

PrepNet: a Framework for Describing Prepositions: Preliminary Investigation Results

Patrick SAINT-DIZIER

IRIT-CNRS, 118, route de Narbonne, 31062 Toulouse France

`stdizier@irit.fr`

1 Introduction

Describing the syntax and the semantics of prepositions, in a way similar to verbs (e.g. in FrameNet (www.icsi.berkeley.edu/framenet/), or VerbNet (www.cis.upenn.edu/verbnet/)) or to nouns (as in WordNet and EuroWordNet) is obviously a very challenging, but necessary task. Prepositions turn out to be a very useful category in a number of applications such as indexing and knowledge extraction since they convey basic meanings of much interest like instruments, means, comparisons, amounts, approximations, localizations, etc. They must necessarily be taken into account—and rendered accurately—for effective machine translation and lexical choice in language generation.

While quite a lot of work has been devoted to nouns and verbs, little has been done in Computational Linguistics circles about prepositions. The reasons are quite clear: prepositions are probably the most polysemic category, possibly more so than adjectives, and linguistic realizations are extremely difficult to predict, not to mention the difficulty of identifying cross-linguistic regularities.

From a linguistic perspective, several investigations have been carried out on quite diverse languages, emphasizing e.g., monolingual and cross-linguistic contrasts or the role of prepositions in syntactic alternations. These observations cover in general a small group of closely related prepositions. The semantic characterization of prepositions has also motivated the emergence of a few dedicated logical frameworks and reasoning procedures. Let us mention projects devoted to prepositions expressing space, time and movement in AI and in NLP, and also the development of formalisms and heuristics to handle PP attachment ambiguities. Let us also mention the large number of studies in psycholinguistics and in ethnolinguistics around specific preposition senses. These results remain quite often relatively theoretical, focussing on cognitive representations. Most of them need a lot of elaborations to be used in NLP systems. From a representation point of view, prepositions seem to reach a very deep level in the cognitive-semantic

structure [3, 4, 9, 12, 13, 17, 20]: cognitive grammar researchers often use prepositions in their metalanguage, in order to express very primitive notions. It is of interest to base our investigations and representations on these issues.

In argument structure, prepositions often play the crucial role of a mediator between the verb's expectations and the semantics of the nominal argument. The verb-preposition-noun semantic interactions are very subtle, but totally crucial for the development of an accurate semantics of the proposition including arguments and modifiers. Prepositions are also closely related to semantic structures such as thematic roles, semantic templates or frames. Finally, languages like English have verbal compounds (verb-particle constructions) [19] that integrate prepositions (compositionally or as collocations) while others, like Romance languages, rather have the preposition as PP head in prepositional phrases or possibly incorporate the preposition in the verb.

In the present work, we only describe prepositions that behave compositionally, as PP-heads, and which have a semantic contents (i.e. we do not consider cases where prepositions just have a grammatical role). This is restrictive, but this is a first step, which, nevertheless, covers a significant set of usages. The semantics of constructions such as verb particule constructions [19] are in general described in lexical entries, because of their relatively ad'hoc character. The same remark holds for idiomatic constructions.

We also make the assumption that preposition behaviors can be described, in a first stage, almost separately from verbs and nouns, although we will have to specify a number of constraints on usages. Prepositions are often said to be almost infinitely polysemous, and, therefore, capturing regularities and introducing categorizations is extremely difficult. We agree with this point of view, however, within a cognitive view of language, with rather broad senses with several levels of stratification, and for most 'standard' usages, we believe it is really possible to construct, via corpus, a relatively fine-grained description of prepositions of interest for a large variety of applications. This remark also holds for the characterization of the numerous metaphors and metonymies prepositions may undergo. However, in particular for metaphors, it is necessary at an early stage to introduce parameters from the verb.

In this paper, we present the results of a preliminary investigation, which is really exploratory, that aims at constructing a repository of preposition syntactic and semantic behaviors. A general, conceptual categorization of prepositions is first presented (formed of abstract notions on which preposition senses are based), followed by a multilevel characterization of preposition usages. For that purpose, we introduce a preliminary frame-based format for representing preposition most prototypical behaviors. Next, we briefly sketch how these prototypical behaviors can be enriched based on a corpus-driven methodology (the direct and indirect realization levels). Fi-

nally, we briefly show how multilinguism can be treated, since this is a crucial facet of prepositions (the multilingual level). In a next stage, we would also like PrepNet to be an open system in which external contributors can enter preposition descriptions of their own language.

2 A general categorization of preposition senses viewed as abstract notions

Here is a categorization proposal of the different senses that prepositions may have. So far, it is essentially based on the observation of French prepositions, with some confirmations from English, Spanish and German. It is largely inspired from thematic role classifications. Senses are here viewed as **abstract notions**; they are associated with lexicalizations under specific restrictions in different languages (the realization levels). The abstract notions we have defined are organized on three levels, from [2]:

1. a first level characterizes a **semantic family**, of a level roughly comparable to thematic roles: localization, manner, quantity, accompaniment, etc.,
2. a second level accounts for the different **facets** of the semantic family, e.g. source, destination, via, fixed position for the localization family,
3. a third level characterizes, roughly speaking, the **modalities of a facet** when appropriate. For example, the facet *manner and attitudes* is decomposed into 3 modalities: *basic manner*, *manner by comparison* and *manner with a reference point*. Due to space limitations, this latter level will not be much developed in this document.

Categorization is as follows:

- **Localization** with subsenses:
 - **source**, - **destination**, - **via/passage**, - **fixed position**.Destination may be decomposed into destination reached or not (possibly vague), but this is often contextual. From an ontological point of view, all of these senses can, a priori, apply to spatial, temporal or to more abstract arguments.
- **Quantity** with subsenses:
 - **numerical or referential quantity**, - **frequency and iterativity**,
 - **proportion or ratio**.Quantity can be either precise (*temperature is 5 degrees above 0*) or vague. Frequency and iterativity, e.g.: *he comes several times per week*.

- **Manner** with subsenses:
 - **manners and attitudes, - means (instrument or abstract), - imitation or analogy.**
 - Imitation: *he walks like a robot; he behaves according to the law,*
- **Accompaniment** with subsenses:
 - **adjunction, - simultaneity of events (co-events), - inclusion, - exclusion.**
 - Adjunction : *flat with terrace / steak with French fries / tea with milk,*
 - Exclusion: *they all came except Paul.*
- **Choice and exchange** with subsenses:
 - **exchange, - choice or alternative, - substitution.**
 - Substitution : *sign for your child,* Choice: *among all my friends, he is the funniest one.*
- **Causality** with subsenses :
 - **cause, - goal or consequence, - intention.**
 - Cause: *the rock fell under the action of frost.*
- **Opposition** with two ontological distinctions: physical opposition and psychological or epistemic opposition. Opposition: *to act contrary to one's interests.*
- **Ordering** with subsenses:
 - **priority, - subordination, - hierarchy, - ranking, - degree of importance.**
 - Ranking : *at school, she is ahead of me.*
- **Minor groups:**
 - **About, - in spite of, - comparison.**
 - About: *a book concerning dinosaurs.*

3 Preposition usage characterization using Frames

3.1 Architecture of PrepNet

In PrepNet, preposition usage characterizations are mainly organized at the lower level of the hierarchy presented above: the modality. Sometimes facets, the level just above, are considered instead because their granularity is sufficient. Each modality is characterized by a rough semantic representation based on dedicated primitives, that we have elaborated. This representation is a simplified version of the Lexical Conceptual Structure (LCS) [9]. Each abstract notion is associated with a set of frames that describe the structure in which the preposition is embedded (much larger than its maximal projection) so that appropriate constraints on usage can be stated.

Representations can be further stratified to account for differences between prepositions. This level can be viewed as an interlingua level, essentially conceptual. At this level, we also plan to introduce inferential patterns associated with semantic representations. A macro level is also foreseen, e.g. to organize preposition senses, for example a hierarchy of prepositions denoting approximation, from the most constrained to the most flexible one.

Within the PrepNet framework, we have described so far 195 senses, using 65 primitives, based on English preposition names (on, near, with, etc.). The 65 primitives identified do reflect the variety of primitive notions conveyed by prepositions. Senses may be a composition of several primitives. There are obviously decisions to make about sense distinctions and their encoding by means of primitives, one can argue on some choices and their ontological or cognitive status. A good test of this preliminary work will be the concrete use of PrepNet in the development of applications.

Work has been so far carried out on French, it is clear that some revisions may be needed when investigating other languages. However, looking at some other languages, we feel to have reach some level of stability. Primitives are viewed here as linguistic macros, which can then be interpreted depending on the environment (e.g. Euclidean geometry for spatial prepositions). The work has been initially designed from a manual analysis of prepositions [2], and then enriched via corpus exploration. Corpora are for a large part fragments of a French encyclopedia (Encyclopedia Universalis), newspapers and scientific papers. A typology of preposition distribution is given in [2].

3.2 Representation of abstract notions

Let us now concentrate on the representations we have settled. An entry at the conceptual level in PrepNet is, at the moment, composed of:

1. **a number, a name and a gloss**, that informally describe the semantics of the abstract notion at stake:
[sense number], name from hierarchy above, 'gloss',
2. **a frame with constraints**, constraints are relatively 'shallow', these are improved by corpus exploration and categorization:
 $X <ACTION> Y$ [sense number] Z ,
where X , Y and Z are the verb (noted as ACTION or STATE) arguments, this frame is followed by the specification of shallow constraints on the verb and the arguments,
3. a **conceptual representation**, in simplified LCS form (in which we essentially keep the semantic field, for which we have developed a richer set of categories). At this level, only the semantics of the preposition is captured. This level introduces a decompositional approach to preposition meaning.

By shallow constraints, we mean (1) the use of a quite generic (or shallow) set of semantic types, (2) the use of the generic verb classes largely derived from WordNet [8] and (3) the use of a number of semantic fields for LCS representations: poss (possession), temp (time), loc (localization), psy (psychological), comm (communication), epist (epistemic), abs (abstract), prop (property), perc (perception), and amount (quantity). We view these features as prototypes, around which uses are grouped. Other uses, such as metaphors or metonymies, will be derived by means of inference rule schemas, among which, type coercion for metonymies. Lexicalization examples (preposition senses), encoded in a kind of synset for a given language, corpus data and more accurate usage constraints are described in modules dedicated to particular languages.

Note that a given verb sense may belong to several classes, such as communication and social interaction. These classes are not postulated to be disjoint. They rather capture generic semantic 'orientations' in verbs. Percentages in the above chart correspond to verb senses.

3.3 A multi-level description of language realizations

The first example below illustrates the main elements given above. The facet VIA of the 'spatial' family describes an action occurring via a passage. Classifications and distinctions are essentially made below from the identification of different selectional restrictions and language realizations (i.e. 1 and 2 below have different sets of surface realizations associated, and different restrictions on the NP argument, which motivates their distinction in the VIA family). It has at least two modalities (examples below can be slightly further decomposed), which are distinguished on the basis of the object shape through which the action <ACTION> occurs:

```
[1] : VIA - narrow passage.
'An entity X moving via / an action that uses
  a narrow passage in an object Y'
X <ACTION> [1] Y
X: concrete entity, ACTION: perception verb,
  Y: location with a narrow passage
representation: X : through(loc or temp, Y)

// illustration: French synset: {a travers, au travers de, dans}
// example: Jean regarde a travers la grille / dans les jumelles.

[2] : VIA - generic.
X <ACTION> [2] Y
'An entity X moving via a location Y'
X: concrete entity, ACTION: movement verb, Y: location
representation: X : via(loc, Y)

// French synset: {par, via}
```

```
// example: Jean rentre par la porte.
```

(litteral translations of examples: 1: John looks through the gate / in binoculars, 2: John comes in by the door).

Preposition sense [2] can be composite when the preposition *par* is combined with a fixed location preposition such as *dessous*, *dessus* etc. to form compounds such as: *par dessus*, *par dessous* (via under, via above). The semantic representation has then an embedded functional structure:

```
X : via(loc, under(loc, Y)).
```

The description of [2] can then be stratified as follows:

```
[2] : VIA - generic.
```

```
X <ACTION> [2] Y
```

```
'An entity X moving via a location Y'
```

```
X: concrete entity, ACTION: movement verb, Y: location
```

```
representation: X : via(loc, Y)
```

```
// synset: {par, via}
```

```
// example: Jean rentre par la porte.
```

```
[2.1] VIA UNDER - generic
```

```
X <ACTION> [2] Y
```

```
'An entity X moving via under a location Y'
```

```
X: concrete entity, ACTION: movement verb,
```

```
Y: location with a passage under it
```

```
representation: X : via(loc, under(loc,Y))
```

```
// French synset: {par dessous}
```

```
// example: Jean passe par dessous le pont.
```

```
[2.2] VIA ABOVE - generic
```

```
etc.
```

The second example below shows, for the abstract notion of 'front position', a distinction made on the semantic domain of the argument: localization on the one hand and psychological or epistemic on the other. This distinction is motivated by the emergence of two very distinct senses, characterized by two different synsets (linguistic realizations) and constraints in the frame:

```
[3] : FIX LOC - in front of object
```

```
'An entity X located in front of another object Y'
```

```
X <ACTION/STATE> [3] Y
```

```
X: concrete entity,
```

```
ACTION/STATE: position verb, movement verb, Y: object
```

```
representation: X: opposite(loc, Y)
```

```
// synset: {en face de, a l'oppose de}
```

```
// example: Il habite en face de la mairie.
```

```
[4] : FIX LOC - front of psy or epist object
```

```
'Someone X against a law, an idea, or a principle Y'
X <ACTION/STATE> [4] Y
X: human, ACT/STATE: psychological or epistemic verb,
    Y: abstract
representation: X: front(psy or epist, Y)

// synset: {contre}
// example: Il proteste contre cette loi.
```

(translations of examples: 3: he lives in front of the town hall, 4: he demonstrates against this law). The third example below introduces another type of stratification motivated by the expression of relations or constraints between arguments, which will determine different lexicalizations. This is, for example, the case for prepositions denoting instruments, as studied in [12]. In this work, we introduced two relations:

- the relation between the actor/agent and the instrument, with 3 levels: Undergo (no control on the instrument or its properties), Select (the actor has some control on the object, but it does not plan to do the action that happens, e.g. like in accidents), Control (the agent has full control on the instrument),
- the relation between the instrument and the Verb-object NP, with 3 levels: Be (the object has some intrinsic properties such that even being passive, it nevertheless participates to the action), React (controlled by the agent for a particular property, the object participates to the action via another property), Act (the instrument fully participates to the action).

In this situation, we postulate a single entry, with different realizations, depending on the above constraints:

```
[5] : MANNER - MEANS - Instrument
'Someone X doing an action Y using instrument Z.'
X <ACTION> Y [5] Z
X: human, ACTION: verb of change, Y: object Z: instrument
representation: X: by-means-of(_, Z)
constraint: +Be +Undergo // synset: {grace a}
constraint: +Be +Select // synset: {par}
constraint: +Select +React // synset: {avec}
constraint: +Act +Control // synset: {avec, au moyen de}
```

(literal translations of the synsets: thanks to, by, with, with + by means of).

4 Populating preposition frames via corpus

PrepNet frames are aimed at being prototypical, with usage constraints based on shallow types. These frames remain conceptual: we view them as

a kind of prelexical level. Frames have been defined from several sources: general semantics considerations in the literature (such as thematic roles or semantic categories), dictionaries (on-line and a concept-based German-French dictionary) and corpus data. Our claim, based on feedback on corpus analysis, is that frames, with the aim of being prototypical, reach a certain level of stability and granularity, from which we can study preposition semantics in more depth, via a stratified approach. Stratification allows for some flexibility when dealing with several languages. Their relevance and usability has been tested by lexicographers, as reported in [2].

The second stage of our work aims at associating these frames with a number of corpus observations, so that restrictions proper to each preposition in a synset can be further analyzed. Exceptions to the general rules stated can also be observed and reported. Standard usages are essentially treated from corpora observations while low frequency cases are analysed via dictionaries or by substituting a preposition by another one of the same class.

Via a semi-automatic bootstrapping method, we collected corpus occurrences of prepositions.

We focussed in a first stage on 14 prepositions (and their different senses): *contre, vers, pour, dans, sous, sur, depuis, apres, autour de, par, des, pres de, aux environs de, aux alentours de* (litt.: against, towards, for, in, on, from, after, around, by, as soon as, near, around) which are among the most frequent ones in French.

We then manually analyzed them (automatic methods are not reliable enough). One of the goals is to promote those uses that correspond the best to the specifications given in the frame. Considering in particular the restrictions on the object Y (the subject X is often much more autonomous), we then introduced a multi-level partitioning of realizations, according to a much more fine-grained set of restrictions essentially so far on Y. We can then capture a whole spectrum of usage norms. This lower level of description, much more accurate, is called the **realization level**. The final step is to establish links between these norms and the other usages found in corpus. Exceptions are so far just listed when found. For derived uses such as metaphors, dedicated operations such as type coercion (modelled by means of rewriting rules on restrictions) need to be developed. This latter level is called the **derived realization level**.

From the corpus, we extracted usages which are prototypical, leaving the others for a later analysis. These latter include less frequent usages (placed in lower usage levels), exceptions and indirect uses such as metaphors or metonymies. The task in PrepNet for the first set of usages is to make a much more detailed categorization of uses. Categorization is characterized by several types of constraints: semantic types, syntactic constraints such as case, type of the verb, etc. We report below our study of the approximation facet, considered for the 3 main semantic fields in which it is used:

time, quantity and location. Prepositions studied are *vers*, *autour de*, *aux environs de*, *aux alentours de*, which all have a sense that expresses the idea of approximation (around, about, near). So far, these restrictions remain somewhat informal, and the abstract notion studied here is among the simplests. This abstract notion is quite regular: it has very few derived uses and metaphors, and it has a relatively high autonomy w.r.t. the verb. Our aim is here to show the principles of the stratification. More complex cases will be shortly available on our web site (www.irit.fr/recherches/ILPL, 'projets' section). Here are the distributions observed (we do not have room to give details about restrictions used):

Detailed usages for the approximation facet		
frame	nb annotated	restrictions on Y
vers Y(+temp)	45	precise time, date, event
autour de Y(+temp)	12	any form of precise date
aux environs de Y(+temp)	15	any date or event
aux alentours de Y(+temp)	23	any form of date
vers Y(+amount)	55	any global quantity: fare, weight, number, size, except time
autour de Y(+amount)	18	any global form of quantity
aux environs de Y(+amount)	21	same as for vers Y
aux alentours de Y(+amount)	23	same as for vers Y
vers Y(+loc)	74	any precise area, direct or by reference
autour de Y(+loc)	18	any form of altitude or precise place
aux environs de Y(+loc)	32	any location defined directly
aux alentours de Y(+loc)	23	any precise location

Precise place: a place explicitly given (Paris, the campus), by reference: via another point (to the left of, north of). Precise event: event with clear boundaries. Global quantity: a rounded number (e.g. 15 kg opposed to 15.675 kg). As the reader can note, lexicalizations of the approximation notion are relatively homogeneous, with some variations, depending on the semantic field (temp, amount or loc).

5 Multilinguality: a few hints

Our hypothesis is that the prototypical frames of PrepNet are stable over a number of languages, at least those of Western Europe, probably with different levels of stratification, involving different type restrictions. These frames are indeed essentially of a conceptual nature. This obviously needs to be tested on a large scale. below is a simple example, to show what we mean, obviously nothing definitive can be concluded from it.

Multilingual variations are very diverse. Our preliminary investigations tend to show that we can specify at the realization level the behaviors of

prepositions in other languages. We do not establish any direct connection between two preposition realizations in two languages. The relation, in terms of translation, is established via the set of restrictions imposed on each lexicalization that corresponds the best to the restrictions imposed on the argument Y. The impact of the other elements (arguments and verb) remains to be explored. Each language has an independent module that establishes a link between the abstract notions and their language realizations and restrictions (as sketched above for French in section 3.3).

Let us consider again the approximation facet presented above for Spanish prepositions ¹. Spanish has 6 prepositions to express this notion, however, 4 of them are really frequent: *cosa de*, *hacia*, *cerca de*, *alrededor de*, while *por* and *mas o menos* are not studied here. Although the approximation facet is not the most complex one, this contrastive study shows the usefulness of PrepNet multilevel description for multilingual purposes. Usage constraints are the following:

Detailed usages for the approximation facet in Spanish	
frame	restrictions on Y
cosa de Y(+temp)	precise time, date, durative event
hacia Y(+temp)	any form of precise time, date, durative event
cerca de Y(+temp)	any hour in the day or any event
alrededor de Y(+temp)	any form of date, time or event
cosa de Y(+amount)	any global quantity: fare, weight, number, size, and time
hacia Y(+amount)	any global form of number, distance or size
cerca de Y(+amount)	same as for cosa de Y
alrededor de Y(+amount)	same as for cosa de Y
cosa de Y(+loc)	any global notion of altitude
hacia Y(+loc)	any global form of area, altitude or place
cerca de Y(+loc)	same as hacia de Y, place must be precise
alrededor de Y(+loc)	same as hacia de Y

From this chart, it is possible, given the type of Y to select a set of realizations for the abstract notion of approximation in Spanish. In this example, which is very simple, the same semantic restrictions (durative, precise date...) are kept. It is unlikely that this is the case for the other preposition families. Consequently, the ontological status and the degree of universality of these restrictions is an open problem that requires further investigations.

6 Related work

As advocated in the introduction, there is quite a lot of literature on prepositions in psycholinguistics circles, and some in AI and in cognitive sciences

¹Many thanks to Silvia Puig Roura for her expertise

[3, 4, 11, 12, 13, 18, 20]. These works need quite a lot of customization to be integrated in PrepNet at this stage of development. We rather keep their results for more in-depth development of PrepNet.

A quite old, but still of interest is [16] which proposes, via cases or roles, a structure for prepositions, and their relations to verbs. This work settled a first understanding of how prepositions work.

The basis and the starting point of our research was developed about 8 years ago by Bonnie Dorr, it is accessible at:

www.umiacs.umd.edu/~bonnie/AZ-preps-English.lcs.

This is a very large database of preposition semantic representations, characterized by their LCS representation and, sometimes, by a thematic grid. It was conceived for machine translation tasks, which explains some of its features. There are about 500 entries (compared to our 170 entries), for probably all English prepositions. Although it is not easy to go into such a huge work dedicated to a different language and to make comparisons, we outline below some differences we feel have some importance.

Each preposition sense in Bonnie Dorr's work receives a comprehensive semantic representation in LCS. Senses are paraphrased by an example, in a way close to synsets in WordNet. Some restrictions are added, and syntactic positions are made explicit.

In our approach, we introduced disjunctions of semantic fields in order to account e.g. for metaphors. This limits the number of entries and prototypical description associated with the multilevel architecture gives a stronger structure to the description. This makes the specification of restrictions more principled-based and accurate. For example, for *behind*, B. Dorr has 3 independent senses (locative, temporal and with movement) whereas we have just one.

We also tried to be compositional, limiting the number of primitives. In Bonnie Dorr's work, there is e.g. a primitive called AWAY-FROM, in addition to AWAY and FROM. We tend to consider that these two primitives can be combined compositionally and that the composite AWAY-FROM is not motivated.

Another difference is that we have considered a kind of 'minimal' semantics for prepositions, without considering potential combinations with verbs a priori. For example, in B. Dorr there is for *against* a sense describing a fixed position and another one describing a movement where the moved object reaches a position against another object. For this latter case, we think that the movement is only in the semantics of the verb and is compositionally induced at the level of the proposition. Same remark for most prepositions expressing positions (north, west, inside, etc.). We have only one representation for the fixed position.

The other point of comparison is FrameNet, which is obviously a much more ambitious challenge than ours. FrameNet says little about prepositions, but it has a few frames such as Accompaniment which are of interest.

Prepositions being much more limited in size, and possibly in complexity than verbs (e.g. no alternations, argument restrictions more stable), it is possible to propose more formal descriptions. The main difficulty with prepositions is to characterize the different senses they may take and the 'variations' around these senses, in terms of metaphors or metonymies w.r.t. the verb they are associated with, and to provide an interpretation.

7 Perspectives

This preliminary study has, obviously, a number of perspectives. Our first aim will be to develop in depth preposition descriptions for French, followed by multilingual work, on Spanish and Catalan, and then on English and German. An idea to test is to have a kind of open system, where, following some guidelines, linguists can enter descriptions and uses of the prepositions of the language they study. Another aspect is to make data accessible in a variety of ways via the Internet.

At a more theoretical level, we plan to study in more depth the relations with the verb, for which we have developed quite a lot of descriptions in the past for French. Another point is the development of various forms of inferential patterns, which characterize the different forms of reasoning prepositions introduce (a classical example are location prepositions), or various relations that hold between prepositions (e.g. inclusion). Finally, also of interest are the integration, via lexical descriptions, verb-particle constructions [18] and collocations.

Acknowledgements We thanks the 4 anonymous reviewers for their useful comments. This project was partly funded by the CNRS TCAN.

References

- [1] Baker, M.C., (1988), *Incorporation: A Theory of Grammatical Function Changing*, Chicago University Press.
- [2] Cannesson, E., Saint-Dizier, P. (2001), *A general framework for the representation of prepositions in French*, ACL01 WSD workshop, Philadelphia.
- [3] Carmen Horno Chéliz, M. del, (2002), *Lo que la preposición esconde*, University of Zaragoza press.
- [4] Cervioni, J., (1991), *La préposition: Etude sémantique et pragmatique*, Duculot, Paris.
- [5] Dorr, B., Olsen, M.B., (1997), *Deriving Verbal and Compositional Verbal Aspect for NLP Applications*, proc. ACL'97, Madrid.

- [6] Dorr, B., (1993), *Machine Translation, a view from the lexicon*, MIT Press.
- [7] Dorr, B. J., Garman, J., and Weinberg, A., (1995), *From Syntactic Encodings to Thematic Roles: Building Lexical Entries for Interlingual MT*, *Machine Translation*, 9:3-4, pp.71-100.
- [8] Fellbaum, C., (1993), *English Verbs as Semantic Net*, *Journal of Lexicography*, vol. 6, Oxford University Press.
- [9] Jackendoff, R., (1990), *Semantic Structures*, MIT Press.
- [10] Levin, B., (1993), *Verb Semantic Classes: a Preliminary Investigation*, Chicago University Press.
- [11] Lindstromberg, S. (1997), *English Prepositions Explained*, John Benjamins.
- [12] Mari, A. (2000), *Polysémie et Décidabilité. Le cas de avec ou l'association par les canaux*, Thèse de Doctorat, EHESS, Paris, collection Langue et Parole, L'Harmattan, 2003.
- [13] Pesetsky, D., (1982), *Paths and Categories*, MIT doctoral dissertation.
- [14] Pustejovsky, J., (1991), *The Generative Lexicon*, *Computational Linguistics*, vol. 17, MIT Press.
- [15] Pustejovsky, J., (1995), *The Generative Lexicon*, MIT Press.
- [16] Spark-Jones, K., Boguraev, B., A note on a study of cases, research note, dec. 85.
- [17] Talmy, L. (1976), *Semantic Causative Types*, In M. Shibatani (ed.), *Syntax and Semantics 6: The Grammar of Causative Constructions*. New York: Academic Press, pp. 43-116.
- [18] Talmy, L., (1985), *Lexicalization Patterns: Semantic Structure in Lexical Forms*, in *Language Typology and Syntactic Description 3: Grammatical Categories and the Lexicon*, T. Shopen (ed.), 57-149, Cambridge University Press.
- [19] Villavicencio, A., (2005) *Verb-particle Constructions in the WWW*, in P. Saint-Dizier (ed), *Syntax and Semantics of Prepositions*, Kluwer Academic, to appear.
- [20] Wierzbicka, A. (1992), *Semantic Primitives and Semantic Fields*, in A. Lehrer and E.F. Kittay (eds.), *Frames, Fields and Contrasts*. Hillsdale: Lawrence Erlbaum Associates, pp. 208-227.