



Entre syntaxe, prosodie et discours : les topiques sujets en français parlé

Lisa Brunetti, Mathieu Avanzi, Cédric Gendrot

► To cite this version:

Lisa Brunetti, Mathieu Avanzi, Cédric Gendrot. Entre syntaxe, prosodie et discours : les topiques sujets en français parlé. France. 1, pp.2041 - 2054, 2012, <10.1051/shsconf/20120100209>. <hal-01118716>

HAL Id: hal-01118716

<https://hal-univ-diderot.archives-ouvertes.fr/hal-01118716>

Submitted on 20 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Entre syntaxe, prosodie et discours : les topiques sujet en français parlé*

Brunettiⁱ, Lisa, Avanziⁱ, Mathieu & Gendrot^k, Cédric

ⁱLaboratoire Dynamique du Langage, Université Lyon 2

^jUniversité de Neuchâtel

^kLaboratoire de Phonétique et Phonologie, Université Paris3

lisa.brunetti@univ-lyon2.fr, mathieu.avanzi@unine.ch, cgendrot@univ-paris3.fr

1 Introduction

En français parlé, un SN en tête de phrase peut être repris ou non par un pronom co-référent dans la construction verbale qui le suit. Les deux possibilités sont illustrées ci-dessous :

(1) **ma mère** était institutrice [CID, Bertrand et al., 2008]

(2) **ma mère**, elle, est née le dix-sept février [CFPP, Branca-Rosoff et al., 2009]

Sur les plans syntaxique, prosodique et informationnel, les propriétés de la construction (2), communément appelée « dislocation à gauche », se distingueraient de celles de la construction (1).

Sur le plan syntaxique, selon des auteurs comme Lambrecht (1994), Delais-Roussarie et al. (2004), De Cat (2007), Riegel et al. (2009), entre autres, le SN en (2) (qui pour des raisons de brièveté nous appellerons désormais « SNx2 ») se trouverait syntaxiquement en dehors du champ de la réaction verbale. Rien de tel lorsque le SN n'est pas repris par un élément pronominal (désormais « SNx1 ») : il se placerait directement sous le même nœud syntaxique que le SV. Les représentations arborescentes ci-dessous donnent une illustration très schématique de cette différence de fonctionnement syntaxique :

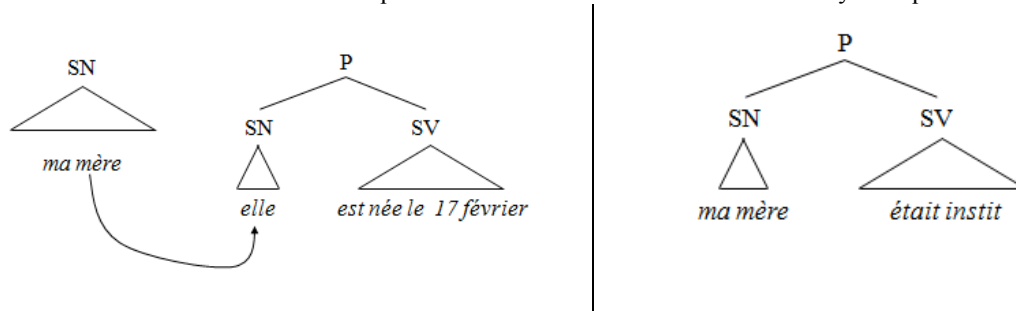


Figure 1. Représentations arborescentes schématiques d'un SNx2, ex. (2), et d'un SNx1, ex. (1).

Si donc en (1) le SN *ma mère* est le sujet de la phrase, en (2) le sujet proprement dit serait le pronom *elle*, qui établit un lien anaphorique avec le SN *ma mère*, situé pour sa part en dehors de la phrase¹. Du point de vue prosodique, des chercheurs comme Bally (1942), Rossi (1999), Delais-Roussarie et al. (2004), Mertens (2008) ou encore De Cat (2007) ont fait l'hypothèse que l'élément disloqué serait systématiquement assorti d'une frontière prosodique majeure². De fait, le SNx2 serait toujours ponctué d'une prééminence forte sur le plan perceptif, actualisée dans la substance par une montée mélodique importante et un allongement syllabique relatif significatif, comme c'est le cas du SNx2 *les parents*, dont la Figure 2 donne une représentation du tracé mélodique. Comme on peut le voir, la syllabe finale du SN est assortie d'un mouvement mélodique montant de 10 dt environ, et fait l'objet d'un allongement relatif (50%) par rapport à son entourage.

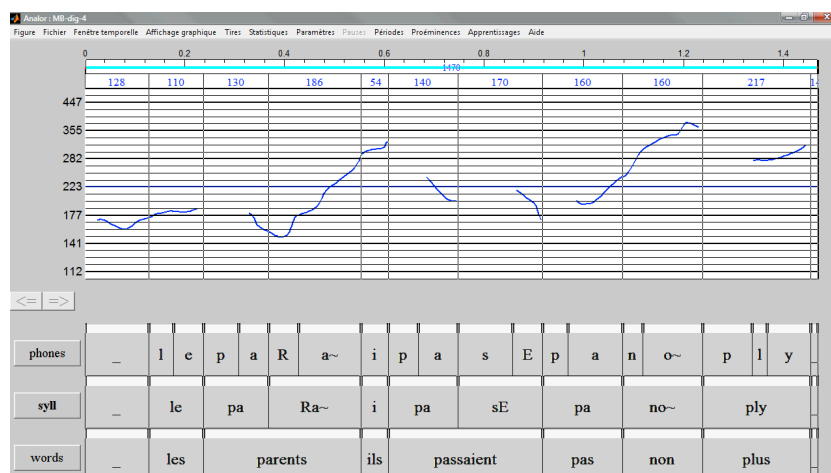


Figure 2. Copie d'écran Analor. Analyse de la séquence *les parents ils passaient pas non plus* [CID]. L'évolution de la f0 (en traits noirs) peut être mesurée en demi-tons (en filigrane, la distance entre deux lignes fines vaut 1 demi-ton, la distance entre deux traits épais une quarte (4 demi-tons)) ; la durée des segments étiquetés est donnée en millisecondes au-dessus de la bande dans laquelle évolue la courbe de f0. Les chiffres plus foncés au-dessus indiquent le temps en secondes du segment par rapport à l'enregistrement du fichier total. Les différentes couches d'alignement, importées directement depuis les fichiers d'alignement au format Textgrid (Praat, Boersma & Weeninck, 2011), sont affichées en dessous de cette bande, en l'occurrence, de haut en bas : les phonèmes et les syllabes en alphabet SAMPA, et la transcription des mots graphiques.

Les SNx1, comme c'est le cas dans l'exemple (1) ci-dessus, ne sont pas aussi contraints, et peuvent ou non être suivis d'une frontière prosodique plus faible (Rossi 1999, Delais-Roussarie & Post, 2008).

Enfin, d'un point de vue pragmatique, un élément disloqué à gauche est généralement considéré comme jouant le rôle de topique de phrase, à savoir le référent à propos duquel porte la phrase (cf. Strawson, 1964, Reinhart, 1981, Lambrecht, 1994) :

There is a general consensus in the literature on dislocation that LD [left dislocations, *ndr*] and RD [right dislocations, *ndr*] are topic marking constructions, i.e. grammatical constructions which serve to mark a constituent as denoting the topic (or theme) with respect to which a given sentence expresses a relevant comment (Lambrecht, 2001: 1072)

Sur ces bases, la construction à SNx2 serait caractérisée par une morphosyntaxe et une prosodie propres, qui la distinguent clairement des constructions à SNx1. Mertens résume cette différence en ces termes :

Dans *l'année universitaire commence demain*, l'accentuation de *année* est facultative : soit on l'accentue, soit on l'intègre au GI [Groupe Intonatif, *ndr*] connexe qui suit. En revanche, dans *l'année, elle commence demain*, non seulement l'accentuation de *année* est obligatoire, mais la construction exige aussi une frontière majeure après ce GI. (Mertens, 2008 : 107)

A date plus ou moins récente, un nombre important de chercheurs ont remis en cause cette description, et ont argué que dans la parole spontanée, les choses étaient plus complexes.

Ainsi, sur le plan morphosyntaxique, des chercheurs comme Berrendonner (1993), Zribi-Hertz (1994), Apothéloz (1997), Fonseca-Greber & Waugh (2003), etc. ont affirmé qu'en français parlé contemporain les propriétés de l'élément de reprise ne seraient plus celles d'un pronom à proprement parler, mais plutôt celles d'un affixe verbal. Selon ces auteurs, la présence du clitique serait due à des raisons sémantiques (Culbertson & Legendre, 2008, Culbertson, 2010) ou variationnistes (Coveney, 2003, Berrendonner, 2007), mais en tout cas pas syntaxiques.

Cette proposition a donc des conséquences syntaxiques importantes : si le clitique n'est plus un pronom, la position de sujet dans la phrase reste libre et peut être occupée par le SN. En d'autres mots, le SNx2 ne serait plus à considérer comme en dehors de la phrase, mais il occuperait la même position que le SNx1 dans la représentation à droite de la Figure 1. Il s'agirait en somme d'une construction à redoublement du sujet.

Quant à la prosodie, on a remarqué que le SNx2 n'était pas toujours suivi d'une frontière prosodique majeure, mais pouvait être suivi d'une frontière prosodique de rang inférieur (Fradin, 1990, Gabriel & Rinke, 2011). Jun & Fougeron (2002) montrent ainsi que la réalisation de frontières plus faibles à l'articulation de constructions segmentées est monnaie courante dans les contextes de débit rapide, où tous les contours de l'énoncé se présentent sous une forme « réduite ». Les raisons qui permettraient d'expliquer pourquoi, dans certains contextes, le SNx2 est suivi d'une frontière prosodique forte ou d'une frontière prosodique plus faible ne semblent toutefois pas être expliquées de façon satisfaisante sur la base de critères rythmiques uniquement. Comme on le voit sur la Figure 3, le contour dont est assorti le dernier terme de la séquence (*sanatorium*) ne se présente pas sous forme réduite. On ne peut donc pas invoquer dans ce cas le débit comme étant à l'origine de l'absence de frontière forte après le SNx2 *mon mari* :

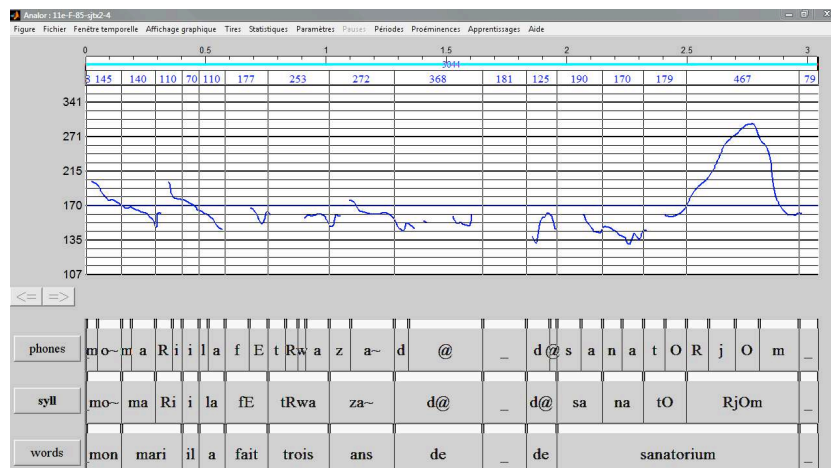


Figure 3. Copie d'écran Anamor. Analyse de la séquence *mon mari il a fait trois ans de de sanatorium* [CFPP].

D'autres facteurs sembleraient donc entrer en ligne de compte pour expliquer le degré de force de la frontière dont sont assortis les SNx2. Selon des auteurs comme Lacheret-Dujour & François (2004), ou plus récemment Grobet & Simon (2009) et Likhacheva (2008), ces facteurs seraient à chercher dans la fonction discursive du SN disloqué. Pour ces auteurs, le degré de force de la frontière droite du SNx2 dépendrait du degré d'activation du topique dans le discours : le degré de proéminence accentuelle serait inversement proportionnel au niveau d'activation. Nous résumons cette hypothèse sous (3) :

(3) Plus le référent dénoté par le SN est actif dans le discours, moins la frontière prosodique dont le SN est assorti est forte sur le plan perceptif, et inversement.

En pratique, cela revient à prédire la présence d'une frontière prosodique de faible niveau quand le topique est actif, une frontière un peu plus importante quand le topique est semi-actif, une frontière prosodique importante dans les contextes où le topique n'est *plus* actif.

A notre connaissance, l'hypothèse (3), n'a pas encore connu de validation empirique. Grobet & Simon, de même que Lacheret-Dujour & François ou Likhacheva, reconnaissent qu'une étude quantitative sur corpus serait nécessaire pour valider cette hypothèse. Le présent travail veut combler ce manque, en analysant quantitativement le rapport entre proéminence prosodique et propriétés discursives du SN avec clitique sujet de reprise.

Une autre caractéristique pragmatique du topique qui a été souvent associée au degré de force de la frontière est le contraste (cf. Buring, 1997 pour l'allemand et l'anglais, Marandin et al., 2002 pour le français, Frascarelli & Hinterhoelzel, 2007 pour l'italien et l'allemand, Brunetti et al., 2010 pour l'italien). Selon ces auteurs, le constituant topical serait signalé par une proéminence prosodique quand la proposition dans laquelle il se trouve est interprétée comme faisant partie d'un ensemble de propositions alternatives dans le discours. Cette interprétation est aussi généralement associée dans la littérature à une dislocation à gauche.

Dans notre travail nous voulons tester également cette hypothèse, et notamment voir si l'interprétation contrastive de l'énoncé est marquée prosodiquement par une frontière prosodique forte sur le SN et/ou syntaxiquement par une construction comme (2), où il y a coréférence entre le SN et un pronom clitique sujet.

Par ailleurs, reste un certain nombre de problèmes en suspens que nous souhaiterions également aborder dans cet article :

- Nous avons vu que pour de nombreux auteurs, les clitiques sujets en français seraient des affixes verbaux, et donc que la présence du clitique ne voudrait pas dire *de facto* que le SN qui co-réfère avec lui est disloqué. Si une telle hypothèse était correcte, et étant donné qu'une proéminence prosodique peut aussi être présente dans les SNx1, alors on doit stipuler que ces SNs seraient également des topiques, et seraient susceptibles d'entrer sous le couvert de la règle (3), ou bien pourraient avoir une interprétation contrastive.
- Nous chercherons également à vérifier l'hypothèse que la présence d'un clitique et la présence d'une frontière prosodique forte constituent deux marquages alternatifs de la même fonction de topique pour le SN. La coprésence de ces deux types de marquage serait témoignage d'une situation transitoire en français parlé contemporain, situation où deux systèmes grammaticaux différents coexistent. Dans ce cas-là, nous nous attendrions que le marquage prosodique soit beaucoup plus présent dans les SN *sans* reprise et beaucoup moins présent dans les SN *avec* reprise.

Cette étude nous permettra, au final, d'apporter des éléments de réflexion à la question fondamentale suivante : quelle est la nature des relations entre morphosyntaxe (présence ou pas du clitique sujet), prosodie (présence ou pas d'une frontière prosodique forte) et propriétés discursives (activation ou réactivation d'un topique, interprétation contrastive) ?

2 Données

Les séquences sur lesquelles nous travaillons ont été extraites du corpus de français parisien CFPP2000 (Branca-Rosoff et al., 2009) et du Corpus d'Interaction Dialogale CID (Bertrand et al., 2008). Le premier corpus est composé d'entretiens entre un linguiste et un ou plusieurs locuteurs à propos du quartier de Paris où ils habitent ; le deuxième comprend des dialogues spontanés d'une heure chacun entre des locuteurs de la région PACA qui se connaissent déjà. Au total, 22 locuteurs (17 femmes, 5 hommes), chacun prononçant entre 1 et 10 énoncés, sont impliqués.

Afin de travailler avec des données aussi homogènes que possible, nous avons limité notre analyse aux SN repris par un clitique de 3^e personne (*il(s)/elle(s)*), constitués de 2 ou 3 syllabes seulement et commençant un nouvel énoncé prosodique déclaratif. Au total, nous avons réalisé deux codages indépendants, l'un pragmatique et l'autre prosodique, sur 101 séquences contenant un SN préverbal qui co-réfère avec un pronom clitique sujet et 101 séquences contenant un SN sans reprise. Nous émettons l'hypothèse ici que les 202 énoncés recueillis sur les corpus utilisés sont représentatifs de leur utilisation en parole spontanée.

3 Codage pragmatique

Le codage pragmatique a été fait par un des auteurs et s'est basé sur la lecture de la transcription et l'écoute des enregistrements des deux corpus. Pour l'analyse des énoncés nous avons pris en compte le

rapport entre l'énoncé et le contexte qui le précède (et parfois qui le suit). La longueur du contexte considéré était celle d'un segment cohérent de discours, celui-ci étant déterminé par des indices de segmentation du discours, comme un changement de tours de parole, un changement explicite de thème de conversation, ou bien, dans le cas des entretiens (corpus CFPP2000), une nouvelle question de l'intervieweur.

3.1 Le degré d'activation du topique

A la suite de Reinhart (1981), Lambrecht (1994), Lambrecht & Michaelis (1998), etc., nous définissons le topique phrastique comme un référent à propos duquel la phrase apporte des informations. Lambrecht & Michaelis, par exemple, affirment que le topique de phrase est « a referent which a proposition is construed to be about in a given discourse situation », où « to be about something » signifie exprimer une information qui est pertinente pour l'auditeur et qui augmente ses connaissances à propos de ce référent (Lambrecht & Michaelis, 1998: 494). Notre supposition est que les SN de notre corpus représentent des topiques de phrase dans ce sens-là.

Généralement, on parle à propos d'individus ou de choses qui sont déjà partagés entre les interlocuteurs, ou que les interlocuteurs peuvent inférer du contexte ou de leur connaissance du monde. Pourtant, un référent, même connu, peut être plus ou moins actif dans l'esprit des interlocuteurs. D'ailleurs, le degré d'activation ne dépend pas seulement de la présence d'un antécédent du référent dans le discours, ou de combien cet antécédent est éloigné du référent, mais aussi de sa fonction en tant que topique. En effet, le référent sera plus actif dans l'esprit des interlocuteurs s'il constitue déjà le topique du discours courant, alors qu'il sera inattendu s'il représente un nouveau topique par rapport au discours précédent (cf. Lambrecht, 1994). Pour cela, dans notre classification des topiques en actifs, semi-actifs et réintroduits, nous prenons en compte à la fois le statut du référent dans le discours et sa fonction en tant que topique.

3.1.1 Topique actif

Nous avons considéré un topique comme étant *actif* (topique A) lorsqu'il a été mentionné dans l'énoncé précédent où, dans la plupart des cas, il représente le topique. Un exemple est illustré par l'énoncé (4B) : le référent du topique est mentionné dans la question qui précède, qui est à propos de la grand-mère.

(4) A : votre grand-mère elle était de langue maternelle...?

B : **ma grand-mère**_i elle_i était d'un village de Piacenza...

Nous considérons toutefois qu'un topique est actif aussi quand, dans l'énoncé précédent, il est introduit comme focus et il est promu au rang de topique seulement dans l'énoncé qui suit. Un exemple est donné en (5), où le topique *ceux-là*, qui a le même référent que *beaucoup d' petits chiens*, est focus dans la construction présentative qui précède.

(5) y a beaucoup d' petits chiens là, **ceux-là**_i ils_i aboient ils font un peu d' bruit mais...

En conclusion, la définition que nous adoptons pour ce topique se chevauche en partie avec celle de *topique continu* (« continuing topic ») donnée par Frascarelli & Hinterhoelzel (2007 :88): un topique qui est « textually given and d-linked with a pre-established aboutness topic ».

3.1.2 Topique semi-actif

Quand le référent du topique n'est pas présent dans le discours, mais il est inférable à partir d'un référent antérieur qui constitue le topique du discours courant, il est considéré comme *semi-actif* (topique B) (cf. Birner & Ward, 1998, Birner, 2006). Dans la définition de Lacheret-Dujour & François, ce topique (*sub-topic*, dans leur terminologie) « sert à pointer un référent disponible en puissance par l'intermédiaire d'une référence antérieure, qui va faire l'objet à son tour d'un développement discursif. [...] (*Justine m'a prêté son cahier pour réviser, mais les pages elles sont toutes tachées*). » (Lacheret-Dujour & François,

2004 :26). Dans l'exemple (6) ci-dessous, le topique *son père* est identifiable grâce au référent *Laurence*, qui est mentionné dans l'énoncé précédent.

(6) A : y a Laurence qui était métisse

B : oui puis attends surtout **son père**_i il_i venait des DOM TOM

Le topique est partiellement activé par son lien avec le topique courant, qui porte en (6) sur les origines ethniques de Laurence.

Nous incluons dans ce type de topique aussi un référent qui est inférable à partir des connaissances du monde partagées par les interlocuteurs. En (7), par exemple, le référent *mes racines* est recouvrable par le fait que les gens normalement ont des racines géographiques. Le topique est partiellement activé par son lien avec le topique courant, à savoir les préférences des locutrices par rapport aux lieux où elles ont vécu.

(7) Aix-en-Provence, j'aime beaucoup mais euh **mes racines**_i elles_i sont à Grenoble.

Dans ce type de topique, le référent est donc nouveau, mais étant lié au topique du discours courant, il est facilement accessible aux interlocuteurs.

3.1.3 Topique réintroduit

Un topique de type C est un topique qui *a été* actif dans l'esprit des interlocuteurs dans un segment du discours précédent, mais il ne l'est plus au moment de l'énonciation de la phrase. Le topique est *réintroduit* après une interruption, donc il n'est pas le topique du discours courant. Il s'agit donc d'un cas de changement de topique (« topic shift », cf. Frascarelli & Hinterhoelzel, 2007, Brunetti, 2009). Bien que le référent ait un antécédent dans le discours, il n'est pas actif dans l'esprit des interlocuteurs. Au contraire, avec ce topique le locuteur veut précisément signaler qu'une rupture dans la continuité du discours s'est produite.

En (8), les locutrices A et B parlent des problèmes de A liés à l'endroit où elle habite. Le problème discuté concerne le fait que A est réveillée par les coqs à cinq heures du matin. Une cinquantaine d'échanges pendant lesquels la locutrice décrit d'autres problèmes liés à son nouvel appartement suivent, jusqu'au moment où la locutrice B pose la question en (8b), réintroduisant ainsi le topique *les coqs*, qui avait été abandonné.

(8) a. A : sinon y a le matin c'est sympa parce que je suis réveillée par les coqs ; comme je le disais à cinq heures et demi [...]

A : puisque euh y a des escaliers qui sont très pointus mais sans absolument aucune rambarde [...]

A : euh ah oui aussi le coup du du va-et-vient c'est-à-dire que quand on allume d'en bas on peut pas éclairer d'en haut [...]

A : non après pour l'eau il faut l'eau chaude n'est pas très chaude donc il faut pas utiliser l'eau froide si on veut avoir une douche à peu près correcte [...]

A : et puis voilà après la porte extérieure ne ferme pas à clef [...]

B : et tu mets combien de temps pour venir jusqu'ici [...]

b. B : et **les coqs**_i ils_i sont dans la pépinière ou...?

Dans d'autres cas, la distance entre l'antécédent et le topique est beaucoup plus courte. En (9), la locutrice commence son discours en parlant d'une vieille dame. Ensuite, elle décrit la scène de la chambre remplie de neige, et finalement elle réintroduit le référent initial (9b).

(9) a. A : et en fait y avait aussi une vieille personne qui était la mère d'une de leurs amies et un matin, parce qu'ils la voyaient pas venir donc ils s'étaient dit tiens on va aller réveiller euh mémé machin chose etc. ils ouvrent la porte et en fait la fenêtre était ouverte, parce que quelqu'un avait dû mal fermer la fenêtre la nuit, donc la chambre était

- B : ah ouais d'accord
 A : à moitié remplie de neige
 B : oh putain ouais d'accord
 b. A : et **la mémé**, elle, était dans son lit et elle dormait tranquillement il parait oui oui

Il est clair que le référent *la mémé* est bien identifiable, mais il n'est plus actif dans l'esprit de l'interlocutrice quand A prononce (9b).

On a remarqué ci-dessus que les topiques semi-actifs peuvent être des référents nouveaux, dans le sens qu'ils n'ont jamais été mentionnés auparavant dans le discours, alors que les topiques réintroduits sont des référents qui ont été mentionnés auparavant, donc ils ne sont pas nouveaux dans le discours. Ce qui rend les premiers plus actifs des derniers c'est le fait que le référent, même nouveau, est lié sémantiquement au topique du discours courant, alors qu'avec un topique réintroduit, ce qu'on veut signaler c'est précisément une rupture dans la continuité du discours.

3.2 Le contraste

En général, l'interprétation contrastive du topique implique la présence (explicite ou implicite) dans le discours de deux (ou plusieurs) alternatives qui s'opposent (*cf.* Buring, 1997). Plus précisément, nous avons considéré nos énoncés comme contrastifs dans les situations suivantes : (a) le topique est une des alternatives qui s'opposent (de façon explicite ou implicite) au niveau de l'information apportée par le prédicat ; (b) le topique est le même référent dans les deux énoncés, mais l'information apportée par le prédicat est, encore une fois, opposée; l'énoncé entier dans lequel se trouve le topique forme une paire d'alternatives avec un autre énoncé.

Le premier cas est illustré par (10), où le topique *ma sœur* de (10b) et le topique *moi* de l'énoncé précédent (10a) s'opposent par le fait qu'on prédique de la dernière qu'elle adorait (faire les courses) et de la première qu'elle était timide (d'où on infère qu'elle n'aimait pas les faire).

- (10) a. Moi j'adorais ça [faire les courses, *ndr*]
 b. ah **ma sœur** elle était plus timide plus réservée

En (11), *mon père* et *ma mère* s'opposent par le fait qu'on prédique du premier, d'avoir toujours vécu dans le quartier dont on parle et de la dernière (implicitement) de n'y avoir pas toujours vécu, mais d'y être arrivée plus tard.

- (11) C'est-à-dire que **mon père** a toujours été dans le quartier et **ma mère** était de Châteauroux.

Le deuxième cas est illustré en (12), où l'information apportée à propos du topique *les enfants* dans le premier énoncé (les enfants allaient dans la rue pour faire des courses) est opposée à l'information apportée par le deuxième prédicat (ils n'allaient pas dans la rue pour jouer).

- (12) **Les enfants** ils allaient dans la rue faire des courses etcetera mais ils jouaient pas dans la rue

Finalement, l'ex. (13) illustre le cas où deux énoncés entiers sont opposés :

- (13) Aix-en-Provence j'aime beaucoup mais **mes racines** elles sont à Grenoble.

La locutrice oppose son appréciation pour la ville d'Aix-en-Provence à celle – qu'on infère être plus grande – pour Grenoble, qui est sa ville natale.

4 Codage prosodique

L'ensemble des énoncés retenus pour l'analyse a été transcrit et aligné automatiquement en phonèmes, syllabes et mots graphiques dans le logiciel Praat (Boersma & Weenink, 2011) avec le script Easyalign (Goldman, 2011). Lors de la phase de correction manuelle des alignements, on a indiqué pour chacun des énoncés quelles étaient les syllabes associées à une disflue (hésitation, faux-départ, interruption, etc.),

afin de ne pas gêner les mesures automatiques des variations prosodiques relatives finalement conduites dans le second logiciel de traitement prosodique, Analor. Sur la base de l'appréciation des variations de paramètres classiques tels que la f_0 , la durée et la présence des pauses, ce logiciel permet de procéder à une détection automatique des proéminences accentuelles et d'attribuer ensuite à chacune des syllabes identifiées comme saillantes un score qui indique son degré de force relative (entre 0/10 et 10/10).

Les valeurs des seuils utilisés pour le traitement des exemples présentés dans les annexes de cet article ont été optimisées à partir d'un entraînement sur un corpus codé manuellement en amont. Nous ne revenons pas ici sur les détails de fonctionnement de l'algorithme, et nous permettons de renvoyer le lecteur intéressé aux travaux mentionnés en bibliographie, dans lesquels il est largement présenté (voir Avanzi et al., 2009, 2011, Avanzi, sous presse). Disons simplement que deux principes fondamentaux ont guidé la mise au point de l'algorithme (Lacheret-Dujour, 2007). Le premier principe est un **principe de quantité**, et stipule que plus le nombre de paramètres acoustiques entrant en jeu dans l'identification des proéminences est important, et plus les seuils fixés sont dépassés, plus la proéminence est perçue comme forte. Le second est un **principe de compensation**, qui rend compte de l'existence de phénomènes de compensation entre les paramètres acoustiques qui participent à la mise en vedette des syllabes. Par exemple, une syllabe qui présente un faible score pour un paramètre donné et un score élevé pour un autre paramètre pourra donner lieu à la perception d'une proéminence de même force qu'une syllabe présentant des scores moyens pour les deux mêmes paramètres.

Les degrés de force ainsi obtenus permettent d'établir une hiérarchie entre les frontières prosodiques actualisant les SN étudiés dans cette étude. A des fins purement pratiques (cf. infra §5.1.2), nous avons considéré qu'au-delà d'un score de 4/10, la syllabe était associée à une frontière prosodique de haut rang (cf. la syllabe finale du SNx2 *les gamins* de la Figure 4, associée à une montée mélodique de 7 dt et à un allongement relatif de 50% environ); qu'en deçà d'un score de 1.9, la syllabe était associée à une frontière prosodique de bas rang (cf. la syllabe finale du SNx2 *le mec* de la Figure 5, qui ne fait ni l'objet d'un allongement sur le plan perceptif, et qui n'est pas non plus porteuse d'un mouvement mélodique perceptible), et qu'avec un score compris entre 2 et 3.9, la syllabe marquait la fin d'un constituant prosodique intermédiaire (cf. le SNx2 *mon mari* de la Figure 6, assortie d'une montée mélodique de 5 dt mais qui ne fait pas l'objet d'un allongement). Dans les figures ci-dessous, les degrés de force sont indiqués dans une couche de transcription juste en dessous de la représentation du signal, les syllabes proéminentes (P) et disfluentes (z) dans une couche d'annotation spécifique, entre la transcription des syllabes et la transcription des mots graphiques :

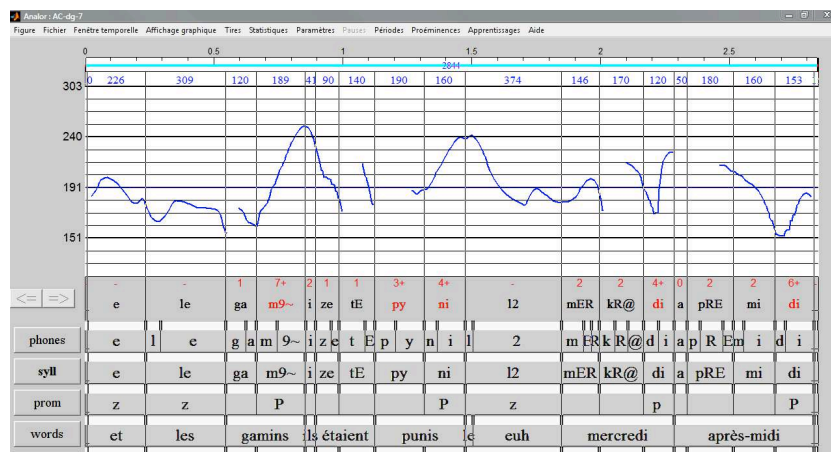


Figure 1. Copie d'écran Analor. Analyse de la séquence *et les gamins ils étaient punis euh mercredi après-midi* [CID].

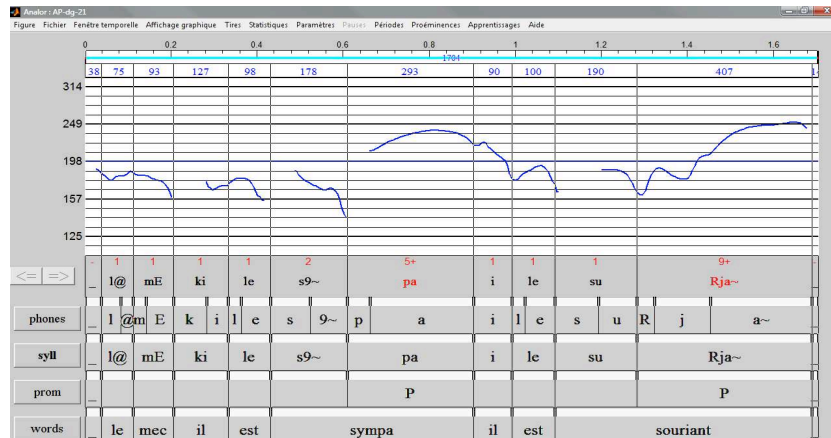


Figure 2. Copie d'écran Anamor. Analyse de la séquence *le mec il est sympa il est souriant* [CID].

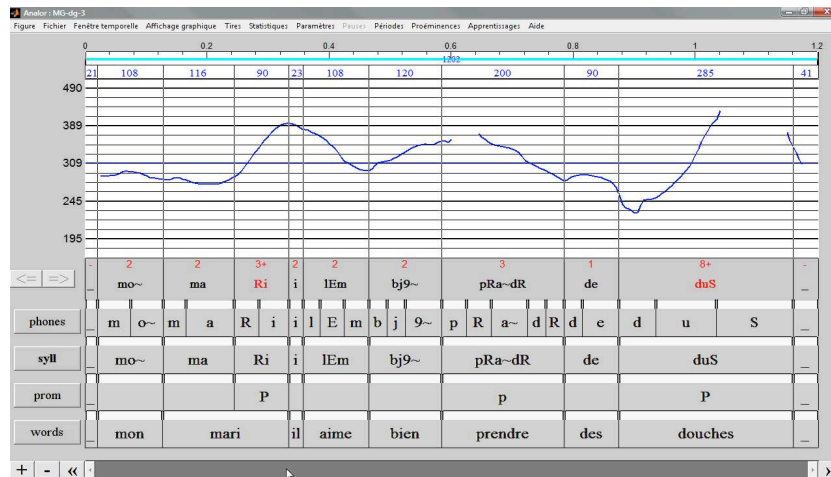


Figure 3. Copie d'écran Anamor. Analyse de la séquence *mon mari il aime bien prendre des douches* [CID].

5 Analyse et résultats

5.1 Différence entre SNx1 et SNx2 en regard de l'interface prosodie / degré d'activation

Nous avons analysé le rapport supposé dans la littérature entre le degré de *frontière prosodique* et le degré d'*activation du topique* sur les SNx2 (*ma mère; elle; est née le dix-sept février*). Les résultats du test ANOVA montrent un effet global significatif du type de topique sur le degré de la frontière prosodique ($F=7.793$ $p=0.000727***$). D'autre part, le test *post-hoc* nous donne une précision intéressante, notamment que la différence est significative seulement entre les topiques de type C et ceux de type A et B ($p=0.004$ et $p=0.0003$ respectivement), mais pas entre A et B ($p=0.94$). En d'autres mots, seulement le topique qui n'est plus actif (topique réintroduit) est significativement plus proéminent que les deux autres (voir Figure 9, graphique 1).

Successivement, nous avons étendu notre analyse aux SNx1 (*ma mère était instit*). Pour ces SN, le test ANOVA ne nous montre aucun effet significatif du degré d'activation topicale sur la proéminence accentuelle (voir Figure 9, graphique 2). On peut seulement observer que – à l'inverse des SNx2 – les

topiques de type C (réintroduits) sont *moins* proéminents que les deux autres. Cela n'empêche que l'effet des topiques réintroduits sur la proéminence accentuelle observée pour les SNx2 soit encore présent – même si de forme atténuée – quand on analyse les SN « x1 » et « x2 » rassemblés (voir Figure 9, graphique 3).

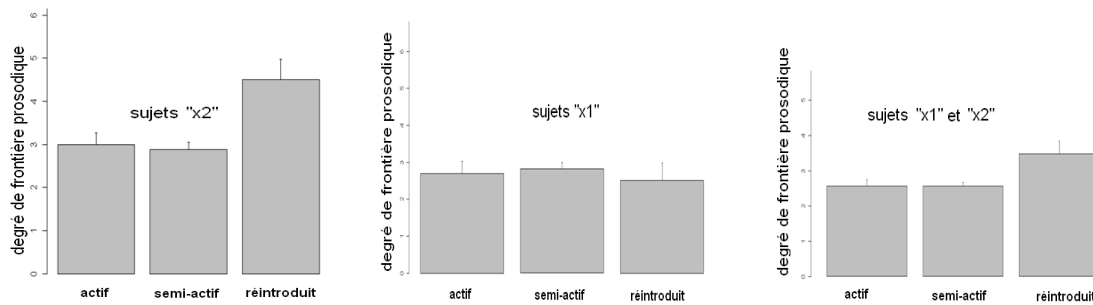


Figure 4. Degré de frontière prosodique mesuré par Analar pour les SNx1 (à gauche), les SNx2 (au milieu), puis pour les SN « x1 » et « x2 » regroupés (à droite).

Il est intéressant aussi de signaler que l'effet d'une frontière majeure avec les topiques réintroduits est beaucoup plus marqué dans les données du corpus CFPP que dans les données du corpus CID. Plus précisément, cet effet est significatif seulement dans le corpus de Paris, alors que dans le corpus CID il ne l'est jamais, ni pour les SNx1 ni pour les SNx2. La raison de cette différence pourrait simplement être le fait que les données du CID sont moins nombreuses et donc moins solides que les données du CFPP. La fiabilité moindre des données du CID pourrait aussi être due au registre plus informel de ce corpus, qui favoriserait une utilisation plus importante de la prosodie pour des raisons expressives et empêcherait donc d'en observer l'effet sur les phénomènes traités dans cet article.

5.2 Différence entre SNx1 et SNx2 en regard de l'interface prosodie / contraste

Pour ce qui concerne le rapport entre contraste, syntaxe et prosodie, nous avons d'abord constaté que l'interprétation contrastive n'est pas liée à la construction à SNx2. Les résultats montrent clairement qu'il n'y a pas une différence significative du nombre d'énoncés contrastifs entre les SNx1 et les SNx2 : parmi les 40 énoncés contrastifs que nous avons codés, 22 étaient parmi les 101 à SNx1 et 18 parmi les 101 à SNx2.

Ce qui est encore plus intéressant c'est que l'interprétation contrastive n'est associée non plus de façon significative à la proéminence accentuelle, ni pour les SNx1, ni pour les SNx2, ni pour l'ensemble des données (SNx1 + SNx2).

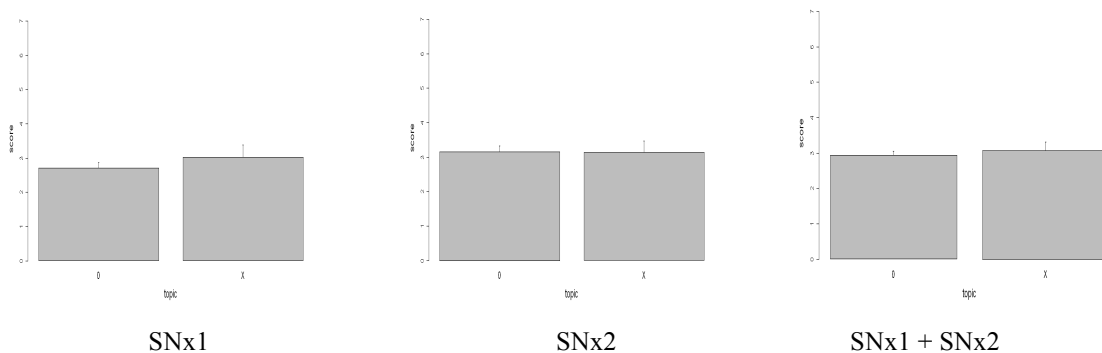


Figure 6. Rapport entre proéminence prosodique et contraste.

Si on considère seulement le corpus CFPP, on s'approche d'un effet significatif du contraste sur la proéminence (la frontière prosodique du SN est plus proéminente dans les énoncés contrastifs) avec les SNx2, mais pas avec les SNx1.

6 Discussion

Le but principal de notre travail était de vérifier l'hypothèse selon laquelle il existerait un rapport entre force de la proéminence prosodique à la frontière d'un SN repris par un clitique sujet et certaines propriétés liées au topique, notamment son degré d'activation et son interprétation contrastive. L'hypothèse concernant le degré d'activation du topique était la suivante : plus le topique est activé ou saillant, moins il est prosodiquement marqué (voir (3) *supra*). Nos résultats concernant les SN avec reprise pronominale ne nous permettent pas de confirmer entièrement cette hypothèse: la proéminence accentuelle n'augmente pas de façon *progressive* en partant des topiques de type A jusqu'à ceux de type C. Le rapport entre degré d'activation et proéminence concerne seulement le troisième type de topiques : ceux qui ne sont plus actifs dans le discours courant (topiques réintroduits). Nos résultats sont donc en accord avec ce que Grobet & Simon (2009) ont proposé dans leur étude, notamment que c'est la réactivation du topique qui est signalée prosodiquement par une proéminence forte.

Nous remarquons que les topiques de type C sont les seuls qui signalent une rupture dans la continuité du discours : les topiques actifs et les semi-actifs ne constituent pas une rupture par rapport au topique en cours. La prosodie semble donc avoir la fonction de signaler cette différence, conformément à ce qu'on a proposé dans la littérature avec des études qualitatives (voir e.g. Frascarelli & Hinterhoelzel, 2007). Avec une proéminence plus forte, le locuteur semble également indiquer qu'une *organisation particulière* des topiques dans le discours a été mise en place, à savoir que le topique courant est lié à un discours précédent, qui avait été interrompu ou abandonné et qui est maintenant repris.

La comparaison avec les SN *sans* clitique de reprise nous a permis de voir si ce rapport entre prosodie et discours était lié à la dislocation à gauche (donc aux constructions avec clitique de reprise) ou bien si la présence du clitique était un phénomène indépendant. Les résultats montrent qu'il n'y a pas de rapport significatif entre frontière prosodique et type de topique pour les SNx1. Nous avons seulement observé une tendance pour les topiques réintroduits, qui va pourtant dans le sens *inverse* à ce que nous avons vu pour les SNx2 : les topiques réintroduits ont des valeurs de frontière *plus basses* parmi les trois types (voir Figure 9, graphique 2). La conclusion que nous pouvons tirer de ces résultats est qu'il existe une tendance à la convergence des trois phénomènes suivants : présence d'un pronom de reprise, réintroduction d'un topique et frontière prosodique forte.

Le fait que le rapport entre frontière forte et topique réintroduit soit restreint aux SNx2 uniquement signifie qu'il y a une différence entre les deux constructions en français parlé contemporain. Pour autant, nous ne pouvons pas exclure que le clitique sujet ait une fonction argumentale et que le SN soit disloqué, au moins dans les énoncés où il y a coïncidence entre frontière forte du SN et réintroduction du topique.

Nos résultats ne nous permettent pas non plus de corroborer l'hypothèse selon laquelle la présence d'un clitique et celle d'une frontière prosodique forte constituent deux marquages alternatifs du topique sujet. Si cela était le cas, le marquage prosodique devrait être beaucoup plus présent dans les SN *sans* reprise et beaucoup moins présent dans les SN avec reprise. Comme on l'a vu, le résultat est plutôt l'inverse : ce sont les SN avec clitique de reprise qui sont marqués prosodiquement, quand leur fonction discursive est celle de topique réintroduit.

Quant aux topiques contrastifs, notre analyse quantitative sur corpus ne nous permet ni de valider l'hypothèse selon laquelle l'interprétation contrastive est liée à une frontière prosodique forte, ni l'hypothèse selon laquelle l'interprétation contrastive est liée à la construction à SN avec reprise. Concernant la prosodie, nos résultats contrastent avec ceux obtenus par d'autres travaux basés sur des expériences de laboratoire (voir par ex. D'Imperio et al., 2008 pour l'italien, Braun, 2004 pour l'allemand). Il faut remarquer d'autre part que les données issues d'expériences de laboratoire permettent de limiter les possibles variables en jeu. Par exemple, nos données pouvaient varier par rapport aux

facteurs suivants : la présence ou non des alternatives dans le discours; le fait que le topique se trouve dans le premier ou bien dans le deuxième énoncé d'un pair d'alternatives, et donc que le contraste soit ou non déjà explicité au moment de l'énonciation du topique (l'importance de ce dernier facteur a été souligné par Myhill & Xing, 1996 à propos de l'hébreu biblique); la présence ou non de mots ou expressions qui explicitent le contraste (*mais, d'autre part*, etc.). La quantité de données que nous avons à disposition pour le moment ne nous permet pas de tester ces différents cas séparément. Ce travail sera donc l'objet d'une recherche future. Pour ce qui concerne le rapport du contraste avec la construction à SNx2, nous pouvons remarquer que la plupart de la littérature qui a souligné le rapport entre contraste et dislocation se basait généralement sur des dislocations de l'objet. Or, ce n'est pas si évident que les deux constructions aient les mêmes caractéristiques. Une étude comparative des deux types de dislocation serait toutefois nécessaire pour vérifier cette hypothèse. Cela aussi sera l'objet d'une recherche future.

7 Conclusion

Au total, les résultats de notre analyse ne correspondent pas entièrement à l'analyse traditionnelle des constructions à SNx1 et à SNx2, où elles sont considérées comme deux constructions avec des propriétés tout à fait différentes. Nos données nous montrent que les deux constructions ne diffèrent pas clairement par rapport à leurs propriétés discursives et prosodiques. On observe une tendance des SNx2 à représenter plus fréquemment des topiques réintroduits et à être plus fréquemment marqués par une frontière prosodique majeure, mais cette différence n'est pas significative.

D'autre part, notre travail n'apporte pas des conclusions similaires aux travaux plus récents, où les différences aussi bien syntaxiques que prosodiques des deux constructions en français parlé ont été complètement remises en question. Au contraire, on a vu qu'il existe une différence entre les deux SN au moins pour une sous-partie de nos données : dans les SNx2, mais *pas* dans les SNx1, il y a un rapport significatif entre frontière prosodique forte et topique réintroduit. Cette différence entre les deux types de SN suggère que ceux-ci diffèrent aussi d'un point de vue syntaxique, notamment que le SNx2 soit disloqué en dehors de la phrase.

En somme, un marquage à la fois syntaxique et prosodique du topique de phrase, en ligne avec l'analyse traditionnelle de ces constructions, est encore présent dans nos données, mais il est limité à un seul type de topiques : les topiques *réintroduits*. Les topiques contrastifs, au contraire, ne montrent aucune relation particulière, dans nos données, ni avec les SNx2 ni avec une proéminence prosodique.

On remarque finalement que la co-présence du clitique et d'une proéminence prosodique forte se réalise seulement pour un *sous-ensemble* des topiques réintroduits analysés : d'autres ne seront pas marqués par la prosodie, ou par la syntaxe, ou ni par l'une ni par l'autre. Cette incohérence nous amène aux considérations suivantes : le marquage linguistique – aussi bien prosodique que syntaxique – des fonctions discursives apparaît moins systématique – à la lumière d'une étude *quantitative* sur corpus, donc dans la réalisation concrète de la langue – que l'on a généralement supposé en observant des phrases isolées ou en faisant des analyses de type qualitatif.

Références bibliographiques

- Apothéloz, D. (1997). Les dislocations à gauche et à droite dans la construction des schématisations. In Berrendonner, A. & Miéville, D. *Logique, discours et pensée. Mélanges offerts à Jean-Blaise Grize*. Berne : Peter Lang, 183-217.
- Avanzi, M. (2012, sous presse). *L'interface prosodie/syntaxe en français. Dislocations, incises et asyndètes*. Bruxelles : Peter Lang
- Avanzi, M., Lacheret-Dujour, A., Obin, N. & Victorri, B. (2009). Vers une modélisation continue de la structure prosodique du français : le cas des proéminences accentuelles, *Journal of French Language Studies*, 21/1, pp. 53-71.

- Avanzi, M., Obin, N., Lacheret-Dujour, A. & Victorri, B. (2011). «Toward a Continuous Modeling of French Prosodic Structure: Using Acoustic Features to Predict Prominence Location and Prominence Degree», *Proceedings of Interspeech*, Firenze, Italy.
- Bally, C. (1942). *Linguistique générale et linguistique française*. Berne : Francke.
- Berrendonner, A. (1993). Sujets zéro. In *Complétude et incomplétude dans les langues romanes et slaves. Actes du VI colloque international*, Cracovie, Sept. 1991, WSP, Cracovie, 17-45.
- Berrendonner, A. (2007). Dislocation et conjugaison en français, *Cahiers de praxématique*, 28, 85-110.
- Bertrand, R., Blache, P., Espesser, R., Ferré, R., Meunier, C., Priego-Valverde, B. & Rauzy, S. (2008). Le CID – Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues*, 49/3, 1-30.
- Birner, B. (2006). Inferential relations and noncanonical word order. In Birner, B. & Ward, G. (éds). *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*. Amsterdam: John Benjamins. 31–51
- Birner, B.J. & Ward, G. (1998). Information Status and Noncanonical Word Order in English. University of Pennsylvania / Northwestern University [*Studies in Language Companion Series*, 40]
- Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.2). www.praat.org.
- Branca-Rosoff, S. Fleury, S., Lefevre, F. & Pires, M. (2009). *Discours sur la ville. Corpus de Français Parlé Parisien des années 2000 (CFPP2000)*. <http://ed268.univ-paris3.fr/CFPP2000/>.
- Braun, B. (2004). Phonetics and Phonology of Contrast-Marking in German – Issues for the Prosody-Semantics Interface. Tagungsband Konvens.
- Brunetti, L. (2009). On links and tails in Italian. *Lingua* 119,5, 756-781.
- Brunetti, L., D’Imperio, M.P. & Cangemi, F. (2010). On the prosodic marking of contrast in Romance sentence topic: evidence from Neapolitan Italian. *Online Proceedings of Speech Prosody*, Chicago, USA.
- Buring, D. (1997). *The Meaning of Topic and Focus - The 59th Street Bridge Accent*, London : Routledge.
- Coveney, A. (2003). Le redoublement du sujet en français parlé. Une approche variationniste. In Hansen, A.B. & Mosegaard-Hansen, M.-B. (éds), *Structures linguistiques et interactionnelles dans le français parlé*. Copenhagen : Museum Tusulanum Press, 110-143.
- Culbertson, J. (2010). Convergent evidence for categorical change in French: from subject clitic to agreement marker. *Language* 86(1), 85-132.
- Culbertson, J. & Legendre, G. (2008). Qu’en est-il des clitiques sujet en français oral contemporain? In *Actes du Ier Congrès mondial de linguistique française*. Paris, France.
- De Cat, C. (2007). *French Dislocation. Interpretation, Syntax, Acquisition*. Oxford : Oxford University Press.
- Delais-Roussarie, E. Doetjes, J. & P. Sleeman (2004). Dislocation. In Corblin, F. & de Swart, H. (éds.), *Handbook of French Semantic*. Stanford : CSLI Publications, 501-528.
- Delais-Roussarie E. & Post, B. (2008). Unités prosodiques et grammaire de l’intonation : vers une nouvelle approche. *Actes des Journées d’étude sur la Parole JEP-TALN 08*.
- D’Imperio, M., Cangemi, F. & Brunetti, L. (2008). The phonetics and phonology of contrastive topic constructions in Italian”, Poster présenté au Colloque *Tone and Intonation TIE3*, Université de Lisbonne.
- D’Imperio, M. & Cangemi, F. (2009). Phrasing, register level downstep and partial topic constructions in Neapolitan Italian. *International Workshop on Intonational Phrasing in Romance & Germanic*, Université d’Hambourg.
- Fonseca-Greber, B. & Waugh, L. (2003). The subject clitics of Conversational European French. In Núñez-Cedeño, R., López, L. & Cameron, R. (éds.), *A Romance Perspective on Language Knowledge and Use, Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL)*, Chicago.
- Fradin, B. (1990). Approche des constructions à détachement. *Inventaire, Revue Romane*, 25/1, 3-34.

- Frascarelli, M. & R. Hinterhölzl (2007). Types of Topics in German and Italian. In Winkler, S. & Schwabe, K. (éds.), *On Information Structure, Meaning and Form*, (Series Editor: Werner Abraham), Amsterdam & Philadelphia, John Benjamins. 87-116.
- Gabriel, C. & Rinke, E. (2011). Structure informationnelle et statut morphosyntaxique des clitiques : évidence diachronique du dédoublement pronominal espagnol et français. In Jacob, D. & Dufter, A. (éds.), *Syntaxe, structure informationnelle et organisation du discours dans les langues romanes*. Frankfurt : Lang, 91-112.
- Goldman, J.-P. (2011). EasyAlign: an automatic phonetic alignment tool under Praat. *Proceedings of Interspeech 2011*, <http://latlcui.unige.ch/phonetique/>.
- Grobet, A. & Simon, A. C. (2009). Constructions à détachement à gauche : les fonctions de la prosodie. In Apothéloz, D., Combettes, B. & Neveu, F. (éds.), *Les linguistiques du détachement*. Berne : Peter Lang, 289-303.
- Jun, S. A. & Fougeron, C. (2002). Realization of Accentual Phrase in French. *Probus*. 14, 147-172.
- Lacheret-Dujour, A. & François, J. (2004). De la notion de détachement topical à celle de constituant thématique extrapropositionnel. *Cahiers de praxématique*, 40, 167-198.
- Lambrecht, K. (1994). *Information structure and sentence form: Topic focus, and the mental representations of discourse referents*. Cambridge : Cambridge University Press.
- Lambrecht, K. (2001) Dislocation. In : Haspelmath, M., König, E., Oesterricher, W. & Raible, W. (éds.), *Language typology and language universals: An international handbook*. Vol. 2, Berlin/New York : Walter de Gruyter, 1050-1078.
- Lambrecht, K. & Michaelis, L.A. (1998). Silence accent in information questions: default and projection. *Linguistics and Philosophy* 21, 477-544.
- Likhacheva, L. (2008). *L'articulation pragmatique du discours. Le cas des constructions détachées à gauche*. Thèse de l'Université de Toulouse.
- Marandin, J.M., Beyssade, C., Delais-Roussarie, E., Rialland, A. (2002). Discourse Marking in French: C Accents and Discourse Moves. *Proceedings of Speech Prosody 2002*. Aix-en-Provence.
- Mertens, P. (2008). Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours, *Travaux de linguistique*, 56, 97-124.
- Myhill, J. & Xing, Z. (1996). Towards an operational definition of discourse contrast. *Studies in Language* 20:303-60.
- Reinhart, T. (1981). Pragmatics and Linguistics: An Analysis of Sentence Topics, *Philosophica*, 27, 53-94.
- Riegel, M., Pellat, J.C. & Rioul, R. (2009). *Grammaire méthodique du français*, Paris : PUF.
- Rizzi, L. (1997). The fine structure of the left periphery. In: Haegeman, L. (ed), *Elements of Grammar: Handbook in Generative Syntax*. Dordrecht : Kluwer. 281-337.
- Rossi, M. (1999). *L'intonation du français : description et modélisation*. Paris/Gap : Ophrys.
- Strawson, P. (1964). Identifying reference and truth-value. *Theoria* 30: 96-118.
- Zribi-Hertz, A. (1994). The Syntax of nominative clitics in Standard and Advanced French. In Cinque, G., Koster, J. Pollock, J.-Y., Rizzi, L. & Zanuttini, R. (éds). *Paths Towards Universal Grammar*. Washington, D.C. : Georgetown University Press, 453-472.

* Nous voulons remercier les relecteurs anonymes pour leurs commentaires détaillés.

¹ Selon certains auteurs dans le cadre de la grammaire générative, le SNx2 occuperait une position syntaxique de Topique au dessus de la Phrase, en formant avec celle-ci un Syntagme Topical (« Topic Phrase » ; voir e.g. Rizzi 1997).

² Dans notre acception, une frontière prosodique majeure est une syllabe porteuse d'une proéminence forte sur le plan perceptif, qui se manifeste dans la substance par un mouvement mélodique d'une certaine amplitude, un allongement syllabique relatif, éventuellement suivie d'une pause