

# PrepNet: a Multilingual Lexical Description of Prepositions

Patrick Saint-Dizier

IRIT-CNRS  
118, route de narbonne  
31062 Toulouse cedex France  
stdizier@irit.fr

## Abstract

In this paper, we present the results of a preliminary investigation that aims at constructing a repository of preposition syntactic and semantic behaviors. A preliminary frame-based format for representing their prototypical behavior is then proposed together with related inferential patterns that describe functional or paradigmatic relations between preposition senses.

## 1. Aims

Describing the syntax and the semantics of prepositions, in a way similar to verbs (e.g. in FrameNet ([www.icsi.berkeley.edu/frame-net/](http://www.icsi.berkeley.edu/frame-net/)), or VerbNet ([www.cis.upenn.edu/verbnet/](http://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. However, so far, prepositions have not been much studied in NLP circles. PrepNet ([www.irit.fr/recherches/ILPL/prepnet.html](http://www.irit.fr/recherches/ILPL/prepnet.html)) is a framework that aims at constructing a repository of preposition syntactic and semantic behaviors. PrepNet is structured in three levels:

- the abstract notion level: global characterization of preposition senses in an abstract way,
- the frame level where frames and shallow restrictions are introduced to represent generic syntactic and semantic aspects of these notions, and
- the language realization levels that deal with realizations for various languages.

Levels (1) and (2) above can be viewed as a kind of structured, multi-layered interlingua form. Due to the complexity of preposition behaviors, this level is more complex than, e.g. ILI in EuroWordNet.

## 2. Main features of PrepNet

Within the PrepNet framework, we have described so far 195 preposition 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. Abstract notion representations may be a composition of several primitives. Primitives are viewed here as linguistic macros, which can then be interpreted depending on the environment (e.g. Euclidean geometry for spatial prepositions). 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 and Spanish, it is clear that some revisions and refinements are needed when investigating other languages. For that purpose, we introduced some flexible means: abstract notions strata and a multi-level description of language realizations. However, we feel to have reached some level of stability for the abstract notion, and the for general architecture of the system.

We have studied in depth preposition realizations around the abstract notions of theme, approximation and instruments for Thai, Bahasa Malaysia, Hindi, Urdu, Kashmiri and Bengali, thanks to a STIC-Asia funding starting (Kawtrakul et al 06). We also studied instrumentality for German, Italian, Arabic and Berber.

An important issue is obviously the characterization of preposition uses over several languages and the methods to deploy to be able to carry out descriptions adequately. This is a major challenge because of the large variety of behaviors one can observe which are not necessarily only superficial, but which may involve different conceptual views. Another problem is that a number of languages use other syntactic categories, incorporation or morphology (such as case marks) instead of prepositions in some situations. A multilingual analysis needs to be somewhat transcategorial and both syntactic and semantic.

## 3. Preposition usage characterization using Frames

In PrepNet, preposition senses are characterized by means of abstract notions which capture senses in a conceptual way. Abstract notions are then characterized by means of a set of syntactic frames, a semantic representation and shallow semantic restrictions. The semantic representation is a simplified version of the Lexical Conceptual Structure (LCS) (Jackendoff 90). Frames describe the structure in which the preposition is embedded (much larger than its maximal projection) so that appropriate constraints on usage can be stated. The abstract notion level can be viewed as an interlingua level, essentially conceptual. Representations are then further stratified to account for differences between prepositions. The lower level, (the realization level)

deals with preposition realizations in different languages. To account for the large variety of realizations, we developed a multi-level partitioning approach, outlining usage norms, groups of exceptions, etc.

### 3.1. General overview of abstract notions

The first level, the abstract notions, is organized as follows:

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*.

### 3.2. Representation of abstract notions

Let us now concentrate on the second level, the representations. An entry at the conceptual level in PrepNet corresponds to an abstract notion; it is 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 further refined by corpus exploration and categorization at the realization level: X <ACTION> Y [sense number] Z, where X, Y and Z are the verb (the verb being noted as ACTION or STATE) arguments, this frame is followed by the specification of shallow constraints on the verb and on 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. This representation can be viewed as a kind of conceptual prototype.

By shallow constraints, we mean (1) the use of a quite generic (or shallow) set of semantic types, (2) the use of generic verb classes largely derived from WordNet (Fellbaum 93) 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. We include in the examples below the synset and an example for French (which must be specified apart) with each frame given, so that they are easier to understand.

### 3.3. Examples: frames and strata

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 shallow selectional restrictions and language realizations. The generic case is numbered 1, 1.1 is a strata, below 1, that deals with a specific situation where the via is a narrow passage:

[1] : VIA - generic.  
 X <ACTION> [1] 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 entre par la porte.

[1.1] : VIA - narrow passage.  
 'An entity X moving via / an action  
 that uses a narrow passage in an object Y'  
 X <ACTION> [1.1] Y  
 X: concrete entity, ACTION: perception verb,  
 Y: location with a narrow passage  
 representation: X : through(loc or temp, Y)  
 French synset:  
 {a travers, au travers de, dans}  
 example: Jean regarde a travers la grille /  
 dans les jumelles.

(literal translations of examples: 1: John comes in by the door, 2: John looks through the gate / in binoculars). Sense [2] is analyzed as a specific case of [1], a particular stratification of meaning, that corresponds to different language realizations.

Preposition sense [1] has also other groups of strata associated. For example, it 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 [1] can then be associated with another group of strata as follows:

[1.2.1] VIA UNDER - generic  
 X <ACTION> [1.2.1] 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.  
 [1.2.2] 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. We have here two representations at the same level of abstraction. We then say that this abstract notion is polymorphic. Representation is as follows:

[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). In (Saint-Dizier 05), we show other types of stratifications 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 (Kawtrakul et alii, 06).

As can be noted, PrepNet frames are aimed at being prototypical, with usage constraints based on shallow types, which allows to have a priori a number of exceptions. These frames remain conceptual: we view them as a kind of prelexical level. Frames have been defined from several sources: general semantics considerations (such as thematic roles or semantic categories), dictionaries, and corpus data. Our claim, based on feedback from 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 (Cannesson et al. 2001).

#### 4. The language realization level

This level is by far the most complex. Let us here, as an illustration, report the different facets of the abstract notion of instrumentality, as reported in (Kawtrakul et alii, 06), where language details are given. In this cited paper, we studied 12 languages from 5 linguistic families: Thai, Malay, Hindi, Urdu, Kashmiri, Bengali, German, Spanish, French, Italian, Arabic and Berber. Each of the classes below originates a strata, with various set of language realizations, attached to each language.

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.

#### 4.1. Concrete instruments

All languages studied have at least one basic instrumental mark operating over concrete objects. Several refinements are identified, for specific types of NPs, or to denote a specific intention:

- the instrument is a recipient or, more generally, conveys an idea of container, the idea being is that the container is used to carry the object along a certain trajectory,
- the instrument is a part of the body. In this case, the instrument is not strictly artifactual.
- the goal is difficult to reach, it requires some efforts from the agent,
- the focus can be emphasized by using dedicated marks.

The second major difficulty is prototypicality (Rosch, 78). When the instrument used is not very prototypical of the action, several languages re-inforce the instrumental prepositions to, sort of, coerce the type of the noun so that it can become an acceptable instrument.

At a conceptual level, it is quite difficult to characterize what is a prototypical instrument for a given action (characterized by subject-verb-object: *John opens the door*). Each event has its own prototypical instrument, making corpus studies extremely large, probably unfeasible. When searching on the web, we find an incredible variety of instruments to open a door, almost impossible to classify. Next, prototypicality is not a boolean notion: instruments are more or less prototypical. Since the instrument is very much dependent on the verb and on the object, we cannot foresee any form of incorporation in the verb that would give us indications. A direction could be to assume Qualia structures (Pustejovsky 91) associated with each potential instrument that describes the function of the object in the telic role. For example, *key(X)* would have *open(X, door)*, with *door* being quite generic. This approach could work via a large lexical development for concrete nouns, it is much more risky when terms are abstract.

#### 4.2. Abstract instruments

Abstract instruments (theorems, regulations, examples, etc.) are realized identically to concrete instruments, but with some typical marks. At this stage, it is difficult to explain why marks are different from concrete instruments. An hypothesis could be that abstract instruments are closer to causes, or to more formal situations for which specific terms were developed.

#### 4.3. Metaphorical instruments

Both concrete and abstract objects can be used metaphorically as instruments. Examples abound in the literature and

on the Web. In 5.6 we examine the path metaphor which is very productive. Besides this case, we have a number of metaphors, such as: *write with your heart*, *fight with your head*, etc. These are not essentially different from metaphors observed in other situations (Lakoff and Johnson 99).

#### 4.4. The overlap instrument-manner

In a number of cases, it is not very easy to make a distinction between instrument and manner. It seems there is a continuum between these two notions or even some form of overlap, where the object is both an instrument and a manner at various degrees, which may depend on context.

#### 4.5. Causality

It is clear that, a priori, instruments can be viewed at various degrees as causes of an event. There is a kind of overlap between these two notions. Instruments are not volitional, so they are under the partial or full control of an agent (humans playing the role of instruments are also controlled by an agent).

#### 4.6. Instruments and paths

Another productive situation is the use of spatial metaphors to express instrumentality. There is a close link between instrumentality and path descriptions (spatial as well as temporal paths). This is a kind of metaphorical use of paths viewed as instruments (as can be seen in (Lakoff et al. 99): 'action is motion, goals are paths, actors are travellers'). Using an instrument parallels the use of paths in the domain of space.

Another interesting phenomenon occurs when an argument is both an instrument and a path, as in *look at the moon in a telescope*. Telescope is indeed the instrument used and also the path through which one looks, or which the light traverses.

#### 4.7. Means of transportation as instruments

Means of transportation (trains, spoons, boxes, envelopes, etc.), sometimes viewed as containers, and mediums of transportation (by air) receive a special treatment in a number of languages

#### 4.8. Language levels

Some marks are proper to formal discourse.

#### 4.9. Positive or negative orientation

The languages we studied also abound in positive-oriented marks that express in a certain way the idea of 'thanks to'. There are also several negative-oriented marks such as the following prepositional compound: F: *de la faute de*, where the term 'faute' conveys a negative orientation.

## 5. 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 [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 work 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 10 years ago by Bonnie Dorr, it is accessible at: [www.umiacs.umd.edu/~bonnie/AZ-preps-English.lcs](http://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 165 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 and to provide an interpretation. While de-

scribing frames in PrepNet, our aim was to have the perspective of using FrameNet for verbs, with descriptions of the same level of detail, and representations that could match at some level. It is clear that verbs and prepositions have intimate connections in compositional frameworks, besides the verb particle constructions in which prepositions play a prominent role.

## 6. 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 (Villavicencio, 06) and collocations.

## 7. References

- Baker, M.C., (1988), *Incorporation: A Theory of Grammatical Function Changing*, Chicago University Press.
- Cannesson, E., Saint-Dizier, P. (2001), *A general framework for the representation of prepositions in French*, ACL01 WSD workshop, Philadelphia.
- Carmen Horno Chéliz, M. del, (2002), *Lo que la preposición esconde*, University of Zaragoza press.
- Cervioni, J., (1991), *La préposition: Etude sémantique et pragmatique*, Duculot, Paris.
- Dorr, B., Olsen, M.B., (1997), *Deriving Verbal and Compositional Verbal Aspect for NLP Applications*, proc. ACL'97, Madrid.
- Dorr, B., (1993), *Machine Translation, a view from the lexicon*, MIT Press.
- 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.
- Fellbaum, C., (1993), *English Verbs as Semantic Net*, Journal of Lexicography, vol. 6, Oxford University Press.
- Jackendoff, R., (1990), *Semantic Structures*, MIT Press.
- Kawtrakul, A. et alii (2006), *A Multilingual Analysis of the Notion of Instrumentality*, proc. EACL workshop on prepositions, Trento.
- Levin, B., (1993), *Verb Semantic Classes: a Preliminary Investigation*, Chicago University Press.
- Lindstromberg, S. (1997), *English Prepositions Explained*, John Benjamins.

- 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.
- Pesetsky, D., (1982), *Paths and Categories*, MIT doctoral dissertation.
- Pustejovsky, J., (1991), *The Generative Lexicon*, Computational Linguistics, vol. 17, MIT Press.
- Pustejovsky, J., (1995), *The Generative Lexicon*, MIT Press.
- Saint-Dizier, P., (2005), *PrepNet: a Framework for Describing Prepositions: Preliminary Investigation Results*, IWCS05, Tilburg.
- Spark-Jones, K., Boguraev, B., A note on a study of cases, research note, dec. 85.
- 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.
- 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.
- Villavicencio, A., (2006) *Verb-particle Constructions in the WWW*, in P. Saint-Dizier (ed), *Syntax and Semantics of Prepositions*, Kluwer Academic.
- 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.