Rolf Kailuweit, Matthias Hartung, Eva Staudinger[1]

University of Heidelberg

Linking Syntax to Semantics: Multivariate Statistical PP-Attachment Disambiguation

Abstract

One of the fundamental problems of the RRG Linking Algorithm from syntax to semantics is that the selection of the appropriate syntactic template for an input sentence is often subjected to several ambiguities. In this paper, we examine PP attachment ambiguities focusing on the French verb *parler*. Within the first section, we deliver an analysis of the possible syntactic templates occurring with *parler*. Based upon a quantitative evaluation of *parler* sentences, we then propose a statistically-driven algorithm for the distinction between core arguments and adjuncts which results in the selection of the most likely template.

Introduction

In the last 15 years, the relation of the syntactic and semantic structure of an utterance (or: the relation of syntactic and semantic structures), i.e. the linking problem, has dominated linguistic discussion. Butt & Holloway (2000: 1) pointed out that "argument realization – how arguments of predicates surface in the clause – is central to linguistic theory." As many other linguists, Butt & Holloway adopt the semantics-to-syntax approach. But, obviously, argument realization is only half of the problem. The exact reverse, i.e. the syntax-to-semantics approach, has been widely neglected in linguistic theory. In comparison with other approaches, one of the advantages of RRG lies in accounting for both sides of the linking problem. Van Valin & LaPolla (1997) provide not only a semantics-to-syntax but also a syntax-to-semantics linking algorithm.

In a recent paper, Van Valin (2003) discusses the contribution of RRG to language processing as it is accounted for by psycholinguistic and computational modeling. Van Valin (ibid.) emphasizes that, from a processing point of view, general principles governing macrorole assignment and general rules which assign the prepositions that mark oblique core arguments should not be part of the linking algorithm itself but that they should be anticipated in a precompiling step at the lexical level. Hence the precompiled logical structures (LS) should contain information on macrorole and preposition assignment. Likewise syntactic templates should be enhanced with macrorole and preposition assignment information. Parsing would then consist in selecting an appropriate template for the input by statistical means. Linking would be reduced to a single step: matching the information on the

1 {rolf.kailuweit, hartung, eva.staudinger}@urz.uni-heidelberg.de

appropriate template to the information on the logical structure, yielding "a very fast and efficient comprehension process."(Van Valin ibid.).

However, this very optimistic proposal faces a lot of empirical problems when we put the processing algorithm to the text. A closer look at the procedure reveals that for a computational implemantation a detailed specification is required. One of the most fundamental problems is the resolution of the syntactic dependencies of the constituents with regard to the distinction between core arguments and adjuncts. The problem is usually referred to with the label PP-attachment ambiguities. Since, at least in languages such as English or French, case-marking prepositions and predicative prepositions cannot be distinguished on morhosyntactic grounds, every PP that does not correspond to an obligatory argument is ambiguous. If the PP fills an optional argument slot of the predicate, the PP must be attached to the core. If it is an adjunct, it has to be attached to the periphery. In addition, if the PP follows a NP, the preceding NP (noun attachment) may function as the syntactic head.

The aim of our paper is to provide an algorithm that is capable of predicting the correct semantic representation for any occurrence of French *parler* ('to talk', 'to speak'). As English *talk* or *speak*, *parler* takes two PP-Arguments. The argument introduced by the dative marking preposition *à* or by the preposition *avec* denotes the ADDRESSEE, the argument introduced by the preposition *de* denotes the TOPIC OF CONVERSATION. Both arguments are optional:

(1) parler à/avec quelqu'un de quelque chose [à/avec PP = ADDRESSEE; de PP = TOPIC OF CONVERSATION]

('to talk/speak to/with somebody about something')

The addressee argument can also be expressed by the preposition *avec* ('with').

In another reading *parler* takes a direct object denoting a language:

(2) parler un langage à/avec quelqu'un

('to talk/speak a language to/with somebody')

Since *de* and *à* are the most current prepositions introducing several kinds of adjuncts and NP-modifiers, *de*-PPs and *à*-PPs co-occurring with *parler* are highly ambiguous.

For example, in:

(3) J'étais meilleur orateur que lui, quand on allait parler aux ouvriers à la sortie des usines (FRANTEXT: DUVIGNAUD, J.)

('I was a better speaker than him, when we went to talk to the workers at the exit of the factories')

des usines is a modifier of the NP *sortie*, but, leaving our world knowledge aside, it could also be the *de*-argument of *parler* or even an adjunct denoting the place where the speaker talks from. In addition, there are two candidates for the ADDRESSEE: *aux ouvriers* and *à la sortie*. The non-ADDRESSEE *à*-complement could be an adjunct denoting the location where the talking takes place, but if *à la sortie* is not the addressee, it could either be an adjunct at the clause level or a modifier of the preceding NP *ouvriers*.

Syntactic Templates

If the syntax-to-semantics linking algorithm consists in matching the information on the appropriate template to the information on the logical structure, we first have to consider the different syntactic templates that could appear with *parler*.

Due to the fact that both oblique arguments of *parler* are optional, the verb appears in a one-argument template:[2]

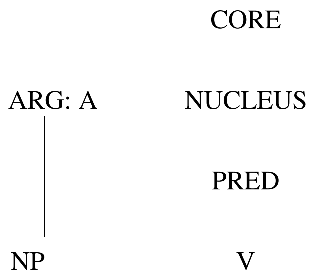


Fig. 1: One-place construction

(4) Je parle et je me tais (FRANTEXT : CLAUDEL, P.)
('I speak and I am silent')

If there is a *de*-PP, a two-arguments template might be the appropriate one:

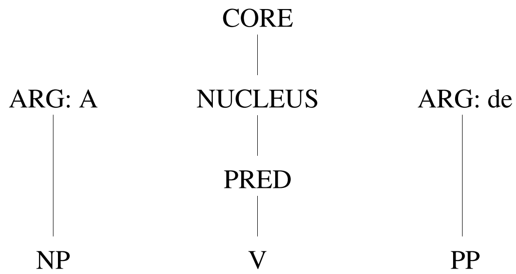


Fig. 2: Two-place construction with *de*-ARG

(5) Je parle de vous (FRANTEXT: FEBVRE, L.)
('I talk about you')

An *à*-PP could also fill a slot of a two-arguments template:

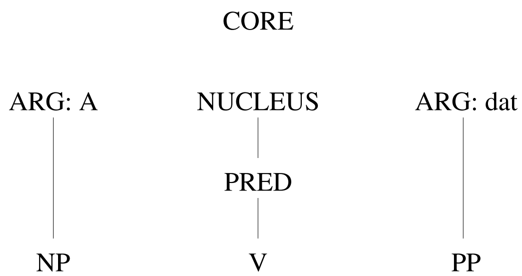


Fig. 3: Two-place construction with *à*-ARG

(6) Je parlai aux prêtres (FRANTEXT: YOURCENAR, M.)
('I talked to the priests')

Instead of the à-PP, an *avec*-PP could be the candidate for one of the arguments:

[2] Following Van Valin' slead (2003), we provide syntactic templates with macrorole and preposition assignment information.

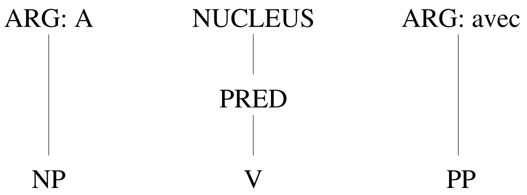


Fig. 4: Two-place construction with *avec*-ARG

(7) Je parlais avec elle (FRANTEXT: GIDE, A.)
(‚I was talking with her')

If a *de*-PP and an *à*-PP appear in the same sentence, a three-arguments template could be the appropriate one:

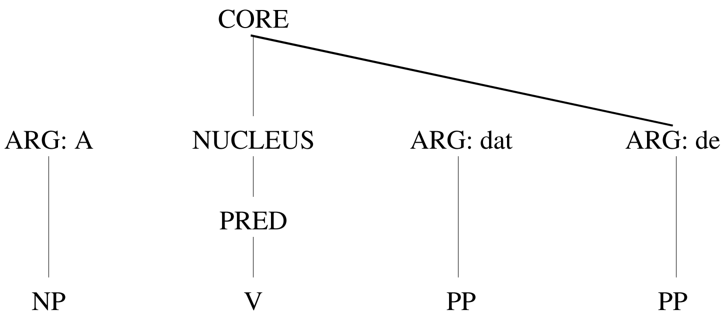


Fig. 5: Three-place construction with *de*-ARG and *à*-ARG

(8) J'a i parlé à ma mère de notre nouvelle vie (FRANTEXT : MOTHERLAND, H.)
('I have talked to my mother about our new life')

But a *de*-PP can also be part of the periphery:

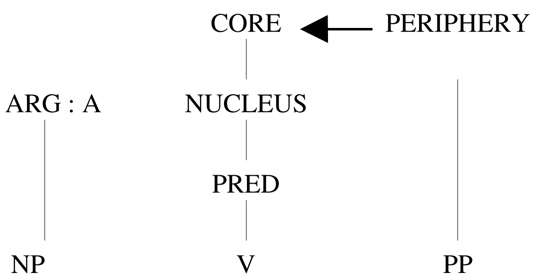


Fig. 6: One-place construction with *de*-ADJ

(9) Il parlait d'une voix ferme (FRANTEXT: GIBEAU, Y.)
('He spoke with a firm voice')

And the same applies to an *à*-PP:

(10) Il était obligé d'en parler au futur (FRANTEXT: CAMUS, A.)
('He was obliged to talk about this in the future')

If a wh-element appears, a precore-slot template is activated. According to the type of the wh-argument, the precore-slot will be marked with the appropriate macrorole and preposition assignment values and the corresponding reduced core template will be chosen:

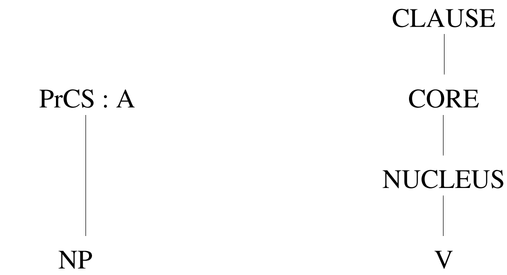


Fig. 7: Precore-slot-construction with actor-wh-element

(11) Qui parle? (FRANTEXT: GRACQ, J.)
('Who is speaking?')

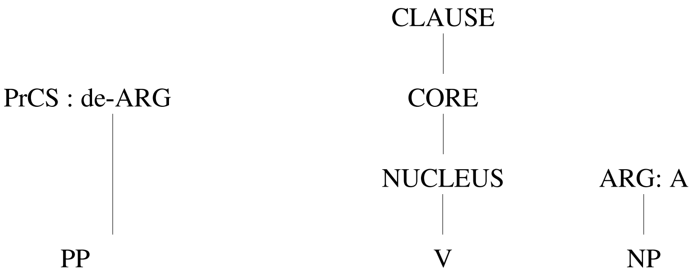


Fig. 8: Precore-slot-construction with de-ARG-wh-element

(12) De quoi parlez-vous (FRANTEXT: CLAUDEL, P.)
('What are you talking about')

Relative pronouns as *dont*, *de quoi* or *à qui* corresponding to the *de*-PP and the *à*-PP can be treated in the same way:

(13) Voilà de quoi je parle (FRANTEXT: PRÉVERT, J.)
('This is what I am talking about')

(14) La vie dont il parlait avec crainte (FRANTEXT: CAMUS, A.)
('The life about which he talked with fear')

(15) Quelqu'un à qui vous pouviez parler (FRANTEXT: MONTHERLANT, H.)
('Somebody you could talk to')

The appearance of a preverbal dative clitic activates the templates containing an *à*-PP, the appearance of the preverbal *en* activates the templates containing a *de*-PP. If both kind of clitics appear, the three-arguments template is activated.

(16) Tu m'en parles si peu dans tes lettres (FRANTEXT: GIBEAU, Y.)
('You talk to me about this so seldom in your letters')

The order of the arguments in the template does not have to be the same as the order in the text. Clitic constructions and relative clauses are not the only examples of a lack of correspondence between the respective orders. Subjects can appear in postverbal position or peripheral material can be inserted between the verb and its arguments:

(17) Ainsi parlent ces poètes (FRANTEXT : ELUARD, P.)
('Thus speak these poets')

(18) J'ai parlé tout à l'heure de fièvre et de maladie (GRACQ, J.)

(‘A few minutes ago I talked about fever and illness’)

This overview of the different syntactic constructions of *parler* is not even exhaustive. Choosing the right template is difficult. As we will see, the choice has to be based on statistic facts, e.g. the probability of each possible template to appear in given text and the probability of the constituent to be an argument or adjunct depending on the nature of its inherent semantic properties.

Logical structure of *parler* and lexical entries

Once we have chosen the right template, the information on the template should be matched with the information on the logical structure of *parler*. However, the LS for *parler* is far from being evident (why not simply:...is not obvious). Van Valin & LaPolla (1997: 116-118) propose the following general LS for verbs meaning "to say:"

(19) **do’**(x, [**express**.( ).**to**.( ).**in.language**.( )’(x, y)])

The internal variables , and refer to the content of the utterance ( ),[3] the addressee ( ) and the language used ( ). Verbs meaning "to say" are described as two-place activities with an optional second argument. The three variables are candidates for the optional y-argument. However, French *parler* as well as English *talk* or *speak* are three-place verbs.[4] The internal and variables can be realized in the same construction. It is easy to prove by the *do-so* test that none of them can be considered an adjunct:

(20) a. J’ai parlé à ma mère de notre nouvelle vie (FRANTEXT : MOTHERLAND, H.)

(‘I have talked to my mother about our new life’)

b. *J’ai parlé à ma mère et je l’ai fait de notre nouvelle vie

(lit.: ‘I have talked to my mother and I did so about our new life’)

c. * J’ai parlé de notre nouvelle vie et je l’ai fait à ma mère

(lit.: ‘I have talked about our new life and I did so to my mother’)

Van Valin & LaPolla (1997: 118) propose a more complex LS for the three-place predicate *to tell*:[5]

(21) [**do’**(x, [**express.**( )**.to.**( )**.in.language.**( )’(x, y)])] CAUSE [BECOME **aware.of’** (y, z)], where y = , z =

Tell differs from *parler* by virtue of being a causative accomplishment. On the contrary, three-place *parler* still is an activity:

---

[3] Van Valin / LaPolla (1997: 118) considered the *about*-argument of *talk* and *speak* an instance of the - variable. González Orta (2004) distinguishes between the content ( ) and the the topic ( ): [**express.**( )**.about**( )**.to.**( )**.in.language.**( )’(x, y)]. While *say* realizes a content, *speak* or *talk* realize a topic. This finer-grained distinction might be appropiate from a semantic point of view. As we will see, it does not help us to solve the linking problem.

[4] This is not taken into consideration by González Orta (2004). She suggests that the **.about.**( ) component in her LS is realized by an argument-adjunct w. But the external variable w does not show up anywhere in the LS she proposes for Old English *secgan* (‚talk’) and *specan* (‚speak’). Argument adjuncts realize a non-macrorole argument of a given verb by means of a predicative preposition. An argument adjunct is not a device for adding an argument to a verb.

[5] By the way, English *tell* is a dative-alternation verb allowing marked undergoer choice. The LS proposed by Van Valin / LaPolla considers the TOPIC OF CONVERSATION as undergoer construction (*tell sth to sb*) the unmarked construction. French *raconter* (‘tell’) supports this analysis, with the TOPIC OF CONVERSATION argument being the only choice for undergoer.

(22) a. *J'a i parlé à ma mère de notre nouvelle vie en 20 minutes
(lit.: 'I have talked to my mother about our new life in 20 minutes')

According to Van Valin (2002), all three-place verbs are causative. Proving that causativity is not restricted to accomplishments, is one of RRG's major contributions to aktionsart classification.[6] There are causative activities such as *rouler*:

(23) Sisyphe roule sa pierre
('Sisiphus roles his stone')
[**do'** (Sisyphus, Ø)] CAUSE [**do'** (stone, [**role'**(stone)])]

Hence, *parler* could have a LS like the following one:

(24) [**do'**(x, [**express.( ).to.( ).in.language.( )**'(x, y)])] CAUSE [**do'**(z, [**listen'**(z)])], where y = , z =

But there are several arguments against such a representation. First, in the case of *rouler*, the verb describes a caused activity of the moving object, while *parler* does not describe a caused activity of the ADDRESSEE. Second, we may safely assume that macrorole intransitive causative verbs don't exist. Causing an activity (or change of state) is a strong agent feature resulting in a high degree of semantic transitivity that should correspond to syntactic macrorole transitivity.

Therefore we claim that the three-place reading of *parler* should be described as a non-causative activity having the following LS:

(25) **do'**(x, [**express.( ).to.( ).in.language.( )**'(x, y, z)]), where y = , z = [MR1]

Three-place activities have not yet been described in RRG's semantic formalism.[7] In our opinion they do not disturb the system in a considerable way. The three arguments correspond to three different degrees of activity in the Actor-Undergoer-Hierarchy:

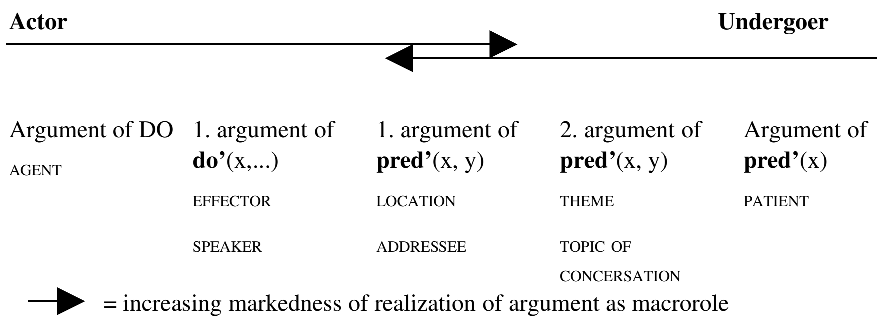


Fig. 9: Actor-Undergoer-Hierarchy (cf. Van Valin / LaPolla 1997: 127; 146)

The ADDRESSEE expressed in French by a dative is located in the center of the hierarchy. The same position is attributed to the dative RECEIVER of *donner* ('give') or the dative LEARNER of *enseigner* ('teach'). This seems to be appropriate from a semantic point of view. Of course, a three-place activity predicate does not embed a two-place **pred'**(x, y) sequence in its logical structure. Hence, the second and the third position of the Actor-Undergoer-Hierarchy have to

[6] Van Valin & LaPolla (1997: 102-109); Kailuweit (2003)

[7] The number of non-causative three-place verbs in French is not easy to determine. For many verbs, the argument status of one PP might be doubtful. Apart from *parler*, there is at least one more three-place example: *en vouloir à quelqu'un de quelque chose* ('to be angry with somebody about something').

be redefined respectively as the second rightmost argument of **pred'**(…) and rightmost argument of **pred'**(…).

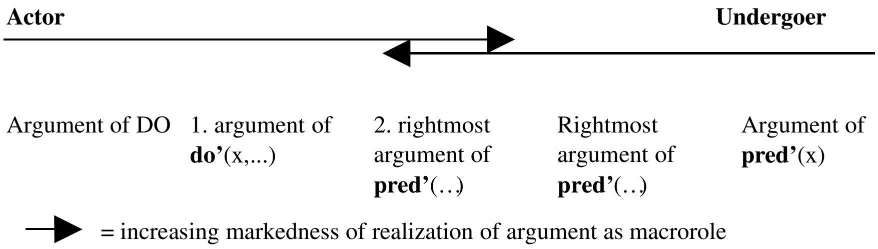


Fig. 10: Actor-Undergoer-Hierarchy redefined

Following Van Valin (2003), our model macrorole and preposition assignment will take place at the lexical level. The actor macrorole is assigned to the x-argument according to the following general rule: activities take an Actor macrorole as leftmost argument of their LS. The oblique case and preposition assignment might follow general rules, too. The dative is the default case for non-macrorole second rightmost arguments of three-place predicates in French. In addition, the preposition *de* seems to be the default marking of non-macrorole rightmost arguments. But this is mere speculation at this stage. Whether by the application of general rules or by idiosyncratic marking, the resulting precompiled LS for the two alternatives of *parler* are the following (round brackets indicate that the y-argument and the z-argument are optional):

(26) a. **do'**(x, [**express.**( )**.to.**( )**.in.language.**( )'(x = A, (y = dat/avec), (z = de))]), where y = , z =

b. **do'**(x, [**express.**( )**.to.**( )**.in.language.**( )'(x = A, (y = dat/avec), (z = U))]), where y = , z =

Currently, lexical entries in RRG only consist of the lemma and its LS. For nouns, the qualia structure (cf. Pustejovsky 1995) is supposed to be added (cf. Van Valin / LaPolla 1997: 184-186). Recently (cf. Van Valin forthcoming: Chap. 2), it has been argued that an argument position of a predicate can be annotated with a qualia type if the predicate requires this particular type of argument. Implementation additionally requires a whole range of morphosyntactic information that should be annotated in terms of features and attribute value matrices such that the technique of unification can be used. First of all, a lexical entry has to be labeled with a part of speech value.

While Van Valin (forthcoming: Chap. 2) only refers to qualia, Pustejovsky (1995) distinguishes two kinds of semantic information: argument selection specifications and qualia. In our context, qualia are not important. In order to distinguish two of the possible arguments of *parler,* we will use argument selection specifications in the sense of Pustejovsky (1995: 67). The ADDRESSEE has the specification x = **animate_individual** including the specification x = **human**. The LANGUAGE has the specification x = **language**. This yields the following entry for *parler*:

$$\begin{bmatrix} \text{parler} & \\ \text{POS} = \text{V} & \\ \text{LS} = \text{do}(x,[\text{express}.(\alpha).\text{to}.(\beta).\text{in}.\text{language}.(\gamma)'(x,y,z)] & \\ \text{ARGSTR} & \begin{bmatrix} x\text{ - ARG}\left[\text{SYN} = \text{A}\right] \\ \left(y - \text{ARG}\begin{bmatrix} \text{SYN} = \text{dat / avec} \\ \text{SEM}\begin{bmatrix} = \beta \\ = \text{animate\_in d} \end{bmatrix} \end{bmatrix} \right) \\ \left(z - \text{ARG}\begin{bmatrix} \text{SYN} = \text{de} \\ \text{SEM}[= \alpha] \end{bmatrix} \right) \\ \text{vel} \\ x\text{ - ARG}\left[\text{SYN} = \text{A}\right] \\ \left(y - \text{ARG}\begin{bmatrix} \text{SYN} = \text{dat} \\ \text{SEM}\begin{bmatrix} = \beta \\ = \text{animate\_in d} \end{bmatrix} \end{bmatrix} \right) \\ \left(z - \text{ARG}\begin{bmatrix} \text{SYN} = \text{U} \\ \text{SEM}\begin{bmatrix} = \gamma \\ = \text{language} \end{bmatrix} \end{bmatrix} \right) \end{bmatrix} \end{bmatrix}$$

Fig. 11: Lexical entry for *parler*

Lexical entries for nouns could be exemplified in the following way:

$$\begin{bmatrix} \text{mère} \\ \text{PSO} = \text{N} \\ \text{ARGSTR}\left[\text{ARG1} = \text{human}\right] \\ \text{QUALIA}[...] \end{bmatrix}$$

Fig. 12: Lexical entry for *mère*

$$\begin{bmatrix} \text{allemand} \\ \text{PSO} = \text{N} \\ \text{ARGSTR}\left[\text{ARG1} = \text{language}\right] \\ \text{QUALIA}[...] \end{bmatrix}$$

Fig. 13: Lexical entry for *allemand*

After having specified these prerequisites, we will now deal with the central question of this paper: how to find the appropriate syntactic template by means of a multivariate probability-driven device.

Previous Work – Why a multivariate approach ?

Much previous work in the field of PP-attachment resolution is restricted to one-sided syntactically or semantically motivated approaches (cf. Ratnaparkhi 1998 for a heuristic approach based on purely syntactic features; Volk 2002, Stetina & Nagao 1997, Hindle & Rooth 1993 for purely semantic criteria). However, empirical data suggest that these problems are best accounted for by integrated models combining syntactic and semantic criteria (Franz

1996: 30ff.), which is underlined by the following examples:

(27) Je parlais [de la mort]$_{ARG}$ (FRANTEXT: FEBVRE, L.)
('I spoke about death.')

(28) Il parlait [d'une petite voix aiguë]$_{ADJ}$ (FRANTEXT: GIBEAU, Y.)
('He spoke with a tiny shrill voice')

Examples and show that neither semantic nor syntactic features alone suffice to resolve PP-attachment ambiguities. As Franz (1996: 32) pointed out, "these principles in isolation do not constitute an empirically adequate theory". Although having the same syntactic construction on the surface, RRG's syntactic trees of these sentences would differ from each other. The weakness of the majority of the semantic approaches put forward up to now lies in their restriction to the computation of the governing node of the respective PP from statistical values. In the case of Verb Attachment, the final decision whether the PP is an argument or an adjunct is omitted. However, if we aim at modeling "true" sentence processing, the analysis should include this final step.

Interactionist models, as they can be found in psycholinguistic research contributions, also provide a strong argument for multi-dimensional approaches to human language processing. They suggest that sentence understanding relies on the parallel processing of probabilistic syntactic and semantic evidence (cf. Jurafsky 1996, Jurafsky & Martin 2000: 471).

Our aim is therefore to develop a probabilistic model for the resolution of PP-attachment ambiguities for the French verb *parler* that integrates several information sources. This model should be cognitively adequate and compatible with the RRG syntax-to-semantics linking algorithm (cf. Van Valin 2003), i.e. it should enable the selection of the appropriate core template for any occurrence of *parler*.

Training Corpus

The training material has been extracted from the FRANTEXT corpus, a compilation of 631 untagged texts, published between 1951 and 2000. Our statistics is based on the first 1000 instances of *parler* (1951-1952). This is certainly not a representative sample; the preparation of the training material, however, is time-consuming. The training corpus has been annotated manually. We only tagged at the constituent-level, i.e. we tagged arguments and adjuncts of *parler,* but not their internal structure. In the case of PPs headed by *à*, *avec* or *de* we also annotated the following:

- <+hum> if the NP contained referred to a person
- <+an> if it referred to something animate
- <+abs> for abstract entities
- <+loc> for locations
- <adv> to mark adverbial expressions
- <prep> to mark prepositional expressions

The <adv> and <prep> tags are used to generate lexical entries for these expressions, so we can identify them as single units during the morpho-syntactical tagging step. In order to reduce the processing load of the parser, adverbial and prepositional expressions headed by *à* and *de*, e.g. *d'égal à égal,* are filtered prior to parsing the PPs, as they obviously have adjunct status. The information extracted from the annotated training corpus forms the basis for the decision whether a PP attached to a verb is likely to be an argument or an adjunct of *parler*. How exactly this decision is taken and how we make use of the training data will be described in more detail as we go on explaining the features of the statistical processing.

The Features of our Model

The multivariate disambiguation model relies on the following features:

Syntagmatic position of the PP (Pos). The hypothesis of "minimal attachment", i.e. PP attachment to the immediately preceding XP, is frequently put forward (cf. Gaussier & Cancedda 2001, Franz 1996: 23). In order to examine the impact of syntagmatic relations on linking, we take into consideration the position of the PP relative to the position of the verb.

Preposition heading the PP (Prep). This feature can be regarded as a probabilistic account of subcategorization frames. The underlying assumption is that specific prepositions indicate the presence of an argument (cf. Van Valin 2003: 18). In the case of *parler*, *à* frequently heads the argument realizing the ADDRESSEE role, whereas *de* tends to mark the TOPIC OF CONVERSATION.

Statistical concordance measures (NounFit, VerbFit). Note, however, that due to their highly ambigious character not every occurrence of a specific preposition can be seen as an argument marker. Generally speaking, the linking problem is intermingled with a second issue concerning the internal structure of NPs. For complex constructions, there exist at least three different attachment possibilities: attachment to the most recent NP, attachment to a higher level NP[8] or attachment to the verb (Franz 1996: 30). The various alternatives are exemplified, in the given order, in -:

(29) Les journaux ne parlaient pas de la couleur de son costume (FRANTEXT: GIDE, A.)
('The newspapers did not talk about the color of her suit')

(30) Nous parlâmes encore de l'abus de la couleur en littérature (FRANTEXT: SAINT-JOHN PERSE)
('We also talked about the abuse of color in literature')

(31) J'ai parlé à ma mère de notre nouvelle vie (FRANTEXT: MONTHERLANT, H. de)
('I have talked to my mother about our new life')

In the case of verb attachment, complete linking must further distinguish verb arguments from adjuncts. Before we can solve the linking problem, we must determine the boundaries of the XPs involved. This decision is based on the comparison of the statistical cooccurence values:

$$VerbFit(V, PP) = \frac{freq(V, PP)}{freq(V)}$$

$$NounFit(N, PP) = \frac{freq(N, PP)}{freq(N)}$$

Note that, following Volk's example (2001), we consider the whole of the PP in order to compute these cooccurrence values. This can be understood as an expansion of the method used by Hindle & Rooth (1993), whose approach was restricted to the head-preposition.
The underlying hypothesis is that the higher the compatibility of their lexical properties, the higher the probability that the phrases in question constitute a complex XP. In other words: The decision whether a PP is part of one of the preceding XPs or should be attached to the main verb instead should be in favour of that alternative with the higher lexical compatibility. Since a sufficient amount of lexical information which is based on an appropriate ontology is

[8] We exclude cross-dependencies, i.e. higher NP attachment is considered as a valid alternative only if the immediately preceding PP has been attached to the higher NP, too. This view is supported by our corpus data.

not available for French, we rely on these co-occurrence counts instead.

Semantic class of the PP head noun (Specifier). Our corpus reveals that the semantic class of the head noun of a PP strongly affects its qualification as an argument. This view is supported by the notion of selectional preferences of verbs, which is a well-known concept stating that verbs tend to impose certain semantic constraints on their arguments (cf. Pustejovsky 1995: 66f., Manning & Schütze 1999: 288). Consequently, we are interested in extracting statistical facts about the nature and the strength of these constraints from our training material. In the case of *parler*, we found that *à*-PPs whose head nouns bear the semantic attribute +human/+animate, have a strong tendency to function as the ADDRESSEE. The TOPIC OF CONVERSATION role, however, is far less restricted in this respect.

Estimating Probabilities

The features of the model can be formally represented as a tuple $T = \langle t_1, t_2, t_3, t_4 \rangle$, where t_1 contains the whole PP, $t_2 \in$ **Pos**, $t_3 \in$ **Prep** and $t_4 \in$ **Specifier**.[9] A shallow parser operating with finite-state techniques (cf. Grefenstette 1996) recognizes NPs and PPs from our training corpus and creates an instance T_i for every PP.

(32) Il parlait [dans une longue interview]$_{PP}$ [d'u n grand nombre]$_{PP}$ [de morts]$_{PP}$ [au cours]$_{PP}$ [des dernières 24 heures]$_{PP}$. (Abeillé & al. 2001)

('He was talking in a long interview about a large number of victims during the last 24 hours')

For this example, our shallow parser produces the following set of T instances:

T = { T_1: <dans une longue interview, +1, dans, None>,
T_2: <d'un grand nombre, +2, de, None>,
T_3: <de morts, +3, de, +hum>,
T_4: <au cours, +4, à, None>,
T_5: <des dernières 24 heures, +5, de, None> }

As our training corpus is manually annotated, we can identify arguments and adjuncts in every training instance taking into account the features involved. As a result, we can compute conditional probabilities for PPs being an argument or adjunct given the evidence we obtain from T_i.

$$P(Arg \mid T_i) = \frac{P(T_i \mid Arg) \cdot P(Arg)}{P(T_i)} = \frac{P(\langle t_1, t_2, t_3, t_4 \rangle \mid Arg) \cdot P(Arg)}{P(T_i)}$$

The Bayesian approach enables us to cover potential dependencies between the features in a statistically adequate manner. Furthermore, it can be shown that this Bayesian approach optimally classifies data with respect to its minimum error rate compared to other decision methods (Fahrmeir 1984: 305f.).

In order to account for sparse data problems, we can not simply equate probabilities with observed frequencies. Thus, we rely on Laplace Smoothing, which is known as a rather simple smoothing technique equally distributing the missing probability mass on unobserved cases (cf. Gaussier & Cancedda 2001, Manning & Schütze 1999).

9 As stated above, the comparison between VerbFit and NounFit precedes the linking procedure and, moreover, follows a different logic of application. Thus, these features are not part of *T*. See the section on the algorithm for details.

Distinguishing Arguments and Adjuncts

In order to obtain the probability that a given PP is an argument or an adjunct, we generate a Finite State Transducer from the annotated training samples.

Finite State Transducers can be regarded as a variation of Finite State Automata, which, apart from accepting an input *X*, simultaneously emit an output *Y* (cf. Klabunde 1998: 72). The first step consisted in generating a data structure *I*, where for each segment of the corpus all contained PPs are described with regard to the criteria introduced above:

(33) $X = \langle x_1, x_2, x_3 \rangle$, where $x_1 \in$ **Prep**[10], $x_2 \in$ **Pos,** $x_3 \in$ **Specifier**[11].

Concerning the verb, the syntactic construction[12] is taken into account.

This information is used to annotate the transitions of the transducer. Its states are labelled by the type of arguments and adjuncts seen so far. The new target state is chosen according to the above specified criteria. When reaching a new state, the transducer emits a tuple containing information on whether the PP is an argument or an adjunct of the verb and on the respective probability, given the state of the transducer, which depends on the PPs seen so far, i.e.:

(34) $Y = \langle y1, y2 \rangle$ where $y_1 \in \{\text{ARG, ADJ}\}$, $y_2 = \text{P}(y_1 \mid \text{currentState})$

In some cases, both the "A RG" and the "ADJ" transition is possible, they just yield different results with different probabilities. Figure 14 shows a simple example for a transducer generated from two training sentences and .

(35) Je parle à l'in dicatif présent
('I speak in the present indicative')

(36) Je parle à ma mère
('I talk to my mother')

The dotted lines are transitions that were not generated from the training samples, but account for unseen instances.[13]

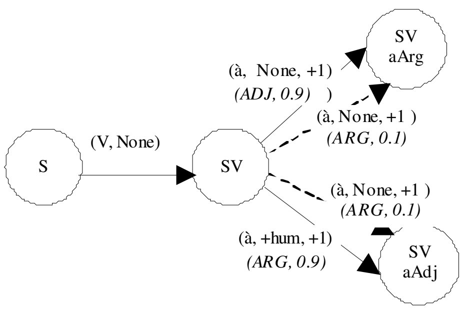


Fig.14: Transducer example

The 68 templates shown in Table 1 (in a reader-friendly form) illustrate the possible

---

10 We do not take into account prepositions that can only head adjuncts, i.e. *en, dans, sur, sous…*The instances of *en* given in the table refer to the pronoun *en* replacing a *de*-PP.

11 If no specifier can be identified for certain nouns, the respective value is *None*.

12 This is important, for example, in the case of imperative constructions, where the dative clitic, which usually appears before the verb, comes after the verb. For simple SVO sentences, the value is None.

13 For the sake of simplicity, only two such transitions have been added; in fact, there would be even more possible combinations that would yield valid transitions.

combinations and thus the terminal states that have been identified on the basis of the training data.

Template	**Frequency**	**Probability Estimations**
V de <ARG>	300	28,2%
V	257	24,2%
dat – clitic V	55	5,20%
dont V	52	5,00%
dat – clitic V de <ARG>	48	4,60%
V <ARG: lg>	45	4,30%
en V	41	3,90%
V à <ARG>	30	2,90%
dat – clitic en V	16	1,60%
V à <ADJ>	15	1,50%
<constr1:Vpour_au_nom_de>	13	1,30%
à qui V	8	0,84%
V avec <ARG>	7	0,75%
dat – clitic V à <ADJ>	6	0,65%
en V à <ADJ>	6	0,65%
V avec <ADJ>	6	0,65%
<constr3:qcVde_refl>	5	0,56%
en V à <ARG>	5	0,56%
V de <ADJ>	5	0,56%
de quoi V	4	0,47%
V de <ARG> à <ARG>	4	0,47%
c'est à <ARG> que V	3	0,37%
de <ARG> V	3	0,37%
refl dat clitic V à ARG	3	0,37%
V de <ARG> avec <ADJ>	3	0,37%
<constr2:Vcontre>	2	0,28%
à qui V de <ARG>	2	0,28%
acc –clitic:lg V	2	0,28%
c'est de <ARG> que V	2	0,28%
dat - clitic V <ARG:lg>	2	0,28%
dat - clitic V à <ARG>	2	0,28%
dat - clitic V de <ADJ>	2	0,28%
dat - clitic V de <ARG> avec <ADJ>	2	0,28%
dont dat - clitic V	2	0,28%
dont V à <ARG>	2	0,28%
dont V avec <ADJ>	2	0,28%
est-ce que à <ARG> que V	2	0,28%
qu <ARG:lg> V	2	0,28%
Refl acc clitic V	2	0,28%
V à <ARG> de <ARG>	2	0,28%
V de <ARG> à <ADJ>	2	0,28%
V:Imp dat - clitic de <ARG>	2	0,28%

<ARG:lg> qu V	1	0,19%
à <ADJ> V	1	0,19%
à <ADJ> V de <ARG>	1	0,19%
à <ARG> V	1	0,19%
à qui V à <ADJ>	1	0,19%
avec <ADJ> V de <ARG>	1	0,19%
c'est de <ARG> que dat - clitic V	1	0,19%
dat - clitic en V à <ARG>	1	0,19%
dat - clitic V avec <ADJ>	1	0,19%
dat - clitic V de <ARG> de <ADJ>	1	0,19%
dont à <ARG> V	1	0,19%
en V avec <ADJ>	1	0,19%
en V avec <ARG>	1	0,19%
en V avec <ARG> à <ADJ>	1	0,19%
qu <ARG:lg> V à <ADJ>	1	0,19%
Refl dat clitic V	1	0,19%
V à <ADJ> à <ADJ> de <ARG>	1	0,19%
V à <ADJ> avec <ADJ>	1	0,19%
V à <ADJ> de <ARG>	1	0,19%
V à <ARG> avec <ADJ>	1	0,19%
V à <ARG> de <ADJ>	1	0,19%
V avec <ADJ> à <ARG>	1	0,19%
V avec <ADJ> de <ARG>	1	0,19%
V de <ARG> avec <ARG>	1	0,19%
V:Imp dat – clitic	1	0,19%
V:Imp en	1	0,19%

Table 1: Templates occurring with *parler*

The algorithm

Due to the probabilities gained from the training corpus, we are able to process input sentences with regard to the linking problem. These sentences have to be kept strictly separate from the training data. In our case, they were extracted from a French corpus (Abeillé et al. 2001) which is morpho-syntactically tagged. Part of the corpus is also available as a fully tagged treebank. As the markup does not reflect whether the PPs have argument or adjunct status, however, we manually added this kind of annotation to the corpus. As a result, this sample serves for evaluation purposes.

Within our algorithm, the PPs contained in the input sentence are incrementally processed according to their linear order. This seems to be the cognitively most adequate procedure, which our linking approach is supposed to take into consideration (cf. Jurafsky 1996, VanValin 2003).

The first step of the algorithm consists in the **NounFit/VerbFit** disambiguation described above. There are several reasons for the separate processing of this step: First, these criteria refer to all preceding XPs, whereas the scope of **Pos**, **Prep** and **Specifier** is limited to the current PP only. As a consequence, the computational approach to **NounFit/VerbFit** diverges from the other features. Second, there is a considerable difference in the probabilistic models as well.[14] Moreover, our training corpus is considered too small for the computation of valid cooccurrence statistics, so that we make use of the World Wide Web as an external resource here (cf. Volk 2001). Third, we believe that the distinction between **NounFit** and **VerbFit** has to be kept separate from the argument-adjunct distinction which logically requires a preceding **NounFit/VerbFit** disambiguation.

The incremental linking process can be formalized as a traversion of the transducer generated from the training data. The states of the transducer can be understood as a representation of the current status of the slot-filling process. Each transition between states is labelled with a tuple *X* = <*Prep*, *Pos*, *Specifier*> representing one PP from the input string.

As a crucial requirement, the transducer must represent all *possible* combinations of arguments and adjuncts within an input sentence. Recall that *parler* opens two argument slots (apart from the PSA-slot) for exactly one *à*-complement and exactly one *de*-complement.[15] Whenever one of these slots is filled by a PP from the input, this is a logical constraint for the following PPs.

The most probable reading of an input sentence concerning the argument or adjunct status of its PPs is thus determined by selecting the chain of transitions with the highest total probability. A complete Bayesian probability model for a sentence consisting of two PP constituents can be exemplified as:

(37) P(V-àArg-deAdj) | <de, +anim, +2>, <à, +hum, +1>)

Note that the complexity of this model generally depends on the number of PP constituents within the respective sentence: *n* PPs necessarily cause an *n*-gram model to be applied. In order to avoid sparse data problems with increasing *n*, we reduce the complete model by introducing an independence assumption:

(38) P(V-àArg-deAdj) | <de, +anim, +2>, <à, +hum, +1>) =
P(V-àArg | <à, +hum, +1>) · P(V-àArg-deAdj | <de, +anim, +2>)

As the algorithm proceeds strictly incrementally, it is reasonable to include a Probabilistic Pruning Step (PPS) similar to the one used in the Beam Search Algorithm proposed by Jurafsky (1996): All chains whose total probability at the current stage is outside the beam width are immediately discarded. The beam width itself is computed as the ratio between the best and the worst path. The pruning threshold has to be determined experimentally (cf. Gibson 1991). Within our algorithm, an alternative is pruned if the ratio between the current total probability of the next more highly ranked alternative and its own current total probability exceeds 10.

Below, our algorithm is described in pseudo-code:

```
for every PP:
        if PP.NounFit > PP.VerbFit:
        merge PP with respective constituent
        continue
        else:
        for every transition in currentState:
           if transition == X.Prep and transition is valid[16]:
                 newState = expand transition
```

[14] As mentioned above, the mathematical background of the linking procedure is Bayesian classification of feature vectors T_i, whereas the **NounFit/VerbFit** disambiguation refers to the comparison of likelihood coefficients.

[15] This knowledge is gained from the Logical Structure of the verb, which is stored in its lexical entry. Note that even if we currently concentrate on *parler*, our approach can be generalized to any verb, provided the subcategorization frame is coded in the lexical entry and there is sufficient training data.

[16] “Valid” means in this case that no argument slot is filled twice (logical constraint).

```
                    compute P(C) based upon y2
                    if P(C) lies outside of beam width:
                            discard newState                    // PPS
                    else:
                            currentState = newState
    return chain with maximum probability
```

This algorithm results in the selection of the appropriate core template from RRG's syntactic inventory according to the number of arguments on the most probable chain of states within the transducer. Now, as described by Van Valin (2003), the ultimate step of the linking procedure consists merely in matching the selected syntactic template with the Logical Structure retrieved from the lexicon. For illustration purposes, consider the following example covering the whole process from template processing to linking.

Input:

(39) Il parlait [dans une longue interview]$_{PP1}$ [d'un grand nombre]$_{PP2}$ [de morts]$_{PP3}$ [au cours]$_{PP4}$ [des dernières 24 heures]$_{PP5}$. (Abeillé & al. 2001)

The shallow parser indicates *dans une longue interview* as a PP, leading to the following transition within the transducer:

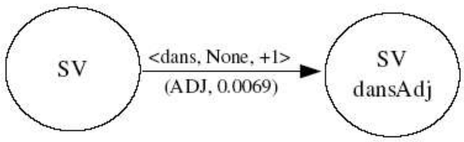
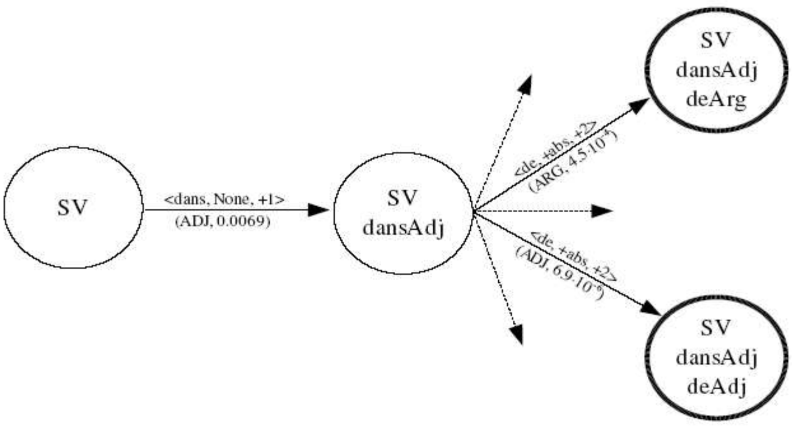


Fig. 15: State of the transducer after the transition <SV, SV-dansAdj>.

Note that the **VerbFit/NounFit** test is omitted here, as the first post-verbal PP can only be attached to the verb. For *dans*, there is only one transition, as this preposition can only head adjunct phrases. As a consequence, the Probabilistic Pruning Step is omitted, too. The parser proceeds to *d'un grand nombre* as the second PP. The result of the **VerbFit/NounFit** comparison suggests that **VerbFit** is the more likely alternative.[17] Thus, *d'un grande nombre* is regarded as an immediate constituent of its own, which leads to the next transitions described below:



[17] P(**VerbFit**)=$2.09\cdot10^{-5}$ vs. P(**NounFit**)= $8.47\cdot10^{-8}$; All results were computed according to the formula stated above, based upon data retrieved from queries of the search engine www.google.fr on June 22, 2004.

Fig. 16: State of the transducer after the transitions <SV-dansAdj, SV-dansAdj-deArg> resp. <SV-dansAdj, SV-dansAdj-deAdj>.

According to our independence assumption, the total probabilities at the current states are computed by multiplication. Up to the current state, the path leading to state SV-dansAdj-deArg is the most likely alternative. Since its probability is about 100 times higher than the probability of the dispreferred one, which clearly exceeds our chosen beam width of 10, SV-dansAdj-deAdj is immediately pruned.

The next step examines the third PP, *de morts*. Applying our theoretical thoughts from above, we have to consider two possible attachment locations for this phrase: attachment to the verb, or attachment to the preceding PP.[18] The **VerbFit/NounFit** comparison results in a strong preference for attachment to the latter alternative[19] so that no further transitions are expanded. Instead, *d'un grand nombre* and *de morts* are merged resulting in one single PP. Its probability to function as an argument or adjunct is already contained in the current state of the transducer.

The next PP, *au cours*, is a priori a candidate for an argument filling the ADDRESSEE slot. However, the **VerbFit/NounFit** test reveals that it should be attached to the preceding complex PP.[20] The same holds for the last PP in the input, *des dernières 24 heures*.[21] As a consequence, our transducer has expanded no further transitions.

Having reached the end of the input, the system returns SV-dansAdj-deArg as the most likely terminal state, revealing the following constituency structure of the input sentence:

(40) Il parlait [dans une longue interview]$_{\text{ADJ}}$ [d'un grand nombre de morts au cours des dernières 24 heures]$_{\text{ARG}}$.

The *de*-PP is identified as the only core argument present in the input. Hence, the following syntactic template is selected:

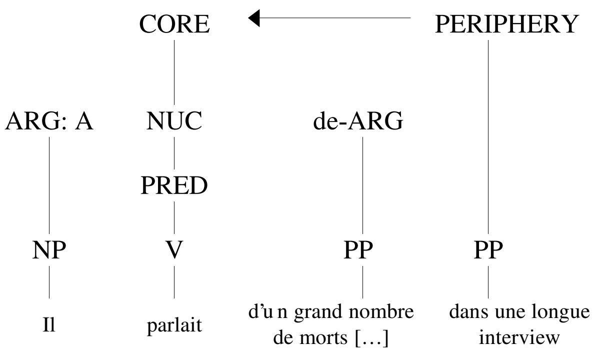


Fig. 17: Syntactic template for

After the appropriate syntactic template has been determined, it can be used as an input for the linking procedure in the sense of Van Valin (2003). We retrieve the precompiled Logical

[18] The third possibility (attachment to higher NP) has been dismissed, because no such higher NP is present here.

[19] P(**VerbFit**)=$2.46\cdot10^{-5}$ vs. P(**NounFit**)= $7.01\cdot10^{-4}$

[20] P(**VerbFit**)=$1.57\cdot10^{-4}$ vs. P(**NounFit**)= $2.07\cdot10^{-3}$

[21] P(**VerbFit**)=$6.08\cdot10^{-7}$ vs. P(**NounFit**)= $1.53\cdot10^{-4}$

Structure from the lexicon, which is repeated below for convenience:

(41) **do'**(x, [**express.**( ).**to.**( ).**in.language.**( )'(x = A, (y = dat/avec), (z = de))]), where y = , z =

The linking algorithm matches the information on the syntactic representation and on the Logical Structure: First, being the ACTOR, *il* is chosen as the x-argument. Second, as only one core argument has been detected in this case, it is linked to the *z*-slot representing the TOPIC OF CONVERSATION role. This step can be propagated via the preposition governing the PP, which is coded as subcategorization information within the lexical entry, and the thematic information about the respective slot, which can be derived from the LS. Third, after all argument-PPs from the syntactic representation have been linked to their respective slot within the LS, not all of the oblique argument slots of *parler* have been filled. Nonetheless, the remaining y-slot is optional. Therefore, there is no violation of the completeness constraint (Van Valin forthcoming) being a necessary condition for accepting the input as grammatically valid. As a last step, the remaining PP, which has been classified as a core adjunct before, is linked to the periphery of the core. This eventually leads to the following semantic representation of the input sentence:

(42) **do'**(il, [**express.**( ).**to.**( ).**in.language.**( )'(il, α= un grand nombre de morts au cours des dernières 24 heures)])

References

Abeillé, Anne; Clément, Lionel & Toussenel, François (2003): "Building a treebank for French", in: Abeillé, Anne (ed.): *Treebanks*. Dordrecht.

Butt, Miriam / Holloway King, Tracy (2000) (eds.): *Argument realization*. Stanford: CSLI.

Fahrmeir, Ludwig (1984): *Multivariate statistische Verfahren*. Berlin.

Franz, Alexander (1996): *Automatic ambiguity resolution in Natural Language Processing. An empirical approach*. Berlin.

Gaussier, Eric & Cancedda, Nicola (2001): "Probabilistic models for PP-attachment resolution and NP analysis", in: *Proceedings of CoNNL-2001*. Toulouse. 45-52.

Gibson, E. (1991): *A computational theory of human linguistic processing. Memory limitations and processing breakdown*. Pittsburgh.

González Orta, Marta Mª (2004): "Argument-Marking and Argument-Adjunct Prepositions within the Lexical Domain of Speech in Old English", in: *Atlantis* 26.1 (June 2004), 11-21.

Hindle, David & Rooth, Mats (1993): "Structural ambiguity and lexical relations", in: *Computational Linguistics* 19, 103-120.

Jurafsky, Daniel (1996): "A probabilistic model of lexical and syntactic access and disambiguation", in: *Cognitive Science* 20, 137-194.

Jurafsky, Daniel & Martin, James H. (2000): *Speech and language processing. An introduction to Natural Language Processing, Speech Recognition and Computational Linguistics.* Prentice Hall.

Kailuweit, Rolf (2003): "Classes de prédications, macrorôles et linking dans la RRG", *Cahiers du CRISCO* 14, 21-35.

Klabunde, Ralf (1998) : *Formale Grundlagen der Linguistik.* Tübingen.

Manning, Christopher D. & Schütze, Hinrich (1999): *Foundations of Statistical Natural Language Processing*. Cambridge.

Pustejovsky, James (1995): *The Generative Lexicon*. Cambridge.

Ratnaparkhi, Adwait (1998): *Unsupervised Statistical Models for Natural Language Ambiguity Resolution*. University of Pennsylvania.

Stetina, Jiri & Nagao, Makoto (1997): "Corpus based PP attachment ambiguity resolution with a semantic dictionary", in: *Proceedings of the 5th Workshop on Very Large Corpora*. 66-80.

Van Valin, Robert D. Jr. / LaPolla, Randy J. (1997): *Syntax. Structure, meaning and function.* Cambridge etc.: Cambridge University Press.
- (2002): *The Role and Reference Grammar analysis of three-place predicates.* [http://wings.buffalo.edu/linguistics/rrg/]
- (2003): *On the relationship between syntactic theory and models of language processing.* [http://wings.buffalo.edu/linguistics/rrg/]
- (forthcoming): *The syntax-semantics-pragmatics interface. An introduction to Role and Reference Grammar.*

Volk, Martin (2001): "Exploiting the WWW as a corpus to resolve PP attachment ambiguities" in: *Proceedings of Corpus Linguistics 2001*. Lancaster.
- (2002): "Combining unsupervised and supervised methods for PP attachment disambiguation", in: *Proceedings of COLING-2002*. Taipeh