A Compositional Framework for Prepositions

Patrick SAINT-DIZIER
IRIT-CNRS, 118, route de Narbonne, 31062 TOULOUSE FRANCE
stdizier@irit.fr

Glòria VAZQUEZ
Universitat de Lleida, pl. Victor Siurana, 1, 25003 LLEIDA, SPAIN
gvazquez@dal.udl.es

Abstract Prepositions have seldom been studied in conjunction with the semantics of predicates (mainly verbs and predicative nouns). We propose here the main lines of a compositional framework incorporating the representation of prepositions and PPs into that of verbs. For that purpose, we use the Lexical Conceptual Structure (LCS) in conjunction with typed λ-calculus and underspecified representations. We also show how the LCS can handle metaphorical uses of prepositions.

1 Introduction

Most prepositions are highly polysemous and they can also be involved in a large number of metaphorical uses. Analyzing the semantics of prepositions is a rather delicate and risky task, but of much importance for any application that requires even an simple form of understanding. Essentially, spatial and temporal prepositions have recieved a relatively in-depth study for a number of languages (e.g. (Boguraev et al. 87), (Verkuyl et al. 92)). The semantics of the other types of prepositions describing manner, instrument, amount or accompaniement remains largely unexplored (with a few exceptions however, such as *avec* (with) (Mari 00)).

Our general application framework is machine translation and information retrieval using linguistic and symbolic techniques. In these frameworks, the treatment of predicative forms is crucial, in particular verbs, but also prepositions, which play a major role in interaction with the verb.

Our aim in this paper is not to develop detailed semantic descriptions for prepositions, but to show how prepositions combine and interfere with their closely related constituents: the verb and the NP, via restrictions percolated upwards or downwards. For that purpose, we use the Lexical Conceptual Structure (LCS) in conjunction with typed λ-calculus and underspecified representations. The LCS allows us to develop preposition semantics at an average degree of granularity, which we found sufficient for our purpose and for the applications we have in mind. The LCS has been widely used as an interlingua in machine translation (e.g. (Dorr 93), (Dorr et al. 97)). The following points are highlighted:

- the degree of polysemy of prepositions and how preposition-senses can be identified,
- a relatively simple, but efficient, way of representing prepositions using a system of primitives, showing how metaphorical transpositions can be accomodated,

- the selection of NPs by prepositions and the selection of PPs by verbs,
- the definition of a compositional framework to handle PP arguments and adjuncts,
- the treatment of complex arguments (e.g. trajectory),
- some operations on the types of the λ-calculus expressions in order to handle more complex interactions between prepositions and their environment.

These elements are developed, to some extent, in the following sections, after a short introduction to the LCS.

2 Overview of the LCS

The Lexical Conceptual Structure (LCS) is an elaborated form of semantic representation, with a strong cognitive dimension. The LCS originated in part from the Lexical Semantics Templates for its structure and from a large number of observations such as those of (Gruber 67) for its ontological typing. The present form of the LCS, under which it gained its popularity, is due to Jackendoff (Jackendoff 83, 90, 97), (Dorr 93). The LCS was designed within a linguistic and cognitive perspective. It has some similarities, but also major differences, with approaches closer to Artificial Intelligence such as semantic nets or conceptual graphs. The LCS is basically designed to represent the meaning of predicative elements and the semantics of propositions, it is therefore substantially different from frames and scripts, which describe situations in the world.

2.1 Main principles and characteristics

The LCS was initially organized around the notion of *motion*. The other semantic/cognitive fields have been derived by analogy (such as change of possession, change of property). This analogy works perfectly in a number of cases, but turns out to be unnatural in a number of others. ¿From that point of view, the LCS should be considered both as a semantic model providing a representational framework and a language of primitives on the one hand, and as a methodology on the other hand, allowing for the introduction of new primitives to the language, whenever justified (e.g. (Pinker 93)).

Another important characteristic of the LCS is its close relations with syntax, allowing for the implementation of a comprehensive system of semantic composition rules. ¿From that point of view, one often compares the LCS with a kind of X-bar semantics.

2.2 The constituents of the LCS

The different elements of the LCS language are mainly: conceptual categories, semantic fields and primitives. Other elements are conceptual variables, semantic features, indexes, and lexical functions (which play low-level roles).

A. Conceptual Categories

(Jackendoff 83) introduces the notion of conceptual category defined from a small set of ontological categories, among which the most important are: *thing, event, state, place, path, property, purpose, manner, amount, time.* These categories may subsume more specific ones, e.g. the category *thing* subsumes: *human, animal, object.*

The assignment of a conceptual category to a lexical item often depends on its context of utterance, for example the noun *meeting* is assigned the category *time* in:
after the meeting
while it is assigned the category *event* in:
the meeting will be held at noon in room B34.
There are constraints on the types of conceptual categories which can be assigned to a lexical item. For example, a color will never be assigned categories such as event or distance.

Conceptual categories are represented as an indice to a bracketed structure:
[$_{<conceptual\ category>}$]
where the content of that structure has the type denoted by the semantic category:
[$_{event}$ The meeting] starts at [$_{time}$ 2 PM].

B. Conceptual primitives

The LCS is based on a small number of conceptual primitives [1]. The main ones are BE, which represents a state, and GO, which represents any event. Other primitives include: STAY (a BE with an idea of duration), CAUSE (for expressing causality), INCH (for inchoative interpretations of events), EXT (spatial extension along something), REACT, EXCH (exchange), ORIENT (orientation of an object), etc. Their number remains small, while covering a quite large number of concepts. A second set of primitives, slightly larger (about 50) describes prepositions: AT, IN, ON, TOWARD, FROM, TO, BEHIND, UNDER, VIA, etc. These primitives are 'lower' in the primitive hierarchy and their number is *a priori* fixed once for all.

C. Semantic Fields

The LCS uses some principles highlighted in (Gruber 65), namely that the primitives used to represent concepts of localization and movement can be transposed and generalized to other fields by analogy.

The main fields considered in the LCS are the following: localization (+loc), time (+temp), possession (+poss) and expression of characteristics of an entity, its properties (+char,+ident) or its material composition (+char,+comp). (Pinker 93) introduces additional fields such as: epistemic (+epist) and psychological (+psy). We have also added +com for communication.

Primitives can then be specialized to a field, e.g. GO_{+loc} describes a change of location, GO_{+temp} a change of time, GO_{+poss} a change of possession, and $GO_{+char,+ident}$ a change in the value of a property (e.g. weight, color). Similarly, BE_{+loc} describes a fixed localization, and BE_{+poss} the possession of something by someone (omitting the arguments). All combinations of a primitive with a semantic field are not relevant.

[1] We won't discuss the use of primitive systems, since it is a very old debate, see e.g. (Wierzbicka 96), and many others such as Shank, Wilks, Jackendoff.

2.3 Construction of LCS representations

Let P be a set of primitives, S a set of semantic fields and T a set of conceptual categories as defined above. Let Var be a set of 'conceptual' variables. LCS representations are constructed formally as follows:
$\forall\ n$:

- (1) $p \in P,\ X,\ Y,\ Z \in Var,\ semantic-field \in S,\ p_{+semantic-field}$ are well-formed LCS representations,
- (2) $p_{+semantic-field}(C_1, C_2,\ ..., C_n)$ is a wff LCS if the C_i are wff LCS representations (semantic fields may be underspecified or absent),
- (3) $[_e\ C_1, C_2,\ ...,\ C_n\ ]$ is a wff LCS if e is a conceptual category ($e \in T$) and if the C_i are defined as in (2).

Condition (2) describes the instantiation of arguments by variable substitution while condition (3) describes the creation of the most external structure of an LCS and the treatment of modifiers by concatenation (or anchoring, as in TAGs in syntax).

The most important structures are the following (where semantic fields have not been included for the sake of readability since there are several possible combinations):

1. $PLACE \rightarrow [_{place}\ PLACE-FUNCTION([_{thing}\ ])]$
2. $PATH \rightarrow [_{path}\ TO/\ FROM/\ TOWARD/$
   $AWAY-FROM/\ VIA([_{thing/place}\ ])]$
3. $EVENT \rightarrow [_{event}\ GO([_{thing}\ ],[_{path}\ ])]\ /$
   $[_{event}\ STAY([_{thing}\ ],[_{place}\ ])]\ /$
   $[_{cause}\ CAUSE([_{thing/event}\ ],[_{event}\ ])]$
4. $STATE \rightarrow [_{state}\ BE([_{thing}\ ],\ [_{place}\ ])]\ /$
   $[_{state}\ ORIENT([_{thing}\ ],\ [_{path}\ ])]\ /$
   $[_{state}\ EXT([_{thing}\ ],\ [_{path}\ ])]$

PLACE-FUNCTIONS are symbols such as ON, UNDER, ABOVE, related to the expression of localization.

2.4 Using the LCS and its combination with other formal devices

It is clearly a huge enterprise to develop a formalism and a set of data sufficient to represent the meaning of a large number of words. The LCS is a vast system designed from a limited number of primitives. The complexity of the system and its expressiveness lie in the powerful, but simple, interactions between its different constituents. It is also a good representation language, with an appropriate structure.

We believe that the LCS framework is a formalism that has a good expressive power to represent the lexical semantics of words of a predicative nature. However, it is necessary to slightly extend its set of primitives, in a principled way, to improve its coverage. For example, we added the primitive FEEL to deal with a number of psychological verbs for which the GO_{+psy} was not appropriate because they do not express

a trajectory. We view LCS primitives as macros which can be further specialized or expanded in a 'pragmatic' interpretation phase. E. g. FROM or TO can be interpreted within an Euclidean geometry framework.

It is also clear (and natural) that the LCS must be paired in lexical entries with other paradigms to represent the different aspects of meaning, e.g.: attribute-value pairs for non-predicative information, lexical semantics relations (Cruse 86), including series, inferences, etc. Representations are organized around verb semantic classes, based on syntactic criteria (as in Levin) or cognitive criteria (as in EuroWordNet).

At the last IWSC3, we proposed (Saint-Dizier 99) to augment the LCS representations with:

- underspecified fragments of representations,
- paired with a typed *lambda*-calculus,
- and fragments of representations used by default.

¿From the point of view of semantic composition, additional devices are used:

- types for primitives related to prepositions,
- additional by-default representations, and rules to manage these defaults in case (very unlikely) of conflicting defaults, and
- Qualia structures of the Generative Lexicon (Pustejovsky 91).

Additional reasoning procedures can be added at this level to take into account the inclusion of domain or pragmatic data.

In conclusion, the LCS is just the kernel of our representation system, to which additional representational or computational devices are addded in order to obtain an operational formal language for natural language semantics.

3 Selectional restrictions associated with prepositions

Most prepositions are highly polysemous, similarly to the other lexical items (e.g. Copestake et al. 95, Ostler et al. 92). Our study being at a preliminary stage, linguistic data and observations are borrowed so far from our intuitions on language and from dictionary entries. In a second stage, we obviously foresee to make more extensive studies, using e.g. corpora.

We view preposition semantics at 2 levels:

- a level where relatively large sense distinctions are identified, allowing for metaphors and for various usage variations within each sense,
- a second level where a sense is decomposed into facets, not necessarily totally coherent, some a priori being more prominent than others in basic usages. Each facet is represented by a slightly different LCS.

The identification of a preposition sense is essentially based on two local criteria: (a) the type of the head noun in the PP, similarly to verbs w.r.t. their arguments, and (b) the restrictions on the nature of the PP argument(s) expected by the verb. Pragmatic or contextual factors may also interfere, but this is outside the scope of this study (e.g. Busa 96).

Let us consider the case of *par*[2]:

- causality: as in passives but also e.g. in *par mauvais temps, je ne ne sors pas* (by bad weather I don't go out),
- origin: *je le sais par des amis* (I know it from friends),
- via: *je passe par ce chemin* (I go via this path),
- tool or means: *je voyage par le train* (I travel by train),
- 'approximate' value: *nous marchons par 3500m d'altitude* (we hike at an altitude of 3500m).

Since most of these senses are subject to a large number of metaphors, the determination of non-overlapping restrictions characterizing each sense is a difficult, if not risky, enterprise.

Let us now examine the restrictions on 3 senses. The 'VIA' sense is basically related to movement verbs; it is characterized as a path, subcategorizing for a noun of type 'way' or 'route' or, by a kind of metonymic extension, any object which can define a trajectory, e.g. an aperture (by the window). It has numerous metaphors in the psychological and epistemic domains (e.g. *je traite ce phénomène par la logique temporelle* (I deal with this phenomena 'by' temporal logic)).

The 'ORIGIN' sense is more narrow, it is essentially used in conjunction with communication (+com) or epistemic verbs, the representation is usually of type place, and the head noun is of type 'human'. This is clearly an *a priori* choice, and we consider that nouns of type e.g. 'object with an informational content'. introduce a metonymic extension, consider e.g. *par la radio / la presse* (from the radio, the press). There is a kind of continuum between Origin and Causality: *I know she wears bracelets from the noise she makes when she moves.*

Finally, the 'TOOLS or MEANS' sense is related to verbs describing concrete actions (e.g. creation and movement verbs, if we refer to verb class systems (e.g. (Levin 93), (Fellbaum 93)). In general it is an adjunct. It is typed as a means, and the object head noun must be e.g. a tool, or, more generally, an object that allows the action to be realized. This object could be found e.g. in the Qualia of the PP head noun, in the encyclopedic knowledge associated with the verb, or via a functional relation in a thesaurus.

[2]This is obviously one possible characterization of the different meanings of *par* which is very much dependent on the theory of meaning one considers.

4 Representing prepositions and PPs in verbs

4.1 Prepositions

The LCS postulates a set of about 50 non-decomposable, low level primitives to represent prepositions. These primitives are directly preposition names in the LCS metalanguage, but they are not necessarily used directly for the corresponding preposition.

For example, 2 major senses of the preposition *avec* (with) are:

- **accompaniement**, represented as: $\lambda I\ [_{manner}\ WITH_{+loc}([_{thing}\ I\ ])]$, +loc indicates a physical accompaniement (*I go to the movies with Maria*), while +psy would metaphorically indicate a psychological accompaniement (*Maria investigated the problem with Joana*).

- **instrument**, represented as: $\lambda I\ [_{manner}\ BY-MEANS-OF([_{thing}\ I\ ])]$ (*they opened the door with a knife*). This is, in fact, a generic representation for most preposition senses introducing instruments.

Note that both senses are characterized as a manner. They are obviously contrasted by different selectional restrictions on the NP, represented by the variable I.

More subtle is the representation of *contre* (approximately 'against'). The direct usage is a physical object positioned against another one:
(1) $\lambda K\ [_{place}\ AGAINST_{+loc,c:+}([_{thing}\ K\ ])]$
where AGAINST indicates a lateral physical (+loc) contact (c:+, see(Jackendoff 90)) [3], between two objects, I and K, where I is against K, and where the agonist force exerted by I on K is balanced by the antagonist force exerted by K on I. The physical contact is the most visible; it is in the foreground, while the forces view (or facet) is rather in the background.

Against can also be used metaphorically in the epistemic or psychological domains as in: *against a theory/ a practice.* The primitive AGAINST is kept since the fundamental idea is the same: something positioned against something else. Only the semantic fields are altered in the representation, translating exactly the idea of a metaphor as being a partial isomorphism between 2 ontological domains (Lakoff et al. 80):
(2) $\lambda K\ [_{place}\ AGAINST_{+psy\vee+epist,c:-,ta:+}([_{thing}\ K\ ])]$.
In that case, however, the physical contact no longer exist (c:-), while the agonist/ antagonist force is present (ta:+, see (Jackendoff 90), slightly simplified here). (2) is a facet of this sense of *contre*, a priori in the background since metaphorical.

4.2 PPs in verbs

The second part of the exercise is to represent PPs in verbs. A first approach is simply to leave an open position with no restrictions on the PP (selectional restrictions are in the subcat frame). For example *walk* is represented as:
$\lambda I,\ \lambda J\ [_{event}\ GO_{+loc}([_{thing}\ I\ ],\ J\ )]$. [4]
However, J must be constrained to be a trajectory, of type path:

[3] In French, our analysis is that *contre* describes a position, not a path.
[4] Sometimes, a CAUSE is added, but it is not crucial for our purpose.

$\lambda I,\ \lambda J\ [_{event}\ GO_{+loc}([_{thing}\ I\ ],\ [_{path}\ J\ ])]$.
but, then, we have composition problems and a duplication of the path when J is instantiated by a PP. The most appropriate way is then to type J, e.g. as a path towards a certain target K:
(3) $\lambda I,\ \lambda J:\ [_{path}\ TOWARDS_{+loc}([_{thing \vee place}\ K\ ])]\ [_{event}\ GO_{+loc}([_{thing}\ I\ ],\ J\ )]$.
The TOWARDS in the type does not exclude any other adjunction, e.g. a source or a VIA. It represents the minimal representation of the meaning of *walk*.

Then, if Rp is the semantic representation of a PP, and Rv that of a verb, the semantic composition is trivially defined as follows, in a DCG + Montagovian style:
(4) VP(Rv(Rp)) $\rightarrow$ V(Rv), PP(Rp).

Metaphorical behaviours are managed by a set of rules describing how and under which conditions a PP can impose its semantic field type to the whole verb representation. For example, for movement verbs, the metaphor +loc $\rightarrow$ +psy is very commonly observed. Each verb class is associated with a specific set of metaphorical shifts. Verbs describing concrete events are more sensitive to metaphors than e.g. epistemic verbs.

Let us now underline a frequently encountered difficulty. In *aller contre un principe* (to go against a principle), Rv is similar to (3) and Rp is equal (2) with K=principle, this facet being activated from the type of the head NP. Rule (4) cannot be applied directly: there is a type conflict between the PP (of type place) and the verb expectation (J is of type path), with the impossibility of using TOWARDS (or TO) for a place since it describes a direction.

First, the representation of the PP (2) must remain as much stable as possible. Therefore, TOWARDS can no longer be present (it is a by-default primitive, see below), AGAINST must then be used. This is the constraint imposed by the PP. Next, an operation on types needs to coerce the conceptual category place into a path because the verb imposes its conceptual category type. This is a very common observation made for many movement verbs (e.g. push, move, throw, etc.). We have also observed that the 'skeleton' of an LCS (the structure, with the primitives and the conceptual categories) of a given verb sense cannot be altered by its arguments. The only possible changes are the semantic fields, to account for true metaphors [5], as seen above. Conversely, the form of the arguments, specified in the type of the λ-abstraction, can change within a relatively narrow domain. According to this observation, if there is no contradiction within the PP, then the conceptual category of the PP is 'coerced' to the category expected by the verb, otherwise the VP is ill-formed. For example, if the PP denotes an object or a place, it becomes the point to reach, via a path. Thus, AGAINST must be embedded into a structure of type path [6]:
(5) $\lambda I, [_{event}\ GO_{+psy \vee +epist}([_{thing}\ I\ ],$
$[_{path}\ AGAINST_{+psy \vee +epist, c:-, ta:+}([_{thing}\ mental\ or\ social\ entities\ ])]$.
There are however relatively few such alternations, indeed we cannot easily find e.g. a place into a means alternation [7].

[5] By true metaphor, we mean constructions where the ontological shift is quite regular, and observed in corpora.

[6] The 2-level structure $[_{path}\ TOWARDS([_{place}\ AGAINST(\ )\ ])]$ is meaningless and cannot be used in the +psy domain. It cannot be constructed by the λ-calculus proposed here, even with a polymorphic typing of the object argument.

[7] *He relies ON Paris to win the elections* is such as a case, but it is rather analysed as a metonymy.

5 Operations on types related to PP processing

The LCS is relatively stable under derived usages of prepositions and its instantiation remains essentially monotonic. We have identified three main phenomena:

- Substitution of a new semantic field in the representation, coming from the object NP, via the preposition, up to the verb representation, to deal with metaphors (e.g. 4.1).
- Introduction of a new primitive for representing the preposition. The variation is however restricted to prepositions of the same conceptual domain, i.e.: localization, manner, instrument (finer grained than conceptual categories of the LCS). For example, a localization cannot be substituted for a manner, but a preposition denoting a fixed position can be substituted for a preposition denoting e.g. a source or a destination. We analyze it as a limited semantic shift which slightly alters, but does not contradict, the verb expectations, as exemplified in 4.2. This shift is essentially motivated by the preposition when the verb expectations do not match with it.
- Weakening of some fragments of representation of the verb. Since the LCS has no means to express the strength of one of its components in a representation, the only solution is to leave largely open or underspecified fragments which are in the background (e.g. 4.2 for walk). Due to space limits this point won't be further elaborated.

5.1 Semantic field substitution

In an structure of type event, a cognitive constraint, proper to the LCS, imposes that all semantic fields are identical (except possibly for manners). Basic verb representations have a semantic field, describing the by-default field of the object NP. When there is a conflict, the NP imposes its semantic field and a shift must be carried out into the verb representation. This shift is limited to the structure of type event being considered, i.e. the node $[_{event}\ ]$ acts as a bounding node. The semantic field shift is constrained by the space of possible metaphorical changes the verb class or subclass may undergo.

An example is the case of *entrer* (enter), with the following basic representation:

$\lambda I, \lambda J:\ [_{path}\ INTO_{+loc}([_{place}])]$

$[_{event}\ CAUSE([_{thing}\ I\ ],\ [_{event}\ GO_{+loc}([_{thing}\ I\ ],\ J)]$.

Metaphorical usages entail substitutions, e.g.:

entrer dans une conversation (enter into a conversation): +loc $\rightarrow$ +com (communication).

entrer dans une entreprise (enter into a company): +loc $\rightarrow$ +epist $\vee$ +loc $\vee$ +psy (since one both physically and intellectually enter into a company). These fields correspond to the different facets at stake in the preposition use, inherited from the facets of *company*. They are all *a priori* foregrounded.

However, as suggested by an anonymous reviewer, in *Daniel uses the lounge as a dining room*, the lounge is a means.

The best way to formally and computationally treat this substitution is to introduce a variable for the semantic field, associated with a by-default assignment procedure. The above example becomes:
(6) $(\lambda I, \lambda J : [_{path}\ INTO_W([_{place}])]$
$[_{event}\ CAUSE([_{thing}\ I\ ],\ [_{event}\ GO_W([_{thing}\ I\ ],\ J)]),\ by-default(W\ =\ +loc).$
The default option allows us to capture the basic ontological domain of the verb, while keeping as much as possible the monotonic character of the composition process. Note that the scope of the default is the whole formula. If the NP controlled by the preposition is of a different ontological domain, then its type is preferred, provided LCS well-formedness constraints are met.

Similarly, we can introduce notations to characterize facets:
$background(P)$
indicates that P is in the background. P is any wff LCS expression. For example, for (6), we have:
$foreground(W\ =\ +loc),\ background(W\ =\ +psy \vee +epist).$
Foreground a priori only partially overlap with the *by-default* operator.

5.2 Primitive substitution from by-default statements

Prepositions can be partitioned into a few 'ontological' subsets: localization, manner, instrument, accompaniement, etc. [8] According to our observations, substitutions are constrained to occur within a subset only. They directly correspond to semantic alternations, such as the place → path seen above since they must be introduced by an appropriate conceptual category.

For example, the verb *naviguer* (sail, navigate) is basically used for boats in French:
$(\lambda I, \lambda J : [_{path}\ TOWARDS_W([_{place}\ ]),\ FIX-POS-PREP_W([_{thing}\ _\ ])])]$
$[_{event}\ GO_W([_{thing}\ I\ ],\ J\ )])$
$by-default(W\ =\ +loc),\ by-default(FIX-POS-PREP\ =\ ON).$
In the path, the direction remains underspecified and the generic preposition type FIX-POS-PREP subsumes the primitive (ON) proper to the medium, water, and infered from the subject argument, *boat*. *Naviguer* can also be used with a light metaphor for the space shuttle (called vessel). In that case the expression of the medium becomes:
$IN_W([_{thing}\ space\ ])$
This element of the path is infered from (or incorporated into, or in the Qualia of) the subject argument, unless a different medium is mentioned in the PP (e.g. *the boat sails in the fuel*).

Now, if we consider more remote metaphors such as *naviguer sur Internet* (navigate on Internet), or *naviguer dans les dossiers* (navigate in the files), we have (1) a different primitive than the by-default one (INSIDE instead of ON), (2) medium becomes the NP itself which needs to be realized since it is no longer either incorporated or inferable from other constituents, and (3) a semantic field substitution (+epist):
$[_{event}\ GO_{+epist}([_{thing}\ I\ ],$
$[_{path}\ TOWARDS_{+epist}([_{place}\ ]),\ INSIDE_{+epist}([_{thing}\ internet\ ])])])].$

[8] The status and nature of these subsets is still under elaboration and evaluation. They come along with a general semantics of prepositions.

So far, the above strategy seems satisfactory. It remains valid for the other conceptual categories postulated by the LCS such as manner, probably involving a smaller choice of prepositions. However, we are now investigating the possibility of verbs which show stronger constraints on those primitive changes.

6 Compositionality, adjuncts and complex arguments

As briefly sketched out in section 2.3, arguments are included into the predicate representation by unification while adjuncts are included by concatenation, without any reference to a λ-operator or to a type. This is parallel to syntax where arguments are direct branches of their mother node, while adjuncts are, in general, adjoined via a duplication of the mother node (or concatenated as in TAGs).

6.1 Concatenation of adjuncts

The concatenation of adjuncts in the LCS is essentially carried out by means of a rewriting rule. If R_1 is the semantic representation of the adjunct and if the structure in which the adjunct has to be embedded into is $[_{conceptual\ cat}\ R\ ]$, then:
$[_{conceptual\ cat}\ R\ ]\ +\ R_1\ \rightarrow\ [_{conceptual\ cat}\ R\ , R_1\ ]$.
The conceptual category is an event [9]. Since a predicate is described in the LCS by at most two embedded events, the application of the rewriting rule is not ambiguous: sentence adjuncts are concatenated within the highest structure of type event, whereas VP adjuncts are concatenated within the lowest structure. This is totally parallel to e.g. X^2 and X^1 adjuncts in syntax. A priori, there are no constraints on R_1, besides those imposed by the general semantics of the verb. This rule can be applied any number of times, there is no order relation or scoping relation a priori between adjuncts in the LCS.

6.2 Complex arguments

Complex arguments cover different situations, e.g.:

- coordination, either freely occuring or subcategorized by the verb (e.g. verbs that convey the idea of putting things together: mix, gather, etc.): *mix butter with eggs, milk and sugar*,
- expression of complex trajectories with e.g. vias,
- idiomatic expressions and prepositions with more than 2 arguments: *9 out of 10*, *x between a and b*.

For example, in *aller de Madrid a Barcelona par l'autoroute via Zaragoza et Lerida* (go from Madrid to Barcelona on the highway via Zaragoza and Lerida), the 'via' and 'par' PPs are analyzed as a part of the spatial trajectory. 'Highway' is viewed here

[9] In complex PPs, we may consider also concatenation of PPs, analyzed as adjuncts to others. In that case, the conceptual category can also be a path or a manner.

as a spatial element, not as a means, which it is also in a different facet. A kind of confirmation in syntax of this equal status of the PPs is the heavy-NP shift (to the right) observed for the 'via'-PP, since heavy-NP shift occurs between constituents with equal syntactic status.

The result is a single representation of type path, where all the parts of the trajectory are concatenated:

$[_{path}\ FROM_{+loc}([_{place}\ Madrid\ ]),\ TO_{+loc}([_{place}\ Barcelona\ ]),$
$\quad VIA_{+loc}([[_{place}\ Zaragoza\ ], [_{place}\ Lerida\ ],\ [_{means}\ highway]])]$

Note how the conjunction is represented. Highway is also included in the list since it is also a VIA. All the elements must be of the same semantic field (+loc) and their conceptual categories must be compatible with the notion of path (here place, but thing would also be acceptable since concrete things have a localization). This representation however fails to capture order: Zaragoza followed by Lerida. We cannot also indicate where the highway portion is. Let us assume that this is in the pragmatics.

7 Conclusion

In this document, we have proposed a computational framework to deal with the semantics of prepositions and their different sense variations and facets. Our aim was not to provide detailed semantic descriptions, but to introduce the basic mechanisms for sense description and its selection, and sense compositions in propositions. We have also outlined the hypotheses, the data and classifications required of such an approach.

Acknowledgements

We thank three anonymous reviewers for their helpful and challenging comments.

References

Boguraev, B., Spark Jones, K., (1987), *Material concerning a study of cases*, Tech. Report 118, University of Cambridge Computer Laboratory.

Busa, F., (1996), *Compositionality and the Semantics of Nominals*, PhD. Dissertation, Brandeis University, MA.

Copestake, A., Briscoe, T., (1995), Semi-Productive polysemy and sense extension, *Journal of Semantics*, vol. 12-1.

Cruse, A., (1986), *Lexical Semantics*, Cambridge University Press.

Dorr, B., (1993) *Machine Translation, a view from the lexicon*, MIT Press.

Dorr, B., Olsen, M. B., (1997), *Deriving Verbal and Compositional Lexical Aspects for NLP Applications*, Proceedings ACL'97, Madrid.

Fellbaum, C., (1993), "English Verbs as Semantic Net", *Journal of Lexicography.*

Gruber, J., (1967), *Studies in Lexical Relations*, MIT doctoral dissertation and in *Lexical Structures in Syntax and Semantics*, North Holland (1976).

Jackendoff, R. (1983), *Semantics and Cognition*, MIT Press, Cambridge.

Jackendoff, R., (1990), *Semantic Structures*, MIT Press.

Jackendoff, R., (1997), *The Architecture of the Language Faculty*, MIT Press.

Lakoff, G., Johnson, M. (1980), *Metaphors we Live By*, University of Chicago Press.

Levin, B., (1993), *English verb Classes and Alternations: A Preliminary Investigation*, Chicago Univ. Press.

Mari, A., (2000), *Polysémie et décidabilité: le cas avec ou l'association par les canaux*, PhD dissertation, EHESS, Paris, October 2000.

Ostler, N., Atkins, S., (1992), Predictable Meaning Shifts: some lexical properties of lexical implication rules, in J. Pustejovsky and S. Bergler (eds.) *Lexical Semantics and Knowledge Representation*, Springer-Verlag.

Pinker, S., (1993), *Learnability and Cognition*, MIT Press.

Pustejovsky, J., (1991), The Generative Lexicon, *Computational Linguistics*, vol 17-4.

Saint-Dizier, P., (1999), Underspecified Lexical Conceptual Structures for Sense Variations, *IWCS3 workshop*, Tilburg.

Verkuyl, H., Zwarts, J., (1992), Time and Space in Conceptual and Logical Semantics: the notion of Path, *Linguitics 30*: 483-511.

Wierzbicka, A., (1996), *Semantics. Primes and Universals*, Oxford University Press.