

Syntax and Semantics of Prepositions

Text, Speech and Language Technology

VOLUME 29

Series Editors

Nancy Ide, *Vassar College, New York*

Jean Véronis, *Université de Provence and CNRS, France*

Editorial Board

Harald Baayen, *Max Planck Institute for Psycholinguistics, The Netherlands*

Kenneth W. Church, *AT & T Bell Labs, New Jersey, USA*

Judith Klavans, *Columbia University, New York, USA*

David T. Barnard, *University of Regina, Canada*

Dan Tufis, *Romanian Academy of Sciences, Romania*

Joaquim Llisterri, *Universitat Autònoma de Barcelona, Spain*

Stig Johansson, *University of Oslo, Norway*

Joseph Mariani, *LIMSI-CNRS, France*

The titles published in this series are listed on www.springeronline.com.

Syntax and Semantics of Prepositions

Edited by

Patrick Saint-Dizier

CNRS, Toulouse, France

 Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3899-2 (PB)
ISBN-13 978-1-4020-3899-0 (PB)
ISBN-10 1-4020-3849-6 (HB)
ISBN-13 978-1-4020-3849-5 (HB)
ISBN-10 1-4020-3873-9 (e-book)
ISBN-13 978-1-4020-3873-0 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springeronline.com

Printed on acid-free paper

All Rights Reserved

© 2006 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed in the Netherlands

Contents

Preface	xi
Acknowledgments	xiv
1	
Introduction to the Syntax and Semantics of Prepositions	1
<i>Patrick Saint-Dizier</i>	
1. The class of prepositions	1
2. About the syntax of prepositions	2
3. Polysemy and sense restrictions	10
4. Representing the semantics of prepositions	12
5. Prepositions and multilingualism	20
6. Overview of the book structure	20
References	23
2	
Preposition Contractions in Quebec French	27
<i>Luc V. Baronian</i>	
1. Introduction	27
2. Facts	28
3. Two possible analyses	31
4. External evidence for Analysis 2	32
5. Core linguistic arguments for Analysis 2	33
6. Conclusion: consequences of the analysis	36
Acknowledgments	40
References	41
3	
The A's and BE's of English Prepositions	43
<i>Andrew McMichael</i>	
1. Some definitions	43
2. Corpus data	44
3. The general formative principle	45
4. Adverb formation	47
5. The A's: a more complicated origin	47

6.	Simplex and compound prepositions: a classification	48
7.	The prefixed Gp as a cognitive functional marker	49
8.	Cognitive schemata of grammaticalising prepositions: an alternative categorisation	49
9.	Extensions of the Pattern	52
10.	Language Typology	52
11.	Conclusion	54
	References	55
4		
	Typological Tendencies and Universal Grammar in the Acquisition of Adpositions	57
	<i>David Stringer</i>	
1.	Introduction	57
2.	A monkey, a parrot and a banana	58
3.	Lexical variation and single syntax	62
4.	Conclusion	66
	References	67
5		
	Multilingual inventory of interpretations for postpositions and prepositions	69
	<i>Mikel Lersundi and Eneko Agirre</i>	
1.	Previous work	70
2.	Method to obtain the inventory and the multilingual table	72
3.	Case study with the Basque instrumental postposition	74
4.	Overall results	77
5.	Remaining problems.	77
6.	Conclusions and future work	79
	Acknowledgments	81
	References	81
6		
	German prepositions and their kin	83
	<i>Martin Volk</i>	
1.	Introduction	83
2.	German prepositions	84
3.	Conclusions	93
	References	94
	Appendix: Prepositions	96
	Appendix: Contracted Prepositions	98
	Appendix: Pronominal Adverbs	98
	Appendix: Reciprocal Pronouns	99

<i>Contents</i>	vii
7	
Directionality Selection	101
<i>Marcus Kracht</i>	
1. Introduction	101
2. Modes	102
3. One Word — Three Meanings	104
4. Selection	106
5. Significance for Interpretation	108
6. Predicting Selectional Properties	110
7. Mode Heads: Evidence from Mari	112
8. Conclusion	113
References	113
8	
Verb-Particle Constructions in the World Wide Web	115
<i>Aline Villavicencio</i>	
1. Introduction	115
2. VPCs in a Nutshell	118
3. VPCs and Dictionaries	118
4. VPCs and Corpora	121
5. VPCs in the Web	122
6. Conclusions	127
Acknowledgments	128
References	128
9	
Prepositional Arguments in a Multilingual Context	131
<i>Valia Kordoni</i>	
1. Introduction	131
2. The Data	132
3. Previous Accounts in HPSG	136
4. Indirect Prepositional Arguments: The Analysis	137
5. Conclusion	142
Acknowledgments	143
References	143
10	
The syntax of French <i>à</i> and <i>de</i> : an HPSG analysis	147
<i>Anne Abeillé, Olivier Bonami, Danièle Godard and Jesse Tseng</i>	
1. Introduction	147
2. Syntactic properties	150
3. Proposed HPSG analysis	154
4. Concluding remarks	159
References	161

11		
In Search of a Systematic Treatment of Determinerless PPs		163
<i>Timothy Baldwin, John Beavers, Leonoor van der Beek, Francis Bond, Dan Flickinger and Ivan A. Sag</i>		
1. Introduction		163
2. The Syntax of Determinerless PPs		165
3. The Semantics of Determinerless PPs		169
4. Analysis		172
5. Conclusion		176
Acknowledgments		177
References		177
12		
Combinatorial Aspects of Collocational Prepositional Phrases		181
<i>Beata Trawiński Manfred Sailer and Jan-Philipp Soehn</i>		
Introduction		182
1. Syntactic Aspects		182
2. Semantic Aspects		185
3. Irregular Combinations		190
4. Summary		193
Acknowledgments		194
References		195
13		
Distributional Similarity and Preposition Semantics		197
<i>Timothy Baldwin</i>		
1. Introduction		197
2. Calculating inter-preposition similarity		199
3. Gold standard sources of inter-preposition similarity		200
4. Evaluation		203
5. Related research		206
6. Conclusion		207
Acknowledgments		207
References		208
14		
A Computational Model of the Referential Semantics of Projective Prepositions		211
<i>John Kelleher and Josef van Genabith</i>		
1. Introduction		212
2. The Challenges		212
3. Previous Computational Work		215
4. The LIVE Model		216
5. Conclusions		226
References		227

<i>Contents</i>	ix
15	
Ontology-Based Semantics for Prepositions	229
<i>Per Anker Jensen and Jörgen Fischer Nilsson</i>	
1. Introduction	229
2. Formal ontologies	231
3. The relation between lexicon and ontology	232
4. Formal meaning ascription	234
5. Generative ontologies with feature structures	235
6. Ontological affinities and generative ontologies	236
7. Compositional ontological semantics for nominals	237
8. Prepositions and semantic roles in Danish	238
9. Identifying paraphrases	241
10. Conclusion	242
Acknowledgments	243
References	243
16	
Analysis and Interpretation of the Japanese Postposition <i>no</i>	245
<i>Ryusuke Kikuchi and Hidetosi Sirai</i>	
1. Introduction	246
2. Syntactic Analysis of <i>No</i>	247
3. Semantic Analysis of <i>No</i>	248
4. Framework of Context-Dependent Interpretation — SDRT	249
5. Case Study	250
6. Discussion	256
7. Conclusion	258
References	259
17	
What do the notions of instrumentality and of manner have in common?	263
<i>Alda Mari</i>	
1. Aim and methodology	263
2. Analysis of <i>avec</i> -instrument and <i>avec</i> -manner	265
3. The model: properties and constraints	274
4. A model for <i>avec</i> -instrument and manner	279
5. Conclusion	283
References	286
18	
A Conceptual Semantics for Prepositions denoting Instrumentality	289
<i>Alda Mari, Patrick Saint-Dizier</i>	
1. An analysis of the primitive notion of instrumentality and its lexicalizations	289
2. Analysing the notion of instrumentality via its lexicalizations	291

3.	The logical model	295
4.	LCS representation of preposition senses and instances	298
5.	Conclusion	302
Acknowledgments		304
References		304
19		
Prepositions in Cooperative Question-Answering Systems: a Preliminary Analysis		307
<i>Farah Benamara and Véronique Moriceau</i>		
1.	Introduction	307
2.	Preposition use in Question Answering : the WEBCOOP system	309
3.	Semantic Representation and Interpretation of Localization Prepositions in WEBCOOP	311
4.	Reasoning with Localization Prepositions	317
5.	Generating Prepositions and PPs	323
6.	Conclusion	328
References		329
Index		331

Preface

A great deal of attention has been devoted in the past ten years in the linguistic and computational linguistics communities to the syntax and the semantics of nouns, verbs and also, but to a lesser extent, to adjectives. Related phenomena such as quantification or tense and aspect have motivated a number of in-depth studies and projects. In contrast, prepositions have received less attention. The reasons are quite clear: prepositions are highly polysemic, possibly more so than adjectives, and linguistic realizations are extremely difficult to predict, not to mention the difficulty of identifying cross-linguistic regularities. Furthermore, a number of languages do not use prepositions or postpositions (or make a limited use of them) and prefer other linguistic forms such as morphological marks, e.g. case marks.

Let us mention, however, projects devoted to prepositions expressing space, time and movement in artificial intelligence and in natural language processing, and also the development of formalisms and heuristics to handle prepositional phrase attachment ambiguities. Prepositions are also present in subcategorization frames of predicative lexical items, but often in an informal and coarse-grained way. Let us also mention the large number of studies in psycholinguistics and in ethnolinguistics around specific preposition senses. Finally, prepositions seem to reach a very deep level in the cognitive-semantic structure of the brain: cognitive grammar developers often use prepositions in their metalanguage, in order to express very primitive notions. An important and difficult question to address, is whether these notions are really primitive or can be decomposed and lexically analysed.

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. Languages like English have verbal compounds that integrate prepositions (compositionally or as collocations) while others, like Romance languages or Hindi either incorporate the preposition or include it in the prepositional phrase. All these configurations are semantically as well as syntactically of much interest.

Prepositions turn out to be a very useful category in language, it does not just play the role of a grammatical marker. Prepositions are essential 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.

Prepositions are also closely related to semantic structures such as thematic roles, semantic templates or frames, and subcategorization frames. 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.

This book emerges from a workshop on the syntax and semantics of prepositions, organized in Toulouse in September 2003. The aim of this workshop was to bring together linguists, NLP researchers and practitioners, and AI people in order to define a common ground, to advance the state-of-the-art, to identify the primary issues and bottlenecks, and to promote future collaborations. The main topics were:

- The syntax of prepositions: formal or descriptive syntax, prepositions in alternations, principles in the syntax of PPs, syntactic and semantic restrictions. General syntactic-semantic principles. Postpositions or other equivalent markers (e.g. case).
- Descriptions: Potential WordNet / EuroWordNet descriptions of preposition uses, productive uses versus collocations, multi-lingual descriptions: mismatches, incorporation, divergences. Prepositions and thematic roles, prepositions in semantic frameworks (e.g. Framenet.).
- Cognitive or logic-based formalisms for the description of the semantics of prepositions, in isolation, and in composition / confrontation with the verb and the NP. Compositional semantics. Logical and reasoning aspects.
- Cognitive or logic-based formalisms for the description of the semantics of prepositions, in isolation, and in composition/confrontation with the verb and the NP. Compositional semantics. Logical and reasoning aspects.
- The role of prepositions in applications, in particular: in machine translation, in information extraction, and in lexicalization in language generation.

- Corpus-based studies that support or challenge any of the approaches described above.
- Lexical knowledge bases and prepositions. Prepositions in AI, KR and in reasoning procedures.

The Program Committee, that we warmly thank, was the following :

Nicholas Asher (Austin)
Pushpak Bhattacharyya (IIIT, Mumbai)
Harry Bunt (Tilburg)
Nicoletta Calzolari (Pisa)
Bonnie Dorr (Maryland)
Christiane Fellbaum (Princeton)
Claire Gardent (CNRS, Nancy)
Betsy Klipple (Upenn)
Alda Mari (ENST, Paris)
Palmira Marraffa (Lisboa)
Martha Palmer (Upenn)
Patrick Saint-Dizier (IRIT, Toulouse)
Gloria Vazquez (Lerida)
Laure Vieu (IRIT, Toulouse)

PATRICK SAINT-DIZIER

Acknowledgments

We thank the programme committee, cited in the preface, for their useful contributions, and two anonymous reviewers who made useful comments on the whole book.

We also thank our 2 sponsors : *Université Paul Sabatier* and the *GDR-CNRS Sémantique* and the Organizing Committee members: Dr. Farah Benamara, Véronique Moriceau and Sara Mendès.

Chapter 1

INTRODUCTION TO THE SYNTAX AND SEMANTICS OF PREPOSITIONS

Patrick Saint-Dizier

IRIT-CNRS

118 Route de Narbonne 31062 Toulouse, France

stdizier@irit.fr

Abstract This first chapter presents basic issues related to preposition syntax and semantics. It introduces different ways to view the syntax of prepositions: relational, functional and lexical. It also shows the high degree of polysemy of a number of prepositions and develops some directions to deal with preposition semantics, in particular designed for natural language processing systems, based on the Lexical Conceptual Structure and underspecification.

Keywords: syntax and semantics of prepositions.

1. The class of prepositions

Prepositions do not exist in all languages. While some languages, such as Indian languages (Hindi, Telugu, Tamil, etc.), have postpositions rather than prepositions, but this may be viewed as a rather minor distinction, other languages do not have prepositions but e.g. morphological marks such as cases, which play an equivalent role. Prepositions do not form a strict closed class of elements, as sometimes hastily presented by grammarians. Most languages with prepositions have a rather limited set of single word prepositions, in general between 40 and 120, although there are divergences among grammarians on the exact nature and definition of a preposition. In addition, there is quite large number of prepositional compounds, i.e. structures that play the role of prepositions, that include nouns (*sur le côté de*, *on the left of*, *al lado de* (Fr., Eng., Sp.)), adjectives (*proche de*, *close to*) or gerundives (*se rapportant à*, *with respect to*). Finally, preposition uses are very different from one language

to another, even within closely related languages in a linguistic family, with often a large number of idiosyncratic constructions: *dream about*, *rêver de* (litt. dream ‘of’ in French), *soñar con* (litt. dream ‘with’ in Spanish and in Portuguese). Not surprisingly, a number of prepositions are highly polysemic, almost comparable to the most polysemous adjectives like *good*.

The fact that cases or other morphemes or affixes are used in some languages instead of prepositions indicates that prepositions have specific relations with other types of linguistic mechanisms. Let us now investigate the different roles played by prepositions from a syntactic and semantic point of view.

Prepositions can first be viewed as a **functional category** in syntax: they are heads of prepositional phrases. The preposition then hierarchically dominates the noun phrase. Prepositions can also be viewed as a **semantic relation** between a structure that precedes it (e.g. a verb) and another one that follows it (e.g. an NP). This relation can be represented as a conceptual relation, as shall be seen below. Finally, prepositions can be viewed as a **lexical category** that imposes both a categorial (structure level) and a semantic selection (semantic restriction level). Similarly to the other predicative categories, prepositions have type restrictions on their arguments, they assign thematic roles, and they have a semantic content, possibly underspecified. The only difference with the other open-class categories like nouns, verbs or adjectives is that they do not have any morphology. These considerations show the central role played by prepositions in the proposition and their fundamental predicative and relational nature.

In the following sections we present some aspects of the syntax and the semantics of prepositions. These are basic notions meant for the reader unfamiliar with prepositions. A number of these notions are further developed in the following chapters for particular classes of prepositions, or for particular languages.

2. About the syntax of prepositions

There are only about 50 prepositions in English (for other languages there is not always a consensus on what a preposition is, e.g. vs. prepositional compounds). Here is a fairly complete list: *aboard, about, above, across, after, against, along, amid, among, anti, around, as, at, before, behind, below, beneath, beside, besides, between, beyond, by, despite, down, during, except, excepting, excluding, following, for, from, in, inside, into, like, near, of, off, on, onto, opposite, outside, over, past, per, plus, round, save, since, than, through, to, toward, towards, under, underneath, unlike, until, up, upon, versus, via, with, within, without*.

In this section, we investigate the different facets of the syntax of prepositions: phrasal constructions with prepositions, prepositions as relations, prepositions as thematic role assignators and prepositions in alternations.

2.1 Preposition distribution in English and French

Before going into the details of the syntax of prepositions, let us say a few words about preposition distribution, illustrated here on English and French.

The WFWSE web site indicates that English prepositions (on the lexeme basis) are distributed as follows in ordinary, everyday English. Among the 30 most frequent words in English, there are 9 prepositions:

Fig. 1a - Preposition uses in English	
preposition	rank
OF	2
IN	5
TO	8
FOR	11
WITH	13
ON	16
BY	18
AT	20
FROM	29

Rank indicates here the usage rank of the term all words considered. For example, of is the second most frequently used word in English.

For French, we have collected 14656 preposition usages from various corpora, their relative occurrence frequencies, within the set of all French prepositions, are distributed as follows:

Fig. 1b - Preposition frequencies in French		
preposition	occurrences	frequency (%)
DE, DES, D', DU (of)	8338	57
A, AU, AUX (at, to)	1649	11.2
EN (of)	856	5.8
POUR (for)	719	4.9
SUR (on)	704	4.8
DANS (in)	462	3.1
PAR (by)	413	2.8
AVEC (with)	280	1.9
ENTRE (between)	85	0.57
VERS (towards)	67	0.46
SOUS (under)	66	0.45
CONTRE (against)	62	0.44

The other prepositions (e.g. *east of*, *above*, *along*) occur less than 50 times, in general less than 10 times. If we do not take into account DE and A and

their morphological variants, frequencies need to be multiplied by 3.14 (no relation with the number π , though). The observation is that 16 prepositions occur more than 1%. They are not necessarily the most polysemic ones (e.g. *entre* (between) is not very polysemic).

2.2 Phrasal verbs

Phrasal verbs, also called prepositional verbs are verb + preposition constructions. These constructions may range from purely idiosyncratic forms (*boil down*) to compositional ones (*switch on*, *run into*). In the first example, the preposition has an intransitive use, whereas in the second it has a transitive use, where the NP is missing, possibly elliptical, but can be reconstructed, e.g. via inference. Other cases include, for example, making explicit an information which would by-default be incorporated. For example, in *climb down*, the preposition ‘down’ is made explicit because the by-default incorporated preposition is *up*.

Non compositional phrasal verbs are common, for example, in English and German (e.g. *ab-stammen*, *auf-nehmen*); they are less frequent in Romance languages, which mainly allow transitive uses (*Il est tombé dessus*, he fell on). Non compositional verb + preposition compounds, also termed verb particle construction (see e.g. Villavicencio, this volume), are often viewed as a lexical unit *per se*, which can subcategorise for a PP or an NP, as in:

(1a) *John switched on the light*

where ‘switched on’ subcategorises for an NP. In the case of a phrasal verb where the association verb + preposition is compositional, a useful (but not systematic) test is that the order of the preposition and the NP can be switched around:

(1b) *John switched the light on.*

Which is neither possible with idiosyncratic forms:

* *This talk boils to very few concrete propositions down.*

nor with prepositions in regular PPs:

* *Mary is waiting John for.*

In most computational linguistics approaches, phrasal verbs are considered as separate lexical units: their subcategorization frame(s), possible alternations and other syntactic properties are described in dedicated lexical entries. It is indeed very difficult to generalize lexical behavior for a given preposition and all the verbs with which it can be combined.

2.3 Prepositions as relations

In general, prepositions introduce a relation between two entities or sets of entities. The first entity is often a kind of external argument while the second

one is headed by the preposition. In *Mary goes to school*, *to* has two arguments: Mary (external) and school: *to*(Mary, school). Mary is an argument shared with the verb *go*.

Prepositions select in general NPs but also sometimes propositions. In some cases, NPs or propositions can be omitted, they are however implicit and can be inferred from the context. Prepositions, as shall be seen below, have their own selectional restrictions. In a VP construction (V PP), the selectional restrictions imposed by the verb on its indirect object (PP) must in some way coincide with the type of the PP (e.g. direction, instrument) and with the type of the NP within the PP. Consider a simple illustration:

(2) *to run to school*

In (2), *run* requires a path, probably underspecified w.r.t. the area in which it occurs. This requirement is met by the preposition *to*. In turn, *to* expects an NP of type: closed, well-delimited, possibly large, space. *School* meets these requirements.

Prepositions such as *around*, *out*, *in*, *away* can be used with empty objects: *go away*, *stroll around*, even if the object is in fact implicit, possibly vague. Prepositions such as *in*, *into*, *without* select an NP complement: *in the room*, *without sugar*, while prepositions such as *out*, *from* can select NP or PP complements: *from under the table*, *out in the streets*. Finally, prepositions such as *between* select a plural NP: *between John and Mary*, *between my 5 best friends*. Finally, a few authors tend also to consider that prepositions such as *from* or *down* select two NPs, as in *from A to B*, *down A to C*. We think this analysis is not correct because e.g. *from* only selects A. The expressions *from A to B* must be analysed as a compound of type *trajectory* where *from* and *to* play an equivalent role.

Besides the NP or PP it selects, a preposition has a kind of ‘external’ argument, possibly shared with another predicate, which is the first element of the relation:

(a book) on (the table).

(3) (Mary) entered into (the opera house).

A more complex case includes two intertwined relations:

(4a) (Max) steals sweets from (behind the counter). and

(4b) Max steals (sweets) from behind (the counter).

In (4a), *steals* indeed expects e.g. a kind of trajectory describing the path followed by the stolen object, whereas in (4b) *sweets* are analysed as being in a fixed position, specified by the preposition *behind*.

Prepositions can, similarly to verbs, be associated with a subcategorization frame where the first element of the frame is in general shared with another predicate, in general a verb. We can then have in a lexical entry the following description for the accompaniment sense of *with*:

with : [NP, NP]

Selectional restrictions can be added to that frame, on each argument position, with the same well-known accuracy problems as for verbs. In particular, a number of preposition senses (at least half of them) can be subject to several forms of metaphors (Moriceau et al. 03).

2.4 Prepositions and thematic roles

Thematic roles are abstract labels that characterize the semantic relations between predicates and their arguments. Each argument of a predicate is marked with a thematic role, which indicates, in a very general way, the ‘semantic’ role played by the argument with respect to the predicate. From this point of view, thematic roles can be considered as a first level of semantic representation, of much interest, for example, in knowledge extraction, where it may not be possible to go much deeper in the semantics, due to the size of the explored documents. Thematic roles have been subject to many controversies, and there is still little agreement on their nature, definition, and role in linguistic theories (Gruber 67), (Jackendoff 87), (Rappaport and Levin 88), (Roca 92), (Ravin 90).

Here is a partial list of roles, which is however generally agreed upon:

- **agent**: the entity who intentionally initiates, makes or originates the action described by the predicate,
- **patient**: the entity that undergoes the action described by the predicate, it is often an animate entity,
- **theme**: the entity moved (in a very general sense) as a consequence of the action expressed by the predicate, it is often a non-animate entity,
- **experiencer**: the entity that experiences some psychological state resulting from the predicate,
- **goal, source, location**: are roles related to spatial, temporal or abstract fields, expressing respectively the goal, the source or the position of a temporal, spatial or abstract entity.

Thematic roles are postulated by a number of authors to be universal, non-ambiguous, and to cover the whole spectrum of the predicate-argument relationships. This is certainly somewhat optimistic. Thematic roles are essentially assigned to NPs, by verbs, prepositions and VPs via predication. Their uses and meanings may be either direct or metaphorical. For example, meteorological forces are often metaphorically assimilated to agents: *the wind broke the window*.

Prepositions can be associated with a thematic grid, which contains in general one role per argument position, but multiple assignments are also possible

when two or more roles are relevant, as for verbs. For example, the following prepositions have the following grids:

on: [theme, location].

between: [theme, location].

towards: [theme \vee agent, goal].

In general, a preposition assigns a thematic role to its ‘object’ argument, i.e. the argument in the scope of the PP it heads. Therefore, *towards* assigns the role goal to its object NP. A preposition being a relation, it is also necessary to take into account another argument, the first argument of the relation, a kind of ‘external’ argument, that the preposition shares in general with the verb of the proposition (or another type of predicate). This latter argument gets thematic roles from at least two sources, which must, obviously, be compatible.

Thematic roles can be defined a priori as by-default roles, which are assigned in sentences when there is no contradiction. However, in a number of situations, they can be revised, in particular in sense extensions, for example goal can become location. This is typical in systems with more refined thematic role typologies (Boguraev 79), (Dowty 89, 91), (Saint-Dizier 99). For example, in:

The arrow moves towards the target: the external argument is a theme,

John runs towards the restaurant: John is an agent.

This problem can be solved by leaving the first role underspecified or by listing all the possibilities in the lexical entry of *towards*. In general, the role(s) mentioned a priori is(are) the most prototypical.

2.5 Prepositions and PP attachment ambiguities

Since the very beginning of language processing techniques, the management of PP-attachment ambiguities has been a real challenge, for which no fully satisfactory solution has ever been proposed. One of the reasons is that resolving such ambiguities often requires non trivial contextual inferences, similarly to e.g. reference resolution (remember the well-known example *I saw a man with a telescope in the park*).

However, the development of large ontologies, used to type in a relatively accurate way predicate arguments and the introduction of heuristics or preferences (based e.g. on statistical analysis and learning techniques) allowed significant progress in this area. If attachment cannot be resolved at parse time, a common approach is to produce a syntactic (or semantic) representation that allows the representation of the ambiguity (e.g. by means of multiple links in syntactic trees which become locally graphs). The ambiguity may then be resolved during the interpretation.

2.6 Prepositions in syntactic alternations

In her book, Beth Levin (Levin 93) shows, for a large set of English verbs (about 3200), the correlations between the semantics of verbs and their syntactic behavior. More precisely, she shows that some facets of the semantics of verbs have strong correlations with the syntactic behavior of these verbs and with the interpretation of their arguments. This very important work emerged from the synthesis of specific investigations on particular sets of verbs (e.g. movement verbs), on specific syntactic behaviors and on various types of information extracted from corpora. Other authors have studied in detail the semantics conveyed by alternations, e.g. (Pinker 89) and the links between them (Goldberg 94).

2.6.1 The alternation system. An alternation, roughly speaking, describes a change in the realization of the argument structure of a verb. The scope of an alternation is the proposition. Modifiers are considered in some cases, but the main structures considered are the arguments, including prepositions, and the verb. Arguments may be deleted or ‘moved’, NPs may become PPs or vice-versa, and some PPs may be introduced by a new preposition. Alternations may also be restricted by means of constraints on their arguments.

Beth Levin has defined 79 alternations for English. They basically describe ‘transformations’ from a ‘basic’ form. However, these alternations have *a priori* little to do with the assumptions of Government and Binding theory and Movement theory, in spite of some similarities. The form assumed to be basic usually corresponds to the direct realization of the argument structure, although this point of view may clearly be subject to debate. Here are now a few types of alternations, among the most common ones. References about works establishing these relations can be found in (Levin 93).

The *Transitivity alternations* introduce a change in the verb’s transitivity. In a number of these alternations the subject NP is deleted and one of the objects becomes the subject, which must be realized in English. The *Middle alternation* is typical of this change:

John cuts the cake → *The cake cuts easily*.

As can be noticed, it is often necessary to add an adverb to make the sentence acceptable. The *Causative/inchoative alternation* (Levin 93) concerns a different set of verbs:

Edith broke the window → *The window broke*.

Verbs undergoing this alternation can roughly be characterized as verbs of change of state or position.

Under the transitivity alternations fall also alternations where an object is unexpressed. This is the case of the *Unexpressed object alternation* where the object1 is not realized. A number of verbs undergo this alternation. In most

cases, the ‘typical’ object is somewhat ‘implicit’ or ‘incorporated’ into the verb, or deductible from the subject and the verb. This is the case, e.g., for the *Characteristic property of agent alternation*:
This dog bites people → *This dog bites*.

2.6.2 Alternations involving prepositions. An interesting alternation, with a heavy semantic impact, is the *conative alternation* that changes the object NP into a PP introduced in English by *at* (*sur* in French), as in:
Edith cuts the bread → *Edith cuts at the bread*.

A second set of alternations deals with changes within the arguments of the VP. One of the most popular alternations is certainly the *Dative alternation* which concerns verbs of giving, of future having, of transfer, etc., as in:

Edith hands the baby a toy ↔ *Edith hands a toy to the baby*

(we use the symbol ↔ for some examples when we feel that both forms have equal status, i.e. one or the other could be considered as basic). The same phenomenon occurs for the *Benefactive alternation*:

I carve a toy for the baby ↔ *I carve the baby a toy*.

The *Spray/Load alternation* involves the permutation of the arguments in the VP and the preposition alternation on ↔ with:

to spray paint on the wall ↔ *to spray the wall with paint*.

English is particularly rich in this type of phenomenon. Let us note also the *Material / product alternation*:

Martha carves a toy out of a piece of wood ↔ *Martha carves a piece of wood into a toy*,

and the *With / Against alternation*:

to hit a stick against the fence ↔ *to hit the fence with a stick*.

2.6.3 The location alternations. The location alternations, a family of alternations which involve a permutation of object1 and object2 and a preposition change, are also of much interest. The participation to certain of these alternations allows one to predict the type of motion and the nature of the end state. Verbs which focus only either on the motion (e.g. *pour*) or on the resulting state (e.g. *fill*) do not alternate. Verbs that alternate constrain in some manner both motion and end state. Let us now specify in more depth these constraints, since in fact quite a few verbs do alternate.

For example, let us consider the *into/with* alternation. (Pinker 89) differentiates among verbs which more naturally accept the *into* form as their basic form and which alternate with a *with* form. Their general form is:

Verb NP(+theme) onto NP(+destination), and they alternate in:

Verb NP(+destination) with NP(+theme).

Load hay into the wagon / Load the wagon with hay.

Other verbs more naturally take the location/container as object (e.g. *stuff*), their basic form is more naturally:

Verb NP(location) with NP(+theme), and alternate in:

Verb NP(+theme) onto NP(+destination).

stuff the truck with hay / stuff hay onto the truck.

Verbs which undergo the ‘into/onto’ alternation, have one of the following properties: simultaneous forceful contact and motion of a mass against a surface (brush, spread, ...), vertical arrangement on a horizontal surface (heap, pile, stack), force is imparted to a mass, causing ballistic motion along a certain trajectory (inject, spray, spatter), etc. Those which do not alternate have, for example, one of the following properties: a mass is enabled to move via gravity (spill, drip, spill), a flexible object extended in one direction is put around another object (coil, spin, twist, wind), a mass is expelled from inside an entity (emit, expectorate, vomit). As can be seen here, the properties at stake are very precise and their identification is not trivial, especially for verbs which can be used in a variety of utterances, with some slight meaning variations.

In general, alternations are described at a global level, and each verb is associated with the alternations it undergoes. Preposition changes are thus specified at this level. A priori, there is no specific information encoded in the preposition lexical entries.

3. Polysemy and sense restrictions

In this section we briefly evoke the problem of polysemy, crucial for prepositions, and the difficulty of characterizing sense boundaries, in particular by means of selectional restrictions. Representation issues are presented in the next section, after these preliminaries.

3.1 Prepositions and polysemy

It is well known that prepositions are highly polysemic and enter into a large number of metonymies and metaphors. However, some prepositions have very restricted uses such as *west of*, *in order to*, *during*, *in spite of*, *in favor of*, *except*, *thanks to*, *concerning*, *via*, etc. We believe that it should be possible to identify a reasonable number of ‘kernel’ senses for each preposition, that accommodate several forms of variations.

The identification of a preposition sense needs to be based on the observation of groups of usages. Two criteria must be taken into account: (a) the nature and the stability within a certain semantic domain related to the head noun type of the PP controlled by the preposition, that confirms the ontological basis of the sense and, concomitantly, (b) the restrictions required by the verb on the

nature of the PP, if it is an argument. Dictionary definitions and multilingual considerations may also help. Pragmatic factors may also interfere, but this is much more ad'hoc.

Although prepositions have some idiosyncratic usages (probably much less in French than in English), most senses are relatively generic and it should be possible to characterize them using relatively consensual and high-level ontology labels.

Let us consider the case of the French preposition *par*. We have identified six senses which can be identified and characterized as shown below. These senses occur in very diverse ontological domains while being all approximately at the same level of abstraction:

- proportion or distribution: *il gagne 1500 Euros par mois* (he earns 1500 Euros per month),
- causality: as in passives but also e.g. in *par mauvais temps, je ne ne sors pas* (by (=in) 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 (=by) this path),
- tool or means: *je voyage par le train* (I travel by train),
- approximation of a value: *nous marchons par 3500m d'altitude* (we hike at an altitude of 3500m).

An important point is that uses of *par* do not necessarily cover all the conceptual field which could 'naturally' be associated with each sense. For example, the expression of the idea of approximation using *par* is rather restricted to localization, speed or movement, it does not include e.g. amounts. One of the tasks is then to characterize, for each sense, what the subset of the conceptual field is. This is done by two means: (1) by a semantic characterization of the NP dominated by the preposition and (2) by the analysis of the restrictions imposed by the verb of the clause on the PP, or, conversely, by the type or the family of the verb (e.g. possession, communication, as in WordNet) the preposition can be combined with, for that particular sense.

Let us now examine the basic restrictions for three senses of *par*. The 'VIA' sense is basically subcategorized by movement verbs; it is 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. *Il passe par des moments difficiles* (He experiences difficult moments)).

The 'ORIGIN' (or 'SOURCE') sense is more narrow, it is essentially used in conjunction with communication or epistemic verbs, the argument is usually

of type place, and the head noun is of type ‘human’: *Il transite par Paris* (he commutes in Paris). We consider that nouns of type e.g. ‘object with an informational content’ or ‘human’ introduce a metonymic extension, as in, e.g. *par la radio / la presse / des amis* (I know the news from the radio / the newspapers / friends).

Finally, the ‘TOOLS or MEANS’ sense is used with verbs describing concrete actions (e.g. creation and movement verbs, if we refer to the verb class system of WordNet (Fellbaum 93)). In general it is an adjunct. It is typed as a means, and the head noun of the PP must be e.g. a tool, or, more generally, an object that allows the action to be realized, which may not necessarily be prototypical. This object could be found e.g. in the encyclopedic knowledge associated with the verb, or via a functional relation in a thesaurus. It has also numerous metaphoric extensions (e.g. *je traite ce phénomène par la logique temporelle* (I deal with this phenomena ‘by’ temporal logic)).

3.2 Some difficulties with selectional restrictions

However, there are many well-known difficulties inherent to the selectional restriction approach, where additional, non-trivial, world knowledge is required to make sense distinctions. Consider the usage:

‘Dans (in) followed by an NP of type location’ (e.g. to be in a drawer).

Location is obviously too general a restriction (**to be in the shelf*). It is then necessary to enter into more complex descriptions, specifying that the location has a (salient) ‘inside’, that is not just a surface, etc. However, as far as only elementary spatial properties are concerned, this remains feasible.

More complex is the case of *boire dans un verre* (literally: drink in a glass). This example highlights the complex interactions between the verb and its PP. The preposition is part of the PP, not part of a verb complex form, this latter construction being quite unusual in French. The recipient is not neutral: while *verre, tasse, bol,...* (glass, cup, bowl) are acceptable arguments, *bouteille, robinet* (bottle, faucet) are not, probably because of their narrow neck, which prevents the drinker from having his mouth inside the recipient. This characterization becomes more complex and, probably, an interpretation for example in terms of Euclidean geometry could be necessary.

4. Representing the semantics of prepositions

A few general purpose classifications for prepositions have been proposed in the past. They tend, in most cases, to converge quite well. In this section, we survey two of them. The first was introduced in the eighties by (Boguraev and Spark Jones 87), while the latter serves as a basis for the PrepNet project (Saint-Dizier 05). Another classification, based on a lexicographic method-

ology, is presented at <http://www.clres.com/prepositions.html> in the Preposition Project (TPP).

In the remainder of this section, we focus on representation and expressivity issues. Basic underspecification techniques are introduced to show the verb-PP interactions in semantic composition.

4.1 A study of cases

Let us present first Boguraev and Sparck Jones classification. It is based on (Woods 79) which provides an extensive list of preposition uses in English. Their study was a purely investigative one, with the aim of characterizing sentence relations. They did not address the question of how the specific assignments, for each individual sentence, could be achieved. We give below the main elements of the list, whis is given in (Boguraev and Spark Jones 87), and accessible via the ACL digital library. Cases are not structured. For each of them, we give the prototypical English prepositions. They are given below in alphabetic order:

Accompaniment (with),
 Activity (at),
 Abstract destination (to),
 After (after),
 Abstract location (in),
 Abstract source (from),
 Attribute (in, with),
 Before (before),
 Comparison (as),
 Destination (to),
 Direction (down, ...),
 Goal (for),
 Instrument (by, with),
 Location (at),
 Manner (with),
 Reason (because of),
 Source (from),
 Time location (at).

4.2 The PrepNet classification

Here is an organization of the different senses for prepositions as implemented in PrepNet, which is still in an early stage of developement (accessible at: www.irit.fr/recherches/ILPL/prepnet.html), with some frequent minor adjustments. Senses are called abstract notions, to dissociate them from linguistic realizations. The classification was initially elaborated from French (Cannesson

et al 02), but seems largely valid for most European languages. It also coincides to a large extent with other classifications, presented in some chapters of this volume.

Senses are organized on three levels:

- 1 a first level characterizes a **semantic family**, 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 developed in this document.

It is also important to note that each preposition sense is considered from the point of view of its basic usage and as the source of numerous metaphors. For example, origin is basically spatial, but has numerous metaphorical transpositions into the temporal, psychological and epistemic domains, to cite just a few generic cases.

Here is the current PrepNet preposition classification, one or more examples follow to illustrate definitions, which cannot be given here in extenso due to space limitations:

■ **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 referencial 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,*

- **Accompagnement** 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.*

Each of the facets described above is associated with a number of preposition lexicalizations. Here is a brief description of the Ordering family, with its 2 subsequent levels:

Fig. 2 - prepositions of the Ordering family		
facet	modality	preposition sense of
Priority	before	before / avant
	after	after / après
Subordination	under	under / sous
	above	on / sur
Hierarchy	under	before / derrière, avant
	above	front of, after / devant, après
Ranking	before	before, ahead of / devant
	after	after / derrière
Degree of importance	proximity	near, close to / à côté de, auprès de,
	comparison	par rapport à, for, with respect to / pour, vis-à-vis de

4.3 Semantic representation and underspecification

Each preposition sense can be associated with a semantic representation, often largely underspecified. Let us consider in this chapter a simple illustration that shows some methodological elements and some basic difficulties, which will be deepened in the next chapters.

4.3.1 Representing preposition senses. Senses are described at two levels: (1) by means of a thematic grid characterizing the ‘standard’ function of each argument as presented in section 2 and, mainly (2) by means of a knowledge representation formalism, for example the Lexical Conceptual Structure (LCS) (Jackendoff 90, 97), which seems to be sufficiently expressive for that purpose. Compared to verbs, representing prepositions in LCS is rather straightforward and much more adequate. The difficulty is to elaborate a minimal, but sufficiently discriminatory set of primitives (55 in (Wierzbicka 92) system, 68 in (Cannesson et al. 02)). Y. Wilks introduces in (Wilks 77) the main arguments for and against the use of primitives, a long, recurring debate during the 70-80s.

A few principles guide this description: (1) the representation of generic senses (e.g. family level) subsumes the representation of their daughters, (2) different senses of a given preposition must receive substantially different semantic representations, (3) metaphoric uses are characterized in part by semantic field substitution in the LCS, not by a different representation with different primitives, and (4) the number of primitives representing prepositions must be as limited as possible. These primitives are lower in the LCS primitive hierarchy than e.g. the GO, CAUSE or BE primitives.

An important feature of the semantic representation of prepositions is the evaluation of an adequate level of genericity, that includes a number of variations related to the semantics of the preposition arguments. A possible solution consists in associating LCS representations with: