

## “Subordinate” clauses and syntactic annotation of spoken French corpora

Frédéric SABIO (Université de Provence, LPL)  
Sylvain KAHANE (Université Paris Ouest, Modyco)  
Paola PIETRANDREA (Università Roma Tre, Lattice)  
Anne LACHERET (Université Paris Ouest, Modyco)

The grammatical analysis of clauses introduced by a “subordinating conjunction” has always been a challenge for linguists because, on one hand, spontaneous spoken data exhibits highly variable syntactic and discursive organizations which have never been properly described through the sentence-based framework of traditional grammar; and on the other hand because the generalized reference to the classic notion of “subordination” tends to unify in an artificial way several types of syntactical configurations that it would be advisable to distinguish carefully.

Within the *Rhapsodie* project (directed by Anne Lacheret, Univ. of Paris Ouest), which is devoted to the syntactic and prosodic tagging of spoken French, we have been directly confronted to such difficulties, and we have had to make some methodological choices which will be the theme of our communication.

Our description is based on a framework that was initially developed by Claire Blanche-Benveniste (Univ. of Aix-en-Provence), according to which two levels of grammatical analysis should be distinguished, namely the microsyntactic level (which accounts for government and dependency relations) and the macrosyntactic level (which investigates the links between successive elements that share a discursive relationship). Our main descriptive options are the following :

1) The sole presence of a “subordinating conjunction” does not prove in itself the existence of any genuine dependency relation in the utterance. In the theoretical frame of “Pronominal approach” (Blanche-Benveniste et al. 1984), we consider that syntactic dependency must necessarily correlate with several paradigmatic properties, such as the equivalence with a pronoun, the possibility to be cleft, and some other criteria. As a matter of fact, besides clearly *dependent* sequences, that can readily be analyzed as subordinated clauses [ex.1-3], there are many others, particularly frequent in spontaneous data, that do not possess any paradigmatic property, and thus appear to be merely *linked* to the neighboring construction, sharing with it a simple “association”, or paratactic, link [4-6]. We will show that the semantic and syntactic properties of both types are quite different (for example, the conjunction in [4] is followed by an imperative clause, which would not be possible with a truly embedded construction).

1. le métier de fleuriste était pas drôle *parce que il fallait avoir les mains dans l'eau* [oral, la fleuriste]  
[lit. working as a florist wasn't funny because you always had to keep your hands in the water]
2. nous avons vu une euh - euh un crépuscule euh - *alors que nous étions d- au au sommet de la mosquée* [AIX-R00PRI001]  
[lit. we saw a er – er a twilight while er we were i- at the at the top of the mosque]
3. il chantait à Saint Laurent à la cathédrale *quand il y avait des fêtes* [Corpaix]  
[lit. he used to sing at Saint-Laurent in the cathedral *when there were parties*]
4. vos clients euh pourront euh à cet endroit admirer la vue sur le lac et le barrage + *parce que n'oubliez pas que le le Muséoscope surplombe le lac de Serre Ponçon* hein [fpubl02]

- [lit. your customers er can er in this place admire the sight on the lake and the dam + *because don't forget that the Muséoscope overhangs the lake of Serre Ponçon*]
5. ici par exemple c'est du corail qu'elle va porter dans sa corne d'abondance - *alors que là-bas ça sera des fruits* [CRFP]  
[lit. here for example it is coral that she is going to carry in her horn of plenty - *while over there that will be fruits*]
6. quand je vois les les les les les élèves qui descendent dans la rue et tout moi je les soutiens [Corpaix]  
[lit. when I see the the the the the pupils who go down in the street and stuff me I support them]

2) The constructions can be organized into grouping of several macrosyntactic units that can be characterized with reference to their linear position, their prosodic contour, and their illocutionary status. Such groupings are described as a [pre-Nucleus / Nucleus / post-Nucleus] sequence in our approach.

3) Those two levels only partially overlap: a given microsyntactic construction can be realized through several macrosyntactic configurations; and inversely, a single macrosyntactic sequence can be associated with a variety of very different microsyntactical structures.

Our communication will present a description of clauses introduced by a “subordinating conjunction” on the basis of both micro and macrosyntactic relationship. We distinguish the five following types:

a- Microsyntactically governed clauses that are realized in the same Nucleus as the verb that embeds them; this is the most canonical type of subordinate clauses [ex. 1-3].

b- Microsyntactically governed clauses that are realized either in the post-Nucleus or in the pré-Nucleus, thus having a “detached” realization [ex.7]:

7. quand j'ai été plus grande alors on m'a appris à faire des bouquets [oral, la fleuriste]  
[lit. when I grew older then I was taught to make bouquets]

c- Microsyntactically governed clauses that are realized as a “delayed” clause, with an “afterthought” effect, as in [8]:

8. faut s'accrocher pour comprendre surtout quand tu es pas anglais [Corpaix]  
[lit. It is necessary to hang on to understand especially when you are not English]

d- Microsyntactically linked sequences (that is: sharing no dependence relationship) realized as a pre- or post-Nucleus [ex. 6].

e- Microsyntactically linked sequences forming an autonomous unit that constitutes a specific type of Nucleus with its own illocutionary force [ex. 4-5].

The tagging system which has been developed within the framework of the *Rhapsodie* project gives an indication of both the microsyntactic dependence relationships and the macrosyntactic groupings. By taking those two levels of analysis into account, it allows us to describe most of the attested uses of conjunctions, including the most problematic ones. Our annotation system will be illustrated with a selection of corpus-drawn utterances.

## References

- ABEILLÉ A., GODARD D. SABIO F., 2008, “Two types of NP preposing in French”, *Proceedings of the 15th HPSG Conference*, Stefan Müller (ed.), CSLI Publications, Stanford, 306-324.

- BALLY C., 1932, *Linguistique française et linguistique générale*, Paris, Leroux.
- BEGUELIN M.-J., AVANZI M., CORMINBOEUF G., (éds.) 2010, *La parataxe*, Berne, Peter Lang.
- BENZITOUN C., 2006, *Description morphosyntaxique du mot quand en français contemporain*, Thèse de doctorat, Université de Provence.
- BENZITOUN C., DISTER A., GERDES K., KAHANE S., MARLET R. (2009), « Annoter du des textes tu te demandes si c'est syntaxique tu vois », *Arena Romanistica*, 4, 16-27.
- BENZITOUN C., DISTER A., GERDES K., KAHANE S., PIETRANDREA P., SABIO F., 2010, « Tu veux couper là faut dire pourquoi. Propositions pour une segmentation syntaxique du français parlé », *Actes du Congrès Mondial de Linguistique Française (CMLF 2010)*, 2075-2090, La Nouvelle Orléans.
- BENZITOUN C., DEBAISIEUX J.-M., DEULOFEU J., DISTER A., GERDES K., KAHANE S., LEFEUVRE F., PIETRANDREA P., ROSSI-GENSANE N., SABIO F., VICTORRI B. (en développement), *Guide d'annotation syntaxique Rhapsodie*, <http://rhapsodie.ilpga.fr/wiki>.
- BERRENDONNER A. à par., *Grammaire de la période*.
- BERTRAND R., PORTES C., SABIO, F., 2007, « Distribution syntaxique, discursive et interactionnelle des contours intonatifs du français dans un corpus de conversation », *Travaux Neuchâtelois de Linguistique*, 47, 59-77.
- BILGER M., 1998, « Le statut micro et macrosyntaxique de *Et* », in Bilger, van den Eynde & Gadet (éds.), *Analyse linguistique et approches de l'oral*, Leuven-Paris, Peeters, p.91-102
- BLANCHE-BENVENISTE C., DEULOFEU J., STEFANINI J. et VAN DEN EYNDE K., 1984, *L'Approche pronominale et son application au français*, Paris : SELAF.
- BLANCHE-BENVENISTE C., BILGER M., ROUGET C., VAN DEN EYNDE K., MERTENS P., 1990, *Le français parlé, études grammaticales*, Paris, Éditions du CNRS.
- CHAROLLES M., VIGIER, D., 2005, « Les adverbiaux en position préverbale : portée cadrative et organisation des discours », *Langue Française*, p.148, 9-30.
- DEBAISIEUX J.-M., 2006-a, «La distinction entre dépendance grammaticale et dépendance macrosyntaxique comme moyen de résoudre les paradoxes de la subordination », *Faits de Langues*, 28, p.119-132.
- DEBAISIEUX J.-M. & DEULOFEU J., 2009, «When a construction constructs the context » in A. Bergs and G. Diewald (eds.) *Context and constructions*, Amsterdam, John Benjamins, p.43-62.
- DEULOFEU J., 2010, « La greffe d'un énoncé sur une construction : une combinaison originale de parataxe et de rection », in Beguelin M.-J. & Corminbeuf G. *La Parataxe*, Berne, Peter Lang, p.175-208.
- DEULOFEU J., 2003, « L'approche macrosyntaxique en syntaxe : un nouveau modèle de rasoir d'Occam contre les notions inutiles », *Scolia*, 16, 77-95.
- MILLER J. et WEINERT R., 1998, *Spontaneous spoken language. Syntax and Discourse*, Clarendon Press, Oxford.
- MITHUM M., 2005, "On the assumption of the sentence as the basic unit of syntactic structure", in *Linguistic diversity and language theories*, Frayzingier Z., Hodges A. and Rood D.S. (eds.), *Studies in language, Companion series*, Amsterdam, John Benjamins, p.169-183
- NARBONA JIMENEZ A., 1990, *Las subordinadas adverbiales improprias en español (II). Causales y finales, comparativas y consecutivas, condicionales y concesivas*, Málaga, Librería Ágora.
- ROUBAUD M.-N., 2000, *Les constructions pseudo-clivées en français contemporain*, Paris, Champion.

- SANDFELD K.R., 1943, *Syntaxe du français contemporain – II Les propositions subordonnées*, Paris, Droz.
- VERSTRAETE J.-C., 2007, *Rethinking the coordinate subordinate dichotomy*, Berlin, Mouton.