

A CONCEPTUAL SEMANTICS FOR PREPOSITIONS DENOTING INSTRUMENTALITY

Alda MARI
ENST – CNRS
46, rue Barrault
75013 PARIS

mari@enst.fr

Patrick SAINT-DIZIER
IRIT-CNRS
118 route de Narbonne
31062 TOULOUSE Cedex FRANCE

stdizier@irit.fr

ABSTRACT

In this paper, we present a semantic analysis and a representation for prepositions denoting instrumentality. The abstract parameters defining instrumentality are elaborated and a model of the interactions agent-object-instrument is proposed and implemented using the Lexical Conceptual Structure.

1 An analysis of the primitive notion of instrumentality and its lexicalisations

The notion of instrument seems to appear at a very early stage of the the semantico-cognitive development of children and has often been considered as a primitive notion (Wierzbicka, 1992). Moreover, at first glance, its content can be expressed by a very intuitive paraphrase: an instrument is an object used to obtain or to reach a certain goal. However, in spite of this apparent simplicity, the parameters defining this notion from a semantic and lexical point of view are extremely complex and subtle. Our study is both analytical and formal; it considers the abstract notions as well as the possible lexicalisations. This work concentrates on four representative prepositions of the notion of instrumentality in French - *par* (~by), *grâce à* (~thanks to), *au moyen de* (~by means of), *avec* (~with) - and is organised in two parts. The first part attempts to identify the abstract parameters that define the notion of instrumentality, the constraints on its lexicalisations and on its contextual values. In the second part, we suggest a model using the Lexical Conceptual Structure (LCS) (Jackendoff, 1990) and elements of the

Generative Lexicon (GL) (Pustejovsky, 1995) that we settle within a compositional framework. Our general aim is to provide under-specified representations for the abstract notion, and LCS for its representation and contextual values (usage examples). A λ -calculus for computing meanings is also provided.

2 Analysing the notion of instrumentality via its lexicalisations

2.1 Three levels of analysis: notions, senses and values

In the semantic literature, there exist two ways to capture the notion of instrumentality: the first one considers the object type expected by the preposition complement (Poncet-Montagne, 1991); the second one consists in identifying the possible relations between the sub-event denoted by the VP and the “causing” sub-event, i.e. the one involving both the subject and the preposition complement: for *Jean mange avec une cuillère* (*John eats with a spoon*) the causing sub-event is “Jean uses a spoon”.

Talmy’s famous work on force relations (Talmy, 1976) relies on this second option. Nevertheless, considering only the cognitive universal aspect of this notion, his work ignores the constraints on the structure of these sub-events, the specificities related to their possible lexicalisations and the parameters defining the control relations among the entities involved within these sub-events.

Our analysis based on event structure considers all of these parameters. Moreover, our contribution extends toward the interpretation of

the complex relations among sub-events and the entities that they involve, with respect to three levels of representation. Given the dichotomy between *notions* (or *senses*) and *values* (or *meanings*) (Mari, 2000) that we understand as a difference between semantic category and contextual instantiations (Poesio, 1996) we consider instrumentality on the following three levels of abstraction:

1. the under-specified form of the **notion** of instrumentality or **cognitive category**,
2. the under-specified forms of the possible lexicalisations of the notion: definition and representations of the **sense(s)** of *par*, *grâce à*, *au moyen de*, *avec*,
3. the contextual **values** of the sense(s) of these prepositions.

The linguistic part of this paper is organised as follows: we first define the event structure related to the notion of instrumentality (section 2.2); we then present the essential features of the data defining the senses of the four prepositions at an informal level (section 2.3); finally, we consider the parameters that serve the instantiation of this notion in order to achieve a deeper and computationally tractable formal model (section 2.4). Section 3 is devoted to model these three levels.

2.2 The event structure of the notion of instrumentality

To see how the notion of instrumentality can be decomposed into three sub-events involving entities having complex relations between each other, let us consider the following example¹:

John cuts the bread = **e3** with a knife

With: *John uses a knife* = **e2** and *the knife has the ability to cut the bread* = **e1**.

The syntactic surface structure is:

$NP_0 V NP_1 PrepInstr NP_2$

In the following discussion, *I*, *J* and *K* represent the denotations of NP_0 , NP_1 and NP_2 .

This analysis assumes the existence of the following sub-events and entities:

1. The sub-event (e_1) implying the instrument (K) and the action (V NP_1).

2. The sub-event (e_2) implying the actor / agent² (I) and the instrument (K).

3. The sub-event (e_3) implying the actor / agent (I) and the action (V NP_1).

The formula (i) makes explicit the relations existing among these sub-events:

$$(i) (e_2 (e_1)) \rightarrow e_3$$

It expresses the fact that “Because the knife has the ability to cut the bread (e_1) and that John uses the knife (on the bread) (e_2), then John cuts the bread (e_3)”. The application of e_2 on e_1 entails e_3 .

2.3 The data: informal definitions, notions of sub-event and control relations

This general instrumentality scheme is instantiated differently in the lexicalisations of the notion of instrumentality, i.e. in preposition senses. Methodologically, we begin by considering these lexicalisations to abstract later over the cognitive notion and representation of instrumentality. For each of the prepositions, we present a typical example, an informal definition, and a sub-event based paraphrase in the lines of (i).

2.3.1 Par (~by)

Typical example:

- (1) Les alpinistes ont atteint le sommet par ce chemin / The alpinists reached the top by this trail

Informal definition: “ $e_1 = K$ is in a certain disposition or state such that it can have a certain effect; $e_2 = I$ does an action on K which is of the type of e_1 or that entails e_1 ; e_3 is obtained”.

Paraphrase in terms of sub-events (1): “the trail has the property of reaching the top of the mountain (e_1), the alpinists take this trail (e_2) and so, they reach the top of the mountain (e_3)”.

2.3.2 Grâce à (~ thanks to)

Typical example:

- (2) Le tourisme prospère grâce au Canal du Midi / The tourism thrives thanks to the Canal du Midi

¹ For the sake of the explanation this example is in English

² Following Talmy (Talmy, 1976), we use the term actor to refer to an unwilling agent.

Informal definition: “ e_1 = the instrument has certain properties (that needs to be positive); e_2 = the actor (I) undergoes these properties; e_3 = the actor is beneficiary”.

Paraphrase in terms of sub-events (2): “the Canal du Midi is a touristic attraction (e_1), tourism benefits from the presence of the Canal du Midi (e_2) and thus tourism thrives (e_3)”.

2.3.3 Au moyen de (~ by means of)

Typical exemple:

(3) Il s’est brûlé au moyen d’huile chaude / He burned himself by means of boiling oil

(4) Ils ont ouvert la porte au moyen d’un cric / They opened the door by means of a jack

Informal definition: “ e_1 = the instrument (K) is such that it can perform an action; e_2 = the agent (I) controls the action that the instrument can perform; e_3 = the agent performs the action”.

Paraphrase in terms of sub-events (3): “a boiling oil can burn (e_1), John uses boiling oil to burn himself (e_2) and thus he burns himself (e_3)”.

2.3.4 Avec

Typical exemple:

(5) Jean s’est brûlé avec de l’huile chaude / John burned himself with boiling oil

Informal definition: “ e_1 = the instrument is such that it can perform an action; e_2 = the actor (or agent) controls the instrument without controlling the action that it can perform; e_3 = the agent does the action that the instrument can perform”.

Let us notice that *avec* has two possible interpretations: either I is an actor (in this case John unwillingly burns himself), or an agent (John is willing to burn himself). In this second case *avec* is a synonym of *au moyen de*. In the remainder of this paper we consider the first interpretation only.

Paraphrase in terms of sub-events (5): “boiling oil can burn (e_1), John uses the boiling oil (e_2) and he burns himself (e_3)”.

2.3.5 Summary

The following table compares the distribution of the four prepositions and the constraints they impose on their environment and that the model will have to implement.

PREP	<i>avec</i>	<i>au moyen de</i>	<i>par</i>	<i>grâce à</i>
Exemple				
Jean s’essuie les mains PREP une serviette / John dries out his hands PREP towel	X	Incompatible because <i>au moyen de</i> rules out the internal instrument.	Incompatible because <i>par</i> rules out any control on the instrument.	= <i>par</i>
Jean séduit Anne PREP sa manière de parler / John seduces Ann PREP his way of talking	Incompatible because <i>avec</i> requires an active control on the instrument.	= <i>avec</i>	X	It is compatible only if there is no control of the actor on the instrument.
Le tourisme prospère PREP au canal du midi / The tourism thrives PREP the Canal du Midi	Incompatible because <i>avec</i> requires an active control on the instrument.	= <i>avec</i>	Incompatible, because <i>par</i> requires the actor to be active in e_1 or e_3	X
Jean s’est brûlé PREP de l’huile / John burned himself PREP boiling oil	X : The control of John on the oil does not aim at the resulting action.	X : The control of John on the oil aims at the resulting action	Incompatible because <i>par</i> rules out any control on the instrument.	= <i>par</i>

3 The logical model

3.1 Main Principles of the Lexical Conceptual Structure (LCS)

The LCS owes much to the former Lexical Semantics Templates. It gained its popularity via Jackendoff's improvements. The LCS is mainly organized around the notion of motion, the other conceptual fields being derived by analogy, in a more or less natural way. We consider the LCS as a semantic representation language and as a methodology for describing the semantics of predicative forms. It is indeed clear that the primitives it is composed of are not comprehensive enough.

The LCS language is composed of three elements: conceptual categories, also called parts of speech: thing, event, state, place, path, property, purpose, manner, amount and time. These are used to type the different LCS structures. The LCS also has a number of conceptual primitives. The most important ones cover the notions of change (GO), state (BE), and cause (CAUSE). Lower level primitives mainly include prepositions: FROM, TO, AT, ON, etc. In our framework, we consider that we need 64 such primitives (Cannesson et al. 01). Finally, the LCS has semantic fields: loc, temp, poss, epist, comm, etc. designed to specialize the above set of primitives to a certain field: GO+loc is a change in the localization domain, while GO+poss is a change of possession.

LCS forms can be read quite easily, for example, the verb run is represented as follows:

[event CAUSE([thing I],
[event GO+loc([thing I], [path]))]]

which can be paraphrased as: I (the subject, and only argument) is the cause of an event which is a change of localization (GO+loc) of itself along a certain path which is left underspecified (possibly instantiated by a PP).

In the representations given below, the LCS is paired with a typed λ -calculus and underspecification, allowing for the introduction of information coming from arguments or from inferences.

3.2 Modelling the actor-agent / action / instrument relations

Let us now model the relations among the e_i presented above. For that purpose, we need to introduce two sets of primitives to characterize (1) the different levels of control that the actor / agent (I) has on the instrument (K) and (2) the degree of commitment of the instrument in the action. For example contrast *cut the meat with a knife* with *eat soup with a spoon*: in the first case the knife does the cutting whereas in the second, the spoon is just used as a tool that facilitates the action, it does not do the eating.

The actor-agent (I) / instrument (K) relation: (e_2)

The control that the actor / agent has on the instrument varies considerably and can be expressed by means of three different primitives in the LCS:

UNDERGO: the *actor* has no control on the instrument or on its properties.

SELECT: the agent uses the instrument and has some control on it. Nevertheless, while doing a certain action with the instrument, the *actor* does not necessarily plan to do the action denoted by $V NP_1$.

CONTROL: the *agent* controls the instrument, in order to obtain the action denoted by $V NP_1$.

The instrument (K) / action ($V NP_1$) relation: (e_1)

According to the commitment of the instrument in the action being performed, this relation can also be instantiated by three different primitives in the LCS:

BE: the instrument has some intrinsic properties such that even being passive, it necessarily participates to the action denoted by $V NP_1$.

REACT: the instrument, while being controlled and activated by the agent with respect to a particular property, participates to the action denoted by the $V NP_1$ via another property, unexpected and uncontrolled by the agent.

ACT: the instrument has an intrinsic property that contributes, via the agent, to the success of the action. The primitive ACT, contrary to the primitive BE, expresses the fact that the

instrument is not passive, but that it participates to the action.

The relation e₃ does not need any additional primitive.

3.3 LCS representation of preposition senses and instances

Let us now show how the meaning of these four prepositions senses can be represented. The LCS provided for the four prepositions has a very regular structure that reflects the sub-event construction typical of instrumentality (i): the first line of the general form associated with the sense of the preposition describes the nature of the control of the subject *I* on the instrument *K* (e₂) the second line accounts for the properties of the instrument (*K*) that are useful for the action to be realized (e₁), while the third line describes the action itself (e₃). As abstracted over in section 3.3, e₂ has wider scope over e₁. Note that the theme *J* is only present in the verb representation within e₃.

It is important to note that, in the calculus given below, PPs are generally analysed as propositional adjuncts: their representation embeds the verb representation and not the reverse as for arguments and adjuncts with a lower scope. Syntactic alternations provide a strong argument in favour of this interpretation: these constructions generally undergo the *Possessor-Subject (transitive)* alternation (2.13.4, Levin, 1993), also valid for French (Saint-Dizier, 1998), which clearly indicates that the PP has wider scope over the whole proposition.

In all of the following LCS, let *I*, *J*, *K* be the variables representing respectively NP₀, NP₁ and NP₂; let *T* be the ontological type of the verb *VERB* of the proposition. For each of the prepositions, we give the general form (or sense representation) and the LCS of the typical example (typical value). Additional operators are introduced below when used.

3.3.1 Par

General form:

$\lambda I, \lambda K, \lambda J$

[event CAUSE([event UNDERGO([I],
[state BE+T([K], AT+T([TELIC-OF(K, J)])])]),
[event BECOME([I], [event GO+T([I],
[path AT+T([VERB([I], [J])])])])])])]

The function TELIC-OF(*K*, *J*) extracts in the telic role of the noun *K* a predicate whose argument types are subsumed respectively by the types of *K* and *J*. The primitive BECOME characterizes accomplishments. It emphasizes the state resulting from the action described by the verb. Its general form is:

[event BECOME([I],
[event GO+T([I],
[path AT+T([VERB([I], [J])])])])])]

It is a function that elaborates the state resulting from the realization of the action described by the verb (*VERB*), within the ontological domain *T*. The *GO+T* and the *AT+T* characterize the evolution of the action, to reach the resulting state via a kind of metaphorical path. *T* is also the ontological domain of the resulting state.

Finally, we leave the verb representation open, indicating only its two arguments *I* and *J*.

Typical example:

(1) *Les alpinistes ont atteint le sommet par ce chemin/ The alpinists reached the top by this trail.*

[event CAUSE([event UNDERGO([alpinistes],
[state BE+loc([chemin],
AT+loc([TELIC-OF(chemin, sommet)])])]),
[event BECOME([alpinistes],
[event GO+loc([alpinistes],
[path AT+loc([atteindre(alpinistes,sommet)
])])])])])]

The TELIC-OF function produces here, for example, the predicate :

Passer-par(chemin, sommet)
(go-via(trail, top))

3.2.2 Grâce à

General form :

$\lambda I, \lambda K, \lambda J$

[event CAUSE([event UNDERGO([thing I],
[state BE+T([K],
[property TELIC-OF(K, J)])]),
[event BECOME ([thing I],
[state VERB([I], [J])])])])]

Typical example:

(2) *Le tourisme prospère grâce au Canal du Midi./ Tourism thrives thanks to the Canal du Midi*

[event E2([I], [event/state E1([K],
[prop TELIC-OF(K, J) or VERB])]),
[event E3([I], [state resulting-state(VERB)])]]

E2 = UNDERGO / SELECT / CONTROL

E1 = BE / REACT / ACT.

E3 = CAUSE / BECOME

As can be noted from the chart in 2.3.5 and from the examples in section 3.2, each preposition sense has its own selectional constraints and representation.

4. Conclusion

In this paper, we proposed an analysis of the notion of instrumentality, going from the abstract notion to its lexicalisations via preposition senses. A symmetric movement has then been suggested, from the representation of examples in LCS with their application constraints to the under-specified representation of instrumentality, via abstract representations of preposition senses. This analysis shows the complexity of the notion and the necessity of using complex knowledge such as the one found in telic roles, among others. A considerable amount of work, systematisation and development of examples (including metaphors and metonymies) remains to be done in this domain. However, we believe that this work, through a concrete study of a complex notion, induces analysis, methods and semantic representation formalisms appropriate for developing a general framework for a proper preposition semantics.

This work is a first effort towards the definition of an accurate semantics for a number of preposition classes which have seldom being studied within a computational linguistics perspective. The next step is to study prepositions denoting means and manners. Similarly to prepositions denoting instrumentality, this study also involves related studies such as metonymic forms (treated here by calls to the telic role of the argument), compositionality (with the verb and the NP), the expression of selectional restrictions, and different forms of knowledge representation and inference, among which, as advocated here, the generative lexicon.

Besides an in-depth analysis of prepositions, our aim is to introduce such an approach in a number of applications where prepositions play or should play a major role. Let us first mention machine translation where it is often useful to go as deep as interlingua forms (Dorr et al 97) to get correct translations. Prepositions should in the future play a major role in knowledge extraction since the compound preposition + noun type is a clear and quite simple trigger of a semantic information such as localization, manner, instrument, accompaniment or expression of an approximation (Cannesson et al. 01). Finally, let us mention the area of natural language generation where preposition choice, an aspect of lexicalisation, is a delicate task. It also interacts much with syntax, in particular with alternations as advocated above, and also with various forms of verbal incorporation.

We believe that such as detailed analysis of prepositions is useful to guarantee a certain level of quality and adequacy of computational linguistics applications which do not rely only e.g. on stochastic observations. Although prepositions have a certain semantic and syntactic autonomy, we also believe that their semantics must be investigated in close connection with the verb and the NP semantics.

Acknowledgements

We thank the numerous native speakers that helped us to constitute a corpus of uses that allowed us to stabilize our analysis.

References

- Cannesson, E., Saint-Dizier, P. (2001), A general framework for the representation of prepositions in French, ACL01 WSD workshop, Philadelphia.
- Dorr, B., Olsen, M.B., (1997) Deriving Verbal and Compositional Verbal Aspect for NLP Applications, proc. ACL'97, Madrid.
- Jackendoff, R., (1990), *Semantic Structures*, MIT Press.
- Levin, B., (1993), *Verb Semantic Classes: a Preliminary Investigation*, Chicago University Press.
- Mari, A. (2000). *Polysémie et Décidabilité. Le cas de avec ou l'association par les canaux*. Thèse de Doctorat, EHESS, Paris. A paraître,

- collection Langue et Parole, L'Harmattan, 2003.
- Poesio, M. (1996). Semantic Ambiguity and Perceived Ambiguity. In K. van Deemter & S. Peters, (eds.), *Semantic Ambiguity and Underspecification*. Stanford: CSLI Lecture Notes, pp. 159-201.
- Poncet-Montange, A. (1991). A propos des noms d'instruments: relations entre forme et sens. *Linguisticae Investigationes XV: 2*, pp. 305-323.
- Pustejovsky, J., (1995), *The Generative Lexicon*, MIT Press.
- Saint-Dizier, P. (1998), Alternations and Verb Semantic Classes for French, in *Predicative Forms for NL and LKB*, P. Saint-Dizier (ed), Kluwer Academic.
- Talmy, L. (1976). Semantic Causative Types. In M. Shibatani (éd.), *Syntax and Semantics 6: The Grammar of Causative Constructions*. New York: Academic Press, pp. 43-116.
- Wierzbicka, A. (1992b). Semantic Primitives and Semantic Fields. In A. Lehrer & E.F. Kittay (eds.), *Frames, Fields and Contrasts*. Hillsdale: Lawrence Erlbaum Associates, pp. 208-227.