and English:

(9) a. O rapaz gostou de cães.
the boy liked PREP dogs

b. O rapaz sonhou com cães.
the boy dreamt of dogs

c. O rapaz aproximou-se dos cães.
the boy came closer to.the dogs.

In what regards the integration of empty prepositions in wordnets, we propose that it is relevant to consider the second case since these prepositions, as illustrated in (9), concern:

i) cases in which the presence of the preposition is language dependent (9a);

[5] The relation established between {from}$_{Prep}$ and {in, at}$_{Prep}$ is that the first causes the negation of the last, and not that the relation between the nodes does not hold. For this reason, it is only possible to express this relation indirectly, linking *from* and *to* as antonyms, which motivates further the relevance of antonymy relation.

ii) cases in which the preposition choice does not correspond to the typical equivalent in other languages (9b, where the Portuguese preposition *com* corresponds to the English preposition *of*, instead of its frequent English translation *with*); and

iii) cases where the argument-marking preposition is homonym of the preposition denoting the opposite semantic content (in 9c, where the argument marking preposition *de* marks a goal location argument, whereas the semantically full preposition *de* denotes an indicator of source location).

On the contrary, truly functional prepositions can be effectively covered by syntactic rules, justifying their absence from the lexicon.

Being idiosyncratic, i.e. language dependent and not permutable by any other preposition, argument-marking prepositions are said to form a semantic component with the verb, since it is the verb+preposition that attributes case to the selected NP (see Neeleman, 1997).

Neeleman proposal results in complex lexical entries for verbs such as *gostar de* (≈ like), *sonhar com* (≈ dream of) and *aproximar-se de* (≈ go closer), for instance, and could motivate their encoding within the node for the verb form.

However, and as underlined by Godoy (2008), at syntactic level these prepositions form constituents with the selected NP and not with the verb, as illustrated in (10), (11) and (12).

(10) a. De cães, o rapaz gosta.
≈ PREP dogs, the boy likes

b. Com cães, o rapaz sonhou.
≈ of dogs, the boy dreamt

c. Dos cães, o rapaz aproximou-se.
≈ to the dogs, the boy moved closer

(11) a. O rapaz gosta de cães e ela também gosta.
≈ the boy likes PREP dogs and so likes she

b. O rapaz sonhou com cães e ela também sonhou.
≈ the boy dreamt of dogs and so dreamt she

c. O rapaz aproximou-se dos cães e ela também se aproximou.
≈ the boy moved closer to the dogs and so moved she

(12) a. O rapaz gosta de cães e de gatos.
≈ the boy likes PREP dogs and PREP cats

b. O rapaz sonhou com cães e com gatos.
≈ the boy dreamt of dogs and of cats

c. O rapaz aproximou-se dos cães e dos gatos.
≈ the boy moved closer to the dogs and to the cats

These examples show that, although required by a given verb, argument-marking prepositions do not form semantic or syntactic components with the verb that subcategorize for them: on the one hand, having no semantic content, these prepositions do not contribute to the semantic content denoted by the VP; on the other, they form syntactic constituents with the NP and not with the verb. Following Godoy's (2008) approach, we consider that these prepositions are not visible at semantic level, existing solely at syntactic level.

Argument-marking prepositions are true grammatical words and semantically empty lexical items, directly related to verbs that subcategorize them, raising the issue of how to represent these items in wordnets, since these prepositions do not denote concepts. Their inclusion in the lexicon, however, can be motivated by different reasons:

i) as idiosyncratic items, these prepositions are acquired by children in a similar process as all other lexical items, since their distribution and/or meaning do not result from the regular application of rules available in natural languages (cf. Godoy, 2008);

ii) argument-marking prepositions constitute a small and closed set of items, necessarily connected to the verbs that require their syntactic realization. So, the collection and treatment of argument-marking prepositions is always related to the collection and treatment of verbs.

iii) their representation as autonomous entries (instead of as part of verbal entries) allows for multiple linking, and avoids multiword expressions that not conform to the properties defined earlier (not undergoing inflection, internal modification or word order variation (as illustrated in (10), (11) and (12)).

These reasons, although strongly of lexicographic nature, motivate the inclusion of these items in wordnets as part of the set of prepositional items, but as extremely underspecified lexical entries. These can be related to other nodes in the net either using the conjunction label with role and involved relations (Vossen, 2012), either using specific selection relations, as proposed in Amaro (2010).

3.3 Informational gain

The cases presented clearly exemplify how the integration of prepositions in wordnets is possible, using mainly available relations with the necessary adaptations to definitions and testing conditions.

In terms of informational gains for these resources, the integration of prepositions allows, for instance, for a more complete description of the lexical items and of properties of POS, such as subcategorization properties of verbs, but also of the computational processes of the lexicon.

For instance, the integration of semantically full prepositions enables the model to represent in a more accurate way the expression of location. This is visible in two specific possibilities:

i) the automatic prediction of which specific lexical units can introduce location, source, goal, etc., considering the percolation of information in the net: if hyponyms inherit their hyperonyms properties, a given argument of a verb can be introduced by the indicated prepositional node or by any of its hyponyms:

(14) He put the books in / under/behind/inside the closet.

ii) the accurate expression of arguments considering the compositionality of PPs: preposition meaning+ complement meaning. For instance, the integration of prepositions makes it possible to encode, through the extension of involve relations, that *put* selects for an argument introduced by a preposition denoting an indicator of location (cf. 15), which is not expressible by the involved location relation as defined in Vossen (2002: 31), which requires a nominal synset (cf. 16).

(15) a. {put}$_{V}$ involve_location {in, at}$_{Prep}$
b. He put the books in the closet.

(16) a. ?{put}$_V$ involved_location {location}$_N$
b. #He put the books in the location.

The specific realization of the argument is naturally conditioned by the semantic properties of the elements in the predicate, corresponding in this case to the physical objects denoted by the direct object of the verb (*the book,* in (15b)) and by the complement of the preposition, in this case *the closet*.
This explains why sentences such as *He put the book inside the table* may be odd, or at least require the assumption that the table in question has an interior compartment, whereas sentences such as *John put the book inside the closet* may seem slightly redundant (as opposed to *John put the book in the closet*), since the container aspect of *closet* constitutes one of its defining semantic properties.
Finally, the integration of semantically full prepositions allows for encoding more accurately the semantic restrictions on argument selection, and thus semantic features, of verbs (Amaro 2009, Amaro et al. 2013). As stated above, *put* can be described as selecting for an argument of the type location (see 16). However, as illustrated below, this semantic type is most of the times built from the semantics of the preposition used: *table, closet, fridge* are hardly thought of as locations, or represented as hyponyms of *location*, but result in well-formed sentences when arguments of the verb *put* introduced by a preposition indicator of location (cf. (17)).

(17) John put the bottle in the table/closet/fridge/window.

For these reasons, the integration of prepositions in wordnets constitutes a relevant informational gain for the model and for the lexicons described.

4. Lexicographic issues: glosses, crosslinguistic equivalences and visual description

Focusing on synonymy and hyponymy relations, we modeled some subsets of Portuguese prepositions directly related to the expression of movement. These subnets concern the expression of location (Figure 1), goal location (Figure 2), source location (Figure 3) and path (Figure 4).
These subnets reveal some underlying issues concerning the description of prepositional meaning, illustrating several strategies to account for them.

4.1 Glosses: using natural language to describe prepositional meaning

The first issue requiring further reflection concerns the use of natural language to describe the meaning of prepositions, starting with the description of the initial node for each subnet.
Considering the network depicted in Figure 1, the Portuguese preposition *em* is the top node for this subnet, roughly corresponding to the English prepositions *in/at*. This preposition denotes the more general and underspecified concept of indicator of location, with regard to a reference location, which is then specified by its hyponyms. But glossing the concept denoted by this preposition as "indicator" is not a coincidence.

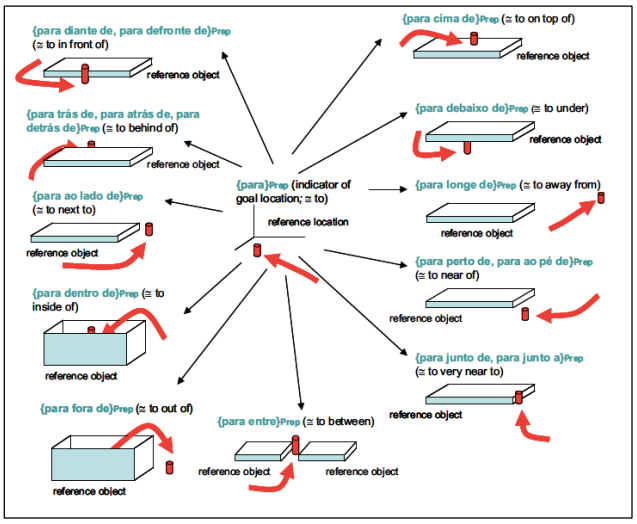

Figure 2: Hyponymy network of prepositional synsets denoting indicators of goal location

From traditional grammars (Cunha & Cintra, 1984) to current linguistic models (Saint-Dizier, 2008; Srikumar & Roth, 2013, Schneider et al., 2015), prepositions are described as items that connect other elements in a sentence. Jensen & Nilsson (2003), for instance, propose a finite set of universal binary role relations to describe the semantic content of prepositions. In their perspective, prepositions denote a relation between the concept denoted by a given lexical item and semantic roles considered in a given ontology. In other words, prepositions can be described as *indicators* of concepts relating to space, temporality, causality, and so on. These ontological analyses can provide us with the top concepts susceptible to be lexicalized by prepositions, but also with an initial proto-hyperonym from which to draw our initial glosses, i.e. the notion or concept of "indicator". Accordingly, we can gloss prepositions as indicators of location, of time, of cause, etc.

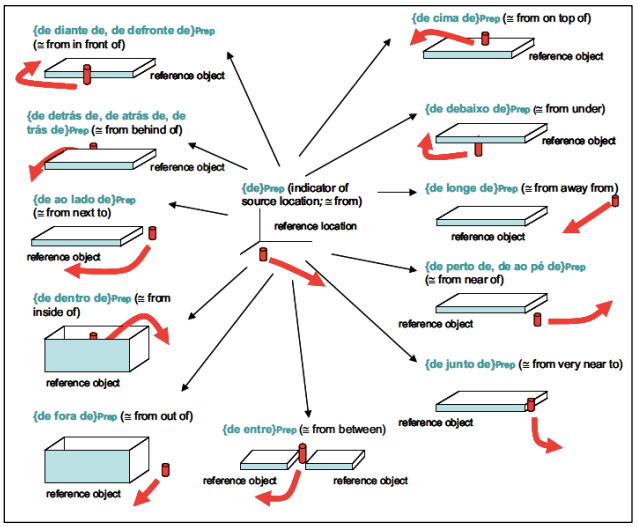

Figure 3: Hyponymy network of prepositional synsets denoting indicators of source location

Building glosses for hyponym prepositional nodes is yet another issue. It is not easy to gloss prepositional concepts

without resorting to the lexical items we intend to describe. For instance, *entre* (≈ between) can be glossed, more or less artificially, as "*em* (in/at) the space that separates objects". However, prepositional expressions such as *debaixo de* (≈ under), *em cima de* (≈ on top of), *ao lado de* (≈ next to), *atrás de* (≈ behind), etc., are not as easily glossed.

Although not as straightforwardly as for other POS, we can gloss the meaning of prepositions using two main strategies. Consider, for instance, the synset {*para fora de*}$_{\text{Prep}}$ (≈ to outside of), hyponym of {*para*}$_{\text{Prep}}$ (≈ to; indicator of goal location). We can build its gloss using:

i) the hyperonym lexical item + NP/VP/Adj concerning the hyponym specific properties (Aristotelian formula). Example: {*para fora de*}$_{\text{Prep}}$ (≈ to outside of) gloss: *para* + uma localização exterior a (*to* + a location exterior to);

ii) the proto-concept of "indicator", providing the specific notion or relation at stake. Example: {*para fora de*}$_{\text{Prep}}$ (≈ to outside of) gloss: indicador de localização final exterior ao objeto ou localização de referência (indicator of final location exterior to the reference object or location).

Both strategies have pros and cons:

i) the first strategy results in regular and direct glosses, although somewhat artificial, that allow the direct replacement of the glossed lexical units. Example: *Ele foi para fora da sala.* --> *Ele foi para uma localização exterior à sala*) (≈ He went to outside of the room --> He went to a location outside of the room);

ii) the second strategy results in more informational descriptions that help to understand more complex concepts, for instance in more abstract cases such as in *He cried in anger*; *The offer was received with fear*.

The decision for one or the other of the strategies must respect the goals and purpose of the resource and its target audience.

Nonetheless, the construction of glosses is directly related to the second issue to be accounted in wordnet model when considering prepositions, namely how to establish crosslinguistic equivalences for prepositional nodes and if these are accurate and feasible using glosses alone.

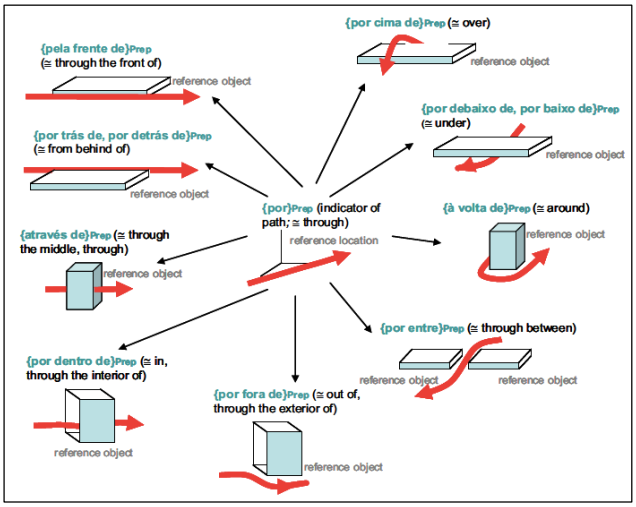

Figure 4: Hyponymy network of prepositional synsets denoting indicators of path

4.2 Crosslinguistic equivalence for prepositions and visual information

As mentioned in previous sections, several authors have studied prepositions and prepositional meaning, departing from different languages, and the concepts denoted can be fairly commonly grouped under notions of temporality, space, cause, etc., organized in several ways. For explanation purposes, Figures 5 and 6 present different proposals concerning different approaches and languages.

AGENT _____ animate being acting intentionally (ex: *treatment **by** physician*)
CAUSE _____ inanimate force/actor
CAUSED_BY ____ inversed CAUSE
PATIENT _____ affected entity/effected entity (ex: *treatment **of** children.*)
PART_OF _____ part of whole/member of (ex: *side **of** the head, cells **in** the eye, agent **from** the CIA*)
COMPRISE ____ inverse PART_OF; whole constitued of parts
BY MEANS OF __ means to end/instrument (ex: *treatment **with** medicine*)
SOURCE _____ source, origin, point of departure (ex: *haemorrhege **from** the intestine*)
PURPOSE _____ purpose
LOCATION ____ place, position (ex: *inflammation **of** the eyes*)
TEMPORALITY __ temporal anchoring, duration, inception, etc. (ex: ***for** two days, **from** last year*)
MATERIAL _____ material (ex: *cushion **of** leather.*)
CHARACTERIZE _ property ascription (ex: *children **with** diabetes*)

Figure 5: Top ontology of prepositional role relations presented in Jensen & Nilsson (2003: 8) for English

- Localization:
  - source
  - destination
  - via/passage
  - fixed position
- Quantity
  - numerical or referential quantity
  - frequency and iterativity
  - proportion or ratio
- Manner
  - manners and attitudes
  - means (instrument or abstract)
  - imitation, agreement or analogy
- Accompaniment
  - adjunction
  - simultaneity of events
  - inclusion
  - exclusion
- Choice and exchange
  - exchange
  - choice or alternative
  - substitution
- Causality
  - cause
  - goal or consequence
  - intention
  - purpose
- Opposition
- Ordering
  - priority
  - subordination
  - hierarchy
  - ranking
  - degree of importance
- Instrument
- Other groups
  - theme
  - in spite of
  - comparison

Figure 6: Abstract notions and facets denoted by prepositions (Saint-Dizier 2008: 764-765) for French

These Figures illustrate sets of notions commonly related to the meaning of prepositions, evidencing that these can be more or less regular across languages. However, even in typologically close languages such as English and French, or Portuguese (as opposed to English and Guarani, for instance), establishing equivalences in prepositional meaning can be tricky. In our perspective, this happens for two main reasons:

i) speakers tend to actualize prepositional meaning considering the distributional properties of the prepositions (i.e., the meaning of the predicates with which semantically full prepositions can occur contributes to the actual definition of their meaning), and distributional properties are inherently language dependent;

ii) prepositions constitute a close class with highly polysemous items, even within the same semantic domain (observe, for instance, the English prepositions *over*, *under*, *through*, *in*, *to*, within the semantic field of movement (cf. Figures 1 to 4)).

The description of sense 1b (of the 17 listed) of the preposition *under* in the *American Heritage Dictionary of the English Language* clearly illustrates this:

(18) 1b. To or into a lower position or place than: *rolled the ball under the couch*.

The equivalence of this sense of *under* in Portuguese has to deal with differences in i) polysemy: *to* = *para* (goal location?) and *into* = *em* (goal position?: not possible in Portuguese); ii) distribution: *rolled the ball*, manner of motion verb (*roll*) non existent in Portuguese.

Also, if we add to this the issues concerning conceptual differences and the description of prepositional meaning through glosses, the potential for inaccuracy and confusion grows further. For instance, *perto de* corresponds to *near* or to *close*? Are these synonyms? And *junto de*? Does it denote a closer location (see Figure 1)? So, how to accurately describe prepositional meaning?

Considering the subset of prepositions studied (Portuguese prepositions concerning the expression of movement), the visual description, as illustrated in the Figures 1 to 4 above, seems to be an efficient strategy. In fact, and given the issues described above, several authors have used spatial models to describe the meaning of prepositions (Galton, 1993, 1997; Herzog, 1995; Asher & Sablayrolles, 1996; Lockwood et al., 2005, among others), thus further motivating our approach.

The visual description proposed uses static elements in the case of location (e.g. in Figure 1) and dynamic ones (arrows) in the cases where there is a component of movement associated to the meaning of the preposition (e.g. in Figure 4), as well as color to highlight the core elements of the descriptions:

- reference objects and locations are depicted in gray and soft colors, with deeper tones whenever 3-dimensional perspective is relevant;
- core objects and representations are depicted in red and bright color, and lines with initial or final arrows are used to represent movement and direction, whenever relevant.

Visual descriptions should be as flat and repetitive as possible, to avoid introducing additional elements and contributing to possible different interpretations.

The use of visual descriptions allows, thus, for straightforwardly representing the meaning of these prepositions[6], while it also illustrates the polysemy of prepositional items (in whatever languages are encoded or 'translated' in the net) and the existence of conceptual voids or gaps, given the fact that visual information is language independent. Naturally, this implies a more complex database able to cope with and display visual information, as well as user-friendly graphic editors for lexicographers.

Also, as less-intensively connected items in a model in which the relations established with the other nodes primarily represent the meaning of a unit, prepositions (as well as of other POS in similar conditions) can profit from the use of visual information for a more rich semantic description. Glosses can, thus, be used for adding useful information of a different nature, such as distributional information, for instance.

[6] The conception of visual descriptions for prepositions related to other notions (causality, manner) may pose specific challenges in itself, which although very interesting are out of the scope of this paper.

5. Final remarks

The integration of prepositions in wordnets, in itself, is currently a non-controversial issue that responds to an identified and open challenge for this model, in particular when it comes to semantically full prepositions. However, the encoding of prepositions reveals further lexicographic challenges concerning the description of their meaning.

In this paper, we aimed at showing that the integration is possible and quite easy, requiring mainly the linguistic adaptation of the tests and conditions that mediate the establishment of the relations of synonymy, antonymy, hyperonymy and cause between prepositional nodes. We demonstrate that the integration of prepositions results in a more complete description of other lexical items, such verbs and verbal selection properties, but it also allows for accounting for computational processes of meaning compositionality.

In what concerns lexicographic strategies, the integration and description of prepositions show the difficulties of establishing equivalences between the concepts denoted by prepositions in different languages, as well as using glosses in natural language to describe their meaning. The use of visual information obviates these issues, while posing issues on implementation.

Finally, the integration of prepositions makes wordnets more useful and usable resources, by augmenting the words described and the quantity of information encoded, and contributes to test other lexicographic strategies, as for instance freeing glosses to serve other lexicographic purposes, instead of being used to describe the meaning of lexical units when the semantic relations available are not sufficient.

6. Acknowledgments

I thank the anonymous reviewers for the careful reading and for all the suggestions and critiques that definitely contributed to the improvement of this paper.

7. Bibliographical References

Amaro, R. (2009). Computation of Verbal Predicates in Portuguese: Relational Network, Lexical-Conceptual Structure and Context - the case of verbs of movement. PhD thesis. University of Lisbon.

Amaro, R., Mendes, S. & P. Marrafa (2010). Encoding Event and Argument Structures in Wordnets. In Sojka P., Horák, A., Kopeˇcek, I. & K. Pala (eds.), Text, Speech and Dialogue, LNAI 6231, Berlim Heidelberg: Springer-Verlag, pp. 21–28.

Amaro, R., Mendes, S., and Marrafa, P. (2013). Increasing Density through New Relations and PoS Encoding in WordNet.PT. *International Journal of Computational Linguistics and Applications*, 4(1): 11–27.

American Heritage Dictionary of the English Language (2016). Fifth Edition, Houghton Mifflin Harcourt Publishing Company.

Asher, N. & P. Sablayrolles (1996). A Typology and Discourse Semantics for Motion Verbs and Spatial PPs in French. In Pustejovsky, J. & B. Boguraev (eds.), Lexical Semantics: the Problem of Polysemy, Oxford: Clarendon Press, pp. 163-209.

Baldwin, T., Kordoni, V. & Villavicencio, A. (2009). Prepositions in applications: a survey and introduction to the special issue. Computational Linguistics, 35(2):119–149.

Bannard, C. & T. Baldwin (2003). Distributional Models of Preposition Semantics. In Proc. of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications, Toulouse, pp. 169–180.

Bond, F. & M. Piasecki (2017). Introduction: Contemporary Challenges for Development and Application of Wordnets. In Proc. of the Workshop on Challenges for Wordnets. Galway. http://ceur-ws.org/Vol-1899/wordnet_preface.pdf.

Cannesson, E. & P. Saint-Dizier (2002). Defining and representing preposition senses: A preliminary analysis. In Proc. of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions, Philadelphia, USA, pp. 25–31.

Crystal, D. (ed.) (1995). The Cambridge Encyclopedia of the English Language. Cambridge University Press, Cambridge.

Cunha C. & L. Cintra (1984). Nova Gramática do Português Contemporâneo. Edições João Sá da Costa, Lda, Lisboa.

Dimitrova, T., E. Tarpomanova & B. Rizov (2014). Coping with Derivation in the Bulgarian Wordnet. In Proc. of the 7th Global Wordnet Conference, pp. 109-117. http://www.aclweb.org/anthology/W14-0115

Dorr, B. J. (1997). Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. In Machine Translation, 12 (4): 271–322.

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

Galton, A. (1997). Space, time, and movement. In Stock, O. (ed.), Spatial and temporal reasoning, Kluwer Academic Publishers, pp. 321-352.

Godoy, L. (2008). Preposições e os verbos transitivos indiretos: interface sintaxe-semântica lexical. In Revista da ABRALIN, vol. VII, pp. 49-68

Gouws, R. H. (2004). State of the art: Lexicology and Lexicography: Milestones in metalexicography. In P. G. J. van Sterkenburg, Piet (ed.) Linguistics Today: Facing a Greater Challenge. John Benjamins Publishing Company. Volume 1, pp. 187-206.

Hernández-Pastor, D., Periñán-Pascual, C. (2016). Developing a knowledge base for preposition sense disambiguation: A view from Role and Reference Grammar and FunGramKB. Onomázein, no. 33, Editorial Pontificia Universidad Católica de Chile.

Herzog, G. (1995). Coping with Static and Dynamic Spatial Relations. In Amsili, P., M. Borillo & L. Vieu (eds.), Proc. of TSM'95, Time, Space, and Movement: Meaning and Knowledge in the Sensible World, Château de Bonas, pp. 47–59.

Jensen, P. & J. F. Nilsson (2003). Ontology-Based Semantics for Prepositions. In Proc. of ACL-SIGSEM workshop: The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications, Toulouse.

Klein, W. (2001). Lexicology and lexicography. In N. Smelser, & P. Baltes (eds.), International encyclopedia of the social & behavioral sciences: vol. 13, pp. 8764-8768.

Lockwood, K., K. Forbus & J. Usher, J. (2005). SpaceCase: A model of spatial preposition use. In Proc. of the 27th Annual Conference of the Cognitive Science Society, Stressa.

Mcshane, M, S. Beale & S. Nirenburg (2005). Disambiguating Homographous Prepositions and Verbal Particles in an Implemented Ontological Semantic Analyzer. Working Paper 01-05, Institute for Language and Information Technologies University of Maryland.

Neeleman, A. (1997). PP-Complements. In Natural language and linguistic theory, n. 15, pp. 89-137.

O'Hara, T. & J. Wiebe (2003). Preposition semantic classification via Treebank and FrameNet. In Proc. of the 7th Conference on Natural Language Learning (CoNLL-2003), Edmonton, Canada, pp. 79–86.

Pribbenow, S. (1993). Computing the meaning of localization expressions involving prepositions: The role of concepts and spatial context. In Zelinsky-Wibbelt, Cornelia (ed.), The Semantics of Prepositions. From Mental Processing to Natural Language Processing. De Gruyter, pp. 441–470.

Sag, I. & Wasow, T. (1999). Syntactic Theory: A Formal Introduction, CSLI Publications, Stanford, CA.

Saint-Dizier, P. (2005). PrepNet: a framework for describing prepositions: preliminary investigation results. In Bunt, H., J. Geerzen & E. Thijse (eds.), Proc. of the 6th International Workshop on Computational Semantics (IWCS'05). ITK, Tilburg, pp. 25-34.

Saint-Dizier, P. (2008). Syntactic and Semantic Frames in PrepNet. In International Joint Conference on Natural Language Processing (IJCNLP 2008), ACL, Hyderabad (Inde), p. 763-768.

Schneider, N., Srikumar, V., Hwang. J. D, & Palmer, M. (2015). A Hierarchy with, of, and for Preposition Supersenses. In Proc. of LAW IX - The 9th Linguistic Annotation Workshop, Denver, Colorado, ACL, pp. 112–123.

Srikumar, V. & Roth, D. (2013). Modeling Semantic Relations Expressed by Prepositions. In Transactions of the Association for Computational Linguistics, 1 pp. 231–242.

Verspoor, C. M. (1997). Contextually-Dependent Lexical Semantics. PhD dissertation, University of Edinburgh.

Vossen, P. (ed.) (2002). EuroWordNet General Document", version 3. University of Amsterdam, http://www.vossen.info/docs/2002/EWNGeneral.pdf