

# Towards a Quantitative Theory of Variability

Philippe Blache  
LPL-CNRS, Université de Provence  
29, Avenue Robert Schuman  
13621 Aix-en-Provence, France  
pb@lpl.univ-aix.fr

## 1 Introduction

Relations between different components of linguistic analysis, such as prosody, morphology, syntax, semantics, discourse, etc. remains a problem for a systematic description (see [Blache03]). However, this is a main challenge not only from a theoretical point of view, but also for natural language processing, especially in human/machine communication systems or speech processing (e.g. synthesis). Several phenomena highlighting such relations has been described. This is typically the case for relations existing between prosody and syntax (see [Selkirk84], [DiCristo85] or [Bear90]). However, explanations are often empirical and exceptionally given in the perspective of an actual theory of language. It is for example possible to specify some relations existing between topicalization and syllable duration (cf. [Doetjes02]) or between prosodic architecture and discourse organization after focus (cf. [DiCristo99]). However, the modularity perspective, which relies on the independence of linguistic components, remains the rule in this kind of description and doesn't support a global vision of the problem.

One of the difficulties in the elaboration of a general account of this problem comes from the fact that there are only few cases of superposition between structures of the different components. It is for example difficult to precise some congruence possibilities between syntax and prosody (see [Hirst98], [Mertens01]). In the same way, and these aspects are related, the problem of variability is not taken into account in a systematic way for example in the framework of a theory. Indeed, we observe situations in which prosody can be realized in a very variable way whereas in some other cases, strong constraints have to be considered.

We think that this question of the interaction between the different linguistic components is usually addressed in the wrong (or incomplete) way. It is impossible to explain relations by means of a bijection superposing structures (for example stipulating direct relations between a syntactic tree and a prosodic hierarchy). One of the problems comes from the fact that the linguistic objects are not the same for syntax and prosody: a word can be formed with several syllables, but a syllable can in turn be formed with different words (cf. [Hirst98]).

More generally, the problem comes from the conception of linguistic information organization. It is difficult (and probably not useful) to try to represent each analysis component (1) in an homogeneous and hierarchized way by means of a total relation and (2) independently from other components. In other words, we think that each component of linguistic analysis is not necessarily fully structured: it is often difficult, or even impossible, to specify a relation between two elements of the same domain. It is for example the case of the utterance [1]

in which two parts are not connected with a precise syntactic relation but with an implicit subordination relation.

(1) *il pleut tu es mouillé* (it rains you are wet)

The same observation can be done for other components of linguistic analysis. There are for example some prosodic phenomena that are typical and recurrent in this domain (we use in this paper a simplified prosodic description limited to the notion of contour presented in [Rossi99]), but it is not necessary to represent them into a hierarchized structure covering the entire input. Generally speaking, each component participates in a partial manner to the elaboration of the informational content of an utterance. We are then far away from the classical modular conception of analysis consisting in describing this process as sequential and relying on a complete and sequential analysis of each domain (organized in level analysis, from phonetics to pragmatic). We think that the interpretation of an utterance is done thanks to pieces of information coming from any component, eventually in a redundant way. There is redundancy when congruence between components exists. But this is not the general case in which part of information can come from prosody, another one from syntax, and another from pragmatic, for example.

We propose in this paper an approach taking advantage of this conception of linguistic analysis and making it possible to describe relations between different components, not at the structure level, but directly between objects belonging to different components involved in the relation. It becomes then possible to describe relations with a variable granularity linking objects that can be at a different level from one component to the other. We can for example describe a relation between an interrogative morpheme and an intonative contour or between a phrase and some prosodic stress.

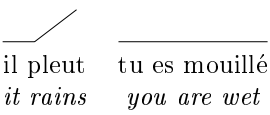
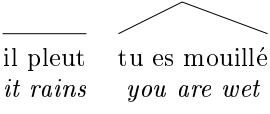
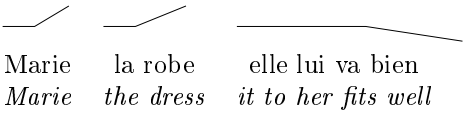
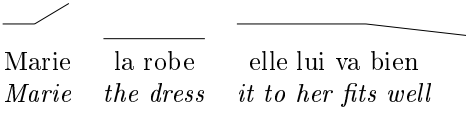
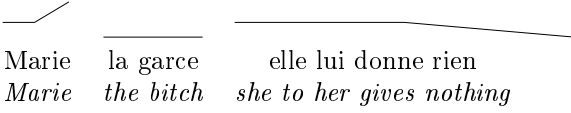
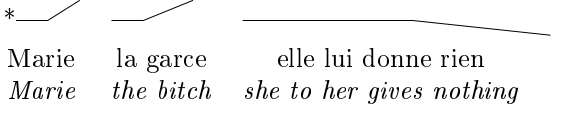
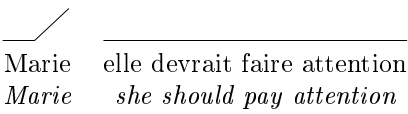
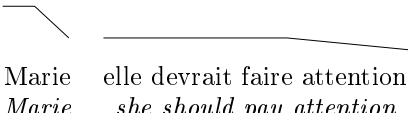
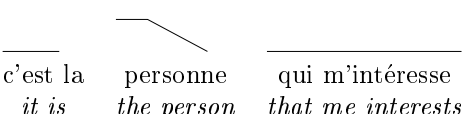
Such relations constitute a basis for describing and explaining variability. This phenomenon cannot be interpreted by means of descriptions coming from a unique domain. We propose an account of variability bringing together information coming from different components and stipulating an equilibrium relation between these components. The idea consists in indicating that as soon as a certain quantity of information (an information threshold) is reached thanks to some linguistic components, then variability can appear in other components. We will see for example that in the case where syntax contains information enough by itself, then prosody becomes variable.

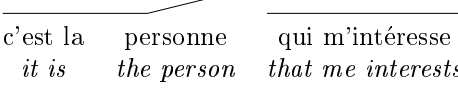
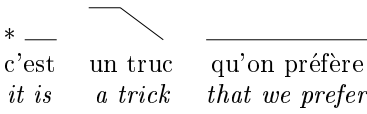
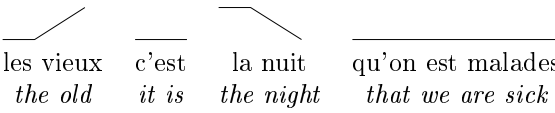
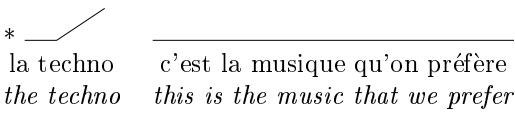
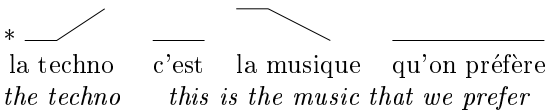
We propose to start from some examples illustrating some variability phenomena in the prosodic realization. We can then provide first some constraint specifying this variability. We define finally a principle providing a general framework for describing variability.

## 2 Some Examples

We present in this section some examples together with a stylized intonative contour. This kind of representation doesn't allow to represent the set of prosodic phenomena that should be taken into account for a precise study. It allows however a first approximation in the perspective of the question addressed in this paper. We use for this some of the notion proposed in [Rossi99], in particular the notion of conclusive, parenthetic and continuative

contours.

- (2)  This utterance is formed by two distinct parts, not linked with any explicit syntactic relation. Intonation gives a correlative interpretation indicating “*it rains, because you are wet*”.
- (3)  The same utterance, with a different intonation, receives a causative interpretation indicating “*if it rains, then you are wet*”.
- (4)  Example of a dislocation of two NPs with an anaphoric relation with two clitics. The intonation follows the same rising schema for each dislocated NP.
- (5)  The same example can be realized with a different intonative contour. In both cases, the interpretation of a double NP dislocation is favored, without many ambiguity (“*the dress fits Mary well*”)
- (6)  The syntactic structure is identical to the one of the previous example. However, the preferred interpretation is that of an apposition more than a multiple dislocation. This interpretation is reinforced by a different intonative contour between the NPs, the second one being parenthetic.
- (7)  The same example seems difficult to realize with an intonative contour typical to a double dislocation.
- (8)  The syntactic structure corresponds to a simple dislocation. The rising contour of the NP constitutes a strong prosodic mark.
- (9)  The form of this example is the same as in the previous example. However, the interpretation is rather a vocative one, more than a dislocated. This interpretation is then driven by the intonation, not by the syntactic structure.
- (10)  The syntactic structure is that of a cleft. In this realization, the intonation of the cleft NP is marked with a fall. The interpretation is something like “*it is the person that interests me (not the clothes)*”.

- (11)  c'est la personne qui m'intéresse  
it is the person that me interests
- In this example, the interpretation is that of a relative more than a cleft (of the kind “*this is her, effectively*”). It is driven by the intonation that presents a continuative contour rather than a conclusive one.
- (12)  \* c'est un truc qu'on préfère  
it is a trick that we prefer
- The interpretation of this example is that of a relative. Such interpretation is natural, without taking into account prosody (it is difficult to associate a cleft interpretation to this element which has a poor semantic content). A typical cleft intonation (with a conclusive contour) cannot be easily realized in this case.
- (13)  les vieux c'est la nuit qu'on est malades  
the old it is the night that we are sick
- In this example, the cleft interpretation is given by the syntactic structure (with a dislocated). This effect is reinforced by a conclusive intonative contour on the cleft NP.
- (14)  \* la techno c'est la musique qu'on préfère  
the techno this is the music that we prefer
- Contrarily to the previous example, the preferred interpretation is that of a relative. Such interpretation is given by the semantic level, the general form being identical with that of the previous example. The intonation reinforces this interpretation.
- (15)  \* la techno c'est la musique qu'on préfère  
the techno this is the music that we prefer
- A conclusive intonation, rather typical of a cleft, cannot be easily realized in this example.

### 3 Basic Constraints

The classical description of prosody/syntax relation is generally done by means of constraints representing either the necessity of a specific realization or its impossibility. In the perspective of a constraint-based approach, this kind of information is represented directly by means of properties of the objects. This is the case for example of Property Grammars, described in [Blache01], that rely on different kinds of constraints (e.g. requirement, exclusion, linear precedence, etc.).

At this stage, it is possible to stipulate a first set of constraints that will constitute a preliminary step in the description of the relations.

#### 3.1 Describing an object with several components

A same linguistic object is described by means of information coming from different components. This characteristics is illustrated by several examples of the previous section. Let's focus more precisely on examples 7-8. This case is apparently simple and regular. Indeed, if the data are verified, each interpretation (being vocative or not) is associated to a specific intonative contour without possibility of variation. We obtain then the several constraints that make it possible to precise some principles.

- [p1]  $SN[detached] \wedge Contour[conclusive] \Rightarrow [-vocative]$   
 [p2]  $SN[detached] \wedge Contour[continuative] \Rightarrow [+vocative]$

Constraint [p1] stipulates that a dislocated phrase, when coming with a marked intonation (typically a conclusive contour), takes a vocative interpretation. We are then in the case of a classical dislocation coming together with an anaphoric relation between the dislocated NP and the clitic. The vocative interpretation described in [p2] implies a detached NP plus an intonational fall. In these constraints, the objects belong to three different components: syntax, semantics and prosody. It is necessary to precise these domains. Moreover, it is also necessary to precise their respective positions. The solution making it possible to build a representation independently from any theoretical presupposition consists in indicating the position of the object in the acoustic signal (cf. [Bird99]). This kind of indication is direct for prosodic information, but difficult to specify for other domains such as syntax, semantics or pragmatics. We propose (see [Blache03]) a general indexation mechanism specifying a different kind of localization for any objects. We propose then to use an anchor containing a different kind of indexation: localization in the signal, in the string or in the context. A complex feature represents such anchor as follows:

$$anchor \rightarrow \left[ \begin{array}{l} \text{TEMPORAL } \langle i, j \rangle \\ \text{POSITION } \langle k, l \rangle \\ \text{CONTEXT } c \end{array} \right]$$

The temporal index is represented by two values (beginning and end). The position is also a couple of indexes (corresponding to nodes in a chart interpretation) localizing an object in the input. The CONTEXT feature implements the notion of *universe* (i.e. a set of discourse referents) as in DRT. An object can then be specified by means of three kinds of information: its *domain*, its *anchor* and its *characterization* (the set of corresponding properties). The following example describes an object from the syntactic domain, with a precise localization both on the temporal and the linear axis:

$$obj \rightarrow \left[ \begin{array}{l} \text{DOMAIN } synt \\ \text{ANCHOR } \left[ \begin{array}{l} \text{TEMP } \langle 880, 1000 \rangle \\ \text{POSITION } \langle 2, 3 \rangle \end{array} \right] \\ \text{CHARAC } [\text{CAT } Det] \end{array} \right]$$

As detailed above, constraints [p1] and [p2] are expressed in terms of implication. However, the kind of relation represented there consists more precisely in a co-variation of the different values. It is moreover necessary, in particular for the representation of information at a finer level than that of the atomic object, to express an element under the form of a set of features, each one being an attribute/value pair. This is the case for example of a phoneme that can be characterized by a set of segments or a syntactic category that corresponds to a set of morphological, syntactic and semantic features. The relation [p2] concerns in fact different features of a same object characterizing a subpart of the utterance. This object is represented as follows:

$$p2 \rightarrow \left[ \begin{array}{l} \text{SYNT } \left[ \begin{array}{l} \text{CAT } [NP_{detached}] \\ \text{POS } \langle i, j \rangle \end{array} \right] \\ \text{SEM } \left[ \begin{array}{l} \text{TYPE } +vocative \\ \text{POS } \langle i, j \rangle \end{array} \right] \\ \text{PROS } \left[ \begin{array}{l} \text{CONTOUR } continuative \\ \text{POS } \langle i, j \rangle \end{array} \right] \end{array} \right]$$

Such a feature structure makes it possible to represent at the same time information coming from different components and participating to the description of a same object or, more generally, a same linguistic phenomenon. Each characteristic is associated with a position in the signal represented by the complex feature ANCHOR. The different information is still represented separately, the feature structure being a way for describing an object containing features connected with some relations.

The covariation relation specified above is expressed by the specification of a simultaneous variation of the value of some features in a structure. There are several ways to represent this kind of relation, one of them being the use of “named disjunctions” (cf. [Kasper95] or [Blache98]). The mechanism consists in enumerating the set of possible values for each feature and indicating the values that are in a mutual dependency. All values belonging to the same part of the disjunction covary: when a value is instantiated, then all other values of the same rank in the named disjunction are also instantiated.

$$detached/dislocated \rightarrow \left[ \begin{array}{l} \text{SYNT} \left[ \begin{array}{l} \text{CAT} \{ [NP_{detached}] \vee_1 [NP_{dislocated}] \} \\ \text{ANCH} \langle i, j \rangle \end{array} \right] \\ \text{SEM} \left[ \begin{array}{l} \text{TYPE} \{ +\text{VOCATIVE} \vee_1 -\text{VOCATIVE} \} \\ \text{ANCH} \langle i, j \rangle \end{array} \right] \\ \text{PROS} \left[ \begin{array}{l} \text{CONTOUR} \{ \text{CONTINUATIVE} \vee_1 \text{CONCLUSIVE} \} \\ \text{ANCH} \langle i, j \rangle \end{array} \right] \end{array} \right]$$

In this example, the named disjunction is represented by  $\vee_1$ . The values  $NP_{detached}$ ,  $+vocative$  and  $continuative$  are then dependent (first part of the disjunction), as well as the values  $NP_{dislocated}$ ,  $-vocative$  and  $conclusive$ . The previous structure works then as a constraint on the concerned objects. As soon as an utterance description needs a set of features specified in this structure, their values have to satisfy the constraint.

### 3.2 Information on different parts of a same object

A quick study of examples 9-14, describing cases of clefts and relatives, exhibits a first property constraining the relative. This one is incompatible with a conclusive contour as in the case of a cleft. This restriction is represented by the following constraint stipulating that a set of categories constituting a NP with a relative cannot be realized with an intonative stress on the name, which corresponds to a parenthetic contour.

$$relative \rightarrow \left[ \begin{array}{l} \text{SYNT} \left[ \left\langle \left[ \begin{array}{l} \text{CAT DET} \\ \text{ANCH} \langle i, j \rangle \end{array} \right] \left[ \begin{array}{l} \text{CAT N} \\ \text{ANCH} \langle k, l \rangle \end{array} \right] \left[ \begin{array}{l} \text{CAT REL} \\ \text{ANCH} \langle m, n \rangle \end{array} \right] \right\rangle \right] \\ \text{SEM} \left[ \begin{array}{l} \text{TYPE } -\text{focus} \\ \text{ANCH} \langle i, l \rangle \end{array} \right] \\ \text{PROS} \left[ \begin{array}{l} \text{CONTOUR } \textit{parenthetic} \\ \text{POS} \langle i, l \rangle \end{array} \right] \end{array} \right]$$

In this example, we can remark on top of the constraint on the relative, the possibility for a same object to represent information on different parts. Syntactic information concerns then the entire structure whereas semantic and prosodic information only concerns a subset. The feature ANCH implements this aspect.

## 4 Variability

The kind of constraints presented above can represent many different relations between components of linguistic analysis. However, it is impossible to provide general descriptions that cannot be captured by covariation. In particular, it is difficult, or even impossible, using such an approach to explain why prosodic realization seems less constrained under certain circumstances. In the case of the distinction between clefts and relatives, a constraint can characterize the general realization of the relative, but nothing can be said as for the cleft: we can remark a great variability for this construction. Different corpus studies show that it seems possible to realize clefts without any specific prosodic mark or with many different marks. The same phenomenon appears when a semantic feature reinforces a syntactic turn. This is the case of the examples 3-6 that present some cases of simple or multiple dislocations. In the case of multiple dislocation (example 3), two clitics are in an anaphoric relation with the detached NP. In this case, we have a morpho-syntactic criterion (two clitics agreeing with the NPs) plus a semantic index (the anaphoric relation). We can then consider that, whatever the prosody, the interpretation is constrained enough by information coming from syntax and semantics. On the contrary, when the anaphoric relation doesn't exist, as in the examples 5 and 6, prosody is strongly constrained and plays an important role in the interpretation. For example, the second realization that would consider the two NPs at the same level (favoring a double dislocation interpretation) is impossible.

Generally speaking, we can consider that, when an utterance cannot be disambiguated with a morpho-syntactic mark, then the prosody would play this role. In the examples 1-2, intonation in itself makes it possible to distinguish between causative and correlative interpretations. More clearly, in the case of the examples 9-11, intonation drives the interpretation as relative or cleft. A salient turn cannot be assigned to a relative, a cleft interpretation is then favored in this case. Cleft variability would then come from the fact that this turn is strongly marked from a morpho-syntactic point of view (more than the relative). In the same way as for double dislocation, morpho-syntactic and semantic constraints are strong enough and the interpretation doesn't need more information, for example a prosodic one. This characteristic is also clear in the examples 12-14. One can observe in the same way a syntactic variability allowed by prosody. For example, a rising intonative contour is classically associated with interrogative turns. We consider in this case that the intonative schema is not very ambiguous, it can be associated with an heavy *weight* for prosody (in the same way as clefts have an heavy weight for syntax). Such a characteristic allows variability of the syntactic form. In a general way, and for any component of linguistic analysis, the weight value is proportional to the ambiguity degree of the form. For example, a conclusive contour, specific to certain constructions, or a major break are associated to heavy weights for prosody. In the same way, marked syntactic constructions with strong morpho-syntactic elements (such as clefts) correspond to heavy syntactic weights.

There exists then a relation between syntax, semantics and prosody that cannot be represented classically in terms of correspondences between the respective structures. More precisely, the only constraints that can be proposed in this perspective are those of cooccurrence restrictions, but without providing an account of variability.

## 5 Equilibrium Principle

We propose here a general framework explaining relations between the different linguistic components and their variability. It is important to remind first some points. We think that it is not necessary to try to describe directly relations between components. Only some constructions or some phenomenon can be explained in this way. In the case in which no explicit relation exists between the domains, the realization of each component can be done independently from the others. Moreover, such a relation can concern the entire component or one of its subpart.

We consider that each phenomenon bears a specific weight encoding in some way its importance in the component. We have seen in the previous section some examples of heavy weights: cleft for syntax, conclusive contours for prosody, anaphoric relations for semantics, etc. But such information taken separately for each component doesn't explain in itself variability or insistence phenomena.

The hypothesis relies on a principle stipulating that an equilibrium between the different components has to be reached. In other words, a certain amount of information has to be given before allowing variability. This information can come from the different domains or components. Concretely, the sum of the weights of each component has to reach a certain threshold. In this perspective, variability becomes possible within a component if the sum of the weights is greater than this threshold. For example, if for some construction, the threshold is reached only by means of a the weights coming from syntax and semantics, then the weight of the prosody is not constrained and can take any value. Concretely, this means that intonation is not anymore constrained and can be realized in a variable way (following however the intrinsic constraints related to the utterance). On the contrary, when the syntactic and the semantic weights are not heavy enough, then the weight of the prosody becomes necessary to reach the threshold and the intonation is less variable. Obviously, as it is shown in the previous section, the threshold can be reached by means of the sum of the weights of any components (including a single one). We can reach for example the threshold by means of prosodic and semantic weights, the syntax becoming variable.

The insistence effects are explained in the same manner: the threshold only needs to be reached in order to observe eventually such phenomenon. For example, if the syntactic and semantic weights are heavy enough to reach the threshold, then the realization of a marked intonation (associated to an heavy prosodic weight) is not necessary. If it occurs, it is interpreted then as an insistence effect whereas the same contour in the configuration of a light syntactic weight (not sufficient to reach the threshold with the semantic one) would not receive the same interpretation.

The equilibrium principle explaining this weight repartition can be described by the following constraint indicating that the weight of a structure is the sum of the syntactic, semantic and prosodic weights and that this value has to reach a given threshold  $t$ :

$$\text{Equilibrium Principle} \rightarrow \left[ \begin{array}{l} \text{SYNT [WEIGHT } a] \\ \text{SEM [WEIGHT } b] \\ \text{PROS [WEIGHT } c] \\ \text{THRESHOLD } (a+b+c) \geq t \end{array} \right]$$

In the examples presented in the first section, we have seen that some constructions such as relatives don't have much possibility of variability. This is explained by the fact that each

component contributes necessarily to the equilibrium of the structure. This is described by a constraint on the respective weights of the components describing the relative:

$$Relative \rightarrow \begin{bmatrix} SYNT [WEIGHT 1] \\ SEM [WEIGHT 1] \\ PROS [WEIGHT 1] \\ THRESHOLD 3 \end{bmatrix}$$

In this constraint, the threshold is set (arbitrarily) to 3. It is reached minimally by the sum of the weights of each component. None of them can then vary. They are all considered as light. On the contrary, the situation is different in the case in which a component has a heavy weight, such as clefts which have a heavy syntactic weight. The weight values of each component are described in this case in terms of local constraints specifying the variability. The syntax bears a weight set to 2. The semantic has a weight at least equal to 1. If the threshold is set to 3, then the prosodic weight can have any value, including 0, for a realization prosodically unmarked.

$$Cleft \rightarrow \begin{bmatrix} SYNT [WEIGHT 2] \\ SEM [WEIGHT b > 0] \\ PROS [WEIGHT c] \\ THRESHOLD 3 \end{bmatrix}$$

Such a constraint describes then a certain variability for semantic and a free variability for prosody.

## 6 Conclusion

The description of the relations between the different linguistic components relies on two propositions presented in this paper: the possibility of representing several kinds of information on a same object and the specification of an equilibrium principle between these components. Such a principle relies on the possibility of assigning a weight to each component plus an equilibrium threshold to be reached for any specific construction.

Several aspects of this proposition still remain to be precised. First, the notion of weight relies essentially on the ambiguity level of the concerned domain. However, several other parameters could be used such as structure frequency, its morphological specificity, etc. In another way, the different objects or constructions also have to be specified. In the same way as the notion of weight is new in linguistics, the use of a description unit common to several linguistic domains comes to propose the idea of a more general object, composed with several linguistic components. Such an object also constitutes a new description tool that remains to be defined, but already provides some elements of analysis for phenomena such as variability. Concretely, the principle proposed here provides a precise framework making it possible to implement variability in some applications such as natural language generation or speech synthesis.

## References

[Bear90] Bear J. & P. Price (1990), "Prosody, Syntax and Parsing", in proceedings of the *ACL-90*.

- [Bird99] Bird S. & M. Liberman (1999) “A formal framework for linguistic annotation”, Technical Report MS-CIS-99-01. Dept of Computer and Information Science, University of Pennsylvania.
- [Blache98] Blache P. (1998) “Parsing Ambiguous Structures using Controlled Disjunctions and Unary Quasi-Trees”, in proceedings of *COLING-ACL'98*.
- [Blache00] Blache P. & D. Hirst (2000) “Multi-level Annotation for Spoken Language Corpora”, in proceedings of *ICSLP-00*.
- [Blache01] Blache P. (2001) *Les Grammaires de Propriétés : Des contraintes pour le traitement automatique des langues naturelles*, Hermès.
- [Blache02] Blache P. (2002) “Variabilité et dépendances des composants linguistiques”, in proceedings of *TALN-02*.
- [Blache03] Blache P. (2003) “Representing Interaction Between Linguistic”, submitted.
- [Blasco99] Blasco-Dulbecco M. (1999) *Les dislocations en français contemporain. Etude syntaxique*, Champion.
- [DiCristo99] Di Cristo A. & L. Jankowski (1999) “Prosodic Organisation and Phrasing after Focus in French”, in proceedings of *ICPhS-99*.
- [Doetjes02] Doetjes J., E. Delais-Roussarie & P. Sleeman (2002) “The Prosody of Left Detached Constituent in French”, in proceedings of *Speech Prosody 2002*.
- [Dorre90] Dvrré J. & A. Eisele (1990) “Feature Logic with Disjunctive Unification”, in proceedings of *COLING-90*.
- [Hirst&DiCristo98] Hirst D. & A. Di Cristo (1998) *Intonation Systems*, Cambridge University Press.
- [Mertens01] Mertens P., J.-P. Goldman, E. Wehrli & A. Gaudinat (2001) “La synthèse de l'intonation ‘ partir de structures syntaxiques riches”, in revue *TAL* 42 :1
- [Pollard94] Pollard C. & I. Sag (1994) *Head-driven Phrase Structure Grammars*, CSLI, Chicago University Press.
- [Rossi99] Rossi M. (1999) *L'intonation, le système du français : description et modélisation*, Ophrys.
- [Sag99] Sag I. & T. Wasow (1999) *Syntactic Theory. A Formal Introduction*, CSLI.
- [Selkirk84] Selkirk E. (1984) *Phonology and Syntax: the Relation between Sound and Structure*, MIT Press.