

LA SEGMENTATION EN UNITÉS MINIMALES EST-ELLE PROPRE AU GENRE ?

QUATRE DISCOURS ACADÉMIQUES À LA LOUPE

A. Dister

Université catholique de Louvain

RÉSUMÉ

En dépit de leur rôle crucial dans la segmentation des discours, il n'existe pas de consensus dans la littérature sur ce que sont les unités minimales du discours (MDU) et sur la manière de les identifier. Pour notre part, nous définissons les MDU par deux types de critères linguistiques qui relèvent de la syntaxe et de la prosodie (Degand et Simon, 2005). Dans cet article, nous détaillons nos critères de segmentation syntaxique et prosodique, et montrons comment ces deux niveaux coïncident ou non. Cette double segmentation nous permet de dégager trois types d'unités minimales du discours. Notre hypothèse est que les genres discursifs se distinguent selon le type de MDU mobilisé par les locuteurs. Nous avons voulu vérifier cette hypothèse sur un type particulier de discours universitaires, les discours prononcés lors de la rentrée académique.

Mots-clés : unité minimale du discours, segmentation prosodique, segmentation syntaxique.

1. INTRODUCTION

En analyse du discours, il est généralement admis que le discours est hiérarchiquement structuré (Grosz et Sidner 1986, Mann et Thompson 1988 ; Polanyi 1988 ; Roulet *et al.* 2001) : un fragment de discours est composé d'un ensemble de segments plus petits reliés les uns aux autres de manière cohérente. Par contre, ce qui diffère d'un modèle à l'autre, c'est la manière dont on définit ces segments constitutifs. Nous considérons qu'une unité minimale du discours est « the smallest interactionally relevant complete linguistic unit, in a given context, that is constructed with syntactic and prosodic resources within its semantic, pragmatic, activity-type specific, and sequential [...] context ». (Selting 2000: 477)

L'objectif de notre travail est de segmenter les énoncés en unités minimales du discours (dorénavant MDU pour *discourse minimal unit*) sur base de critères fiables et reproductibles afin d'étudier leur contribution dans la structuration des discours. Pour ce faire, seuls des critères linguistiques objectivables ont été retenus : ils concernent d'une part la syntaxe et d'autre part la prosodie (Degand et Simon 2005 ; Degand *et al.* 2007).

Dans un travail dans lequel nous analysons trois types de discours différents (Degand *et al.* soumis), nous avons pu constater que les liens entre syntaxe et prosodie varient d'un

discours à l'autre. Néanmoins, il est parfois difficile d'attribuer les différences relevées à des « genres » particuliers plutôt qu'à une variation inter-locuteurs.

Afin de valider ou d'infirmer l'hypothèse d'une différence qui relèverait du « genre », nous proposons d'analyser si, à l'intérieur d'un même type de textes, les MDU possèdent des propriétés similaires.

Parmi les nombreux discours tenus dans le cadre de l'institution universitaire, nous avons choisi de nous pencher sur des discours prononcés lors de l'évènement que constitue une rentrée académique¹. Ces discours présentent, a priori, de fortes caractéristiques convergentes. C'est le cas en ce qui concerne le mode de production. En effet, chaque texte a préalablement été écrit pour être lu (*written to be spoken*). De plus, les contextes sociaux et référentiels sont également les mêmes : ces discours sont non seulement tenus dans l'enceinte de l'université, lors d'une cérémonie très formelle, mais tous portent également sur l'université. La cérémonie a aussi la particularité de se dérouler devant un public de pairs et devant les représentants politiques.

Ainsi, on a avec nos données non seulement une convergence dans les conditions de production et d'énonciation, mais également dans le thème traité, convergence que nous n'aurions sans doute pas en choisissant des discours tenus lors de séances de remise de doctorats honoris causa ou avec des conférences ou des cours. Dans le premier cas, si le contexte est également très formel, les thèmes traités peuvent être différents ; dans le cadre des conférences, les modes de production peuvent varier puisque les locuteurs ont un degré de liberté plus ou moins grand par rapport à leur texte écrit (qui peut parfois ne pas exister), tandis que lors d'une rentrée académique, les orateurs suivent leur texte à la lettre². En choisissant les discours de la rentrée académique, nous limitons donc le nombre de facteurs qui auraient généré de la variation et rendu nos résultats plus difficilement interprétables.

Si nous posons l'hypothèse que le genre discursif se marque au travers de la segmentation en unités minimales du discours, les quatre textes devraient présenter de fortes similitudes en ce qui concerne les stratégies de segmentation, différentes néanmoins des trois genres déjà étudiés par ailleurs. Si cette hypothèse ne se vérifie pas, une autre explication pourrait se trouver dans la variation individuelle liée aux locuteurs dont les profils, en termes de « professionnels du genre », sont différents. En effet, lors de la rentrée académique, prennent la parole le recteur de l'Université, le président du conseil d'administration, le président du personnel scientifique et le président de l'assemblée générale des étudiants.

Nous allons dans un premier temps décrire nos choix d'annotation syntaxique et prosodique, avant de passer à l'analyse des données proprement dite.

2. L'ANNOTATION SYNTAXIQUE

¹ Il s'agit des quatre discours prononcés lors de la rentrée académique 2006 de l'UCL

² Nous pouvons l'affirmer avec certitude dans la mesure où les orateurs nous ont fait parvenir leur texte écrit.

Pour l'annotation syntaxique de nos données, nous nous inscrivons dans le cadre d'une grammaire de dépendance développée depuis les années 1970 par le groupe aixois de recherche en syntaxe (GARS) autour de Claire Blanche-Benveniste (Blanche-Benveniste *et al.* 1990). Il s'agit pour nous de segmenter le flux oral en séquences plus petites, que nous appelons *unités de rection* (UR). Celles-ci sont ensuite elles-mêmes segmentées en séquences fonctionnelles, selon un découpage décrit dans Bilger et Campione (2002). Les critères de découpage que nous avons développés sont reproductibles, et doivent pouvoir s'appliquer à n'importe quel énoncé de français parlé.

La segmentation syntaxique que nous effectuons est largement détaillée dans Dister *et al.* (2008a et 2008b). Elle se fait de manière totalement manuelle. Nous présentons ici brièvement le découpage en unités de rection, qui seul intervient dans notre définition de l'unité minimale du discours³.

Selon leur composition interne, nous dégageons cinq types d'unités de rection⁴ : les unités de rection complètes (urc), les unités de rection inachevées (uri), les unités de rection elliptiques (ure), les unités de rection averbales (ura) et les unités de rection « plus » (urc+).

Les unités de rection sont de taille très variable : cela peut aller d'un seul mot (le cas des impératifs par exemple) à plusieurs dizaines de mots. La longueur ne détermine pas l'appartenance d'une UR à l'un ou l'autre type.

2.1. Les unités de rection complètes (urc)

Une unité de rection complète est une unité qui apparait comme achevée, syntaxiquement et sémantiquement : les compléments obligatoires sont présents, les séquences entamées sont complètes, etc.

ces premiers pas sont prometteurs (acaCB2r)

c'est parce qu'⁵ il leur est difficile de valoriser leur expérience scientifique en dehors du monde académique que de nombreux chercheurs voient leur motivation initiale se dissiper (acaDL1r)

2.2. Les unités de rection inachevées (uri)

³ Le découpage en séquences, ou encore le codage du dispositif de la rection ou des listes, n'intervient pas dans la définition de nos unités minimales de discours. Nous renvoyons le lecteur aux références citées pour plus de détails sur ces niveaux d'analyse.

⁴ Le code qui suit chaque séquence citée correspond au code du texte tel qu'il est encodé dans notre base de données. Sauf mention explicite, tous les exemples que nous citons proviennent tous du corpus aca, analysé dans cet article.

⁵ Les « propositions subordonnées » sont intégrées dans les UR lorsqu'elles dépendent d'un verbe recteur. Elles ne constituent pas, dans ce cas, des UR autonomes mais bien des séquences régies ou associées (cf. ci-dessous), selon les cas. Les complétives dépendent d'un verbe recteur. Elles sont des compléments dans la rection d'un verbe, même si elles contiennent elles-mêmes un verbe qui peut régir des compléments. De la même manière, à ce niveau d'analyse, on n'entre pas dans la composition interne des relatives. Celles-ci sont intégrées aux UR donc elles dépendent.

Une unité de rection est inachevée soit parce qu'une des places obligatoires de la valence n'est pas instanciée, soit parce que l'un des syntagmes de l'UR (qu'il soit obligatoire ou non) est inachevé.

Dans le cas des données ici analysées, aucune unité de rection n'est inachevée. Cela s'explique aisément par le type particulier de discours concerné. Comme nous l'avons dit, les locuteurs lisent un texte, qu'ils ont sans doute maintes fois répété. Ainsi, si les disfluences ne sont pas absentes de ces discours (on a des cas de répétitions de mots grammaticaux, des achoppements dans la linéarité de l'énoncé, etc.), elles sont en tout cas nettement moins fréquentes que dans l'oral spontané, et aucun énoncé n'apparaît comme inachevé (ni dans les portions de textes que nous avons analysées pour ce travail, ni d'ailleurs dans l'ensemble des quatre discours).

2.3. Les unités de rection elliptiques (ure)

Une unité de rection elliptique est une unité de rection dans laquelle l'un des éléments est omis, sans pour autant faire de cette UR une unité inachevée.

Les URE recouvrent deux grands cas de figure : le sujet est omis ; un parallélisme de construction fait qu'un (ou plusieurs) élément(s) n'est (ne sont) pas repris dans l'URE, comme c'est le cas dans les deux exemples ci-dessous. Dans le premier cas, le sujet de *annoncent*, *ces premiers pas*, n'est pas repris après la coordination ; dans le deuxième exemple, la proposition *que si le doctorat est mieux valorisé* se rattache à deux verbes recteurs, *deviendront* et *portera ses fruits*. Nous considérons que l'unité de rection construite autour de *deviendront* est elliptique.

ces premiers pas sont prometteurs et annoncent d'autres initiatives tant en matière d'enseignement que de recherche (acaCB2r)

nous ne deviendrons une société de la connaissance le financement annoncé de la recherche ne portera ses fruits que si le doctorat est mieux valorisé (acaDL1r)

2.4. Les unités de rection averbales (ura)

À la différence d'une *urc*, le noyau d'une *ura* n'est pas un verbe conjugué mais un nom, un pronom, un adjectif, un adverbe, un verbe à l'infinitif ou encore une interjection.

mesdames |⁶ *messieurs* | *nous ne sommes pas au Liban* (acaVJ1r) → *ura* | *ura* | *urc*

(...) *nos liens avec l'Université sont à la fois forts et fragiles forts* | *forts parce que nous sommes passionnés par notre travail* (acaDL1r) → *urc+* | *ura*

2.5. Les unités de rections complètes « plus » (urc+)

Une unité de rection complète « plus » est une unité de rection complète qui contient des éléments qui ne sont pas dans la rection du verbe mais qui sont associés à celle-ci. C'est

⁶ Le symbole de la barre verticale (|) indique la frontière entre deux UR.

typiquement le cas des adverbes de phrases, qui sont associés à toute une séquence et ne sont pas sous la dépendance unique du verbe recteur. C'est aussi le cas des marqueurs du discours, des éléments disloqués ou encore de certaines subordonnées.

Cette catégorie est créée principalement pour des raisons pratiques, afin de ne pas scinder trop fréquemment les unités de la tirade. Les *urc+* correspondent à ce que l'on peut appeler une séquence maximale. Dans les deux exemples suivants, nous soulignons l'élément associé.

la plupart d'entre nous y a d'ailleurs déjà effectué un séjour de recherche (acaDL1r)

si la formation exigeante de nos étudiants est évidemment notre première raison d'être
nous contribuons bien plus largement au développement de toute la société (acaVJ1r)

3. L'ANNOTATION PROSODIQUE

L'annotation prosodique se fait de manière semi-automatisée grâce au logiciel ProsoProm (Avanzi *et al.* 2007) qui permet de détecter les syllabes proéminentes, et de découper les énoncés en unités intonatives majeures (UIM).

Toute syllabe proéminente en position finale de mot lexical (plein) est potentiellement une frontière d'UIM si elle répond aux contraintes suivantes :

- elle est allongée (1,5 fois plus longue) ;
- elle est suivie par une pause (100ms) ;
- l'on a un des 3 contours majeurs suivants : level : contour plat ; rising : contour montant (continuatif) ; falling : contour descendant (conclusif).

Certaines syllabes proéminentes finales ne répondent pas à ces deux contraintes mais nous les considérons néanmoins comme des frontières d'UIM :

- un contour « rising » particulièrement haut (entre 5 et 10 DT plus haut que les syllabes environnantes), même s'il n'est pas suivi d'une pause ;
- un allongement très marqué de la syllabe finale (plus de 2,5 fois plus longue que les syllabes environnantes), même sans pause subséquente.

4. CORRESPONDANCE ENTRE UNITÉ SYNTAXIQUE ET UNITÉ PROSODIQUE

Lorsque les deux annotations syntaxique et prosodique sont effectuées, nous les mettons en parallèle afin de les confronter. Nous avons ainsi pu dégager trois types d'unités minimales du discours, qui se distinguent du point de vue de leur structuration interne et mettent en jeu des stratégies différentes de la part des locuteurs :

- 1) *MDU canoniques*⁷ : à une unité de rection correspond une unité intonative majeure ;
- 2) *MDU condensées* : plusieurs unités de rection sont regroupées dans une seule unité intonative majeure ;
- 3) *MDU splittées* : une unité de rection est répartie sur plusieurs unités intonatives majeures distinctes.

Voici quelques exemples de chacun de ces types d'unités minimales du discours rencontrés dans le corpus :

MDU canoniques : 1 UR = 1 UIM

je m'explique (acaDD1r)
mesdames (acaDL1r)
messieurs (acaDL1r)
chers collègues (acaDL1r)
chers étudiants (acaDL1r)
mais pour autant que nous pouvons en juger elle a encore du mal à se traduire dans les faits (acaDD1r)
mais nos liens avec l'université sont aussi fragiles (acaDL1r)

MDU condensées : n UR = 1 UIM

j'approuve | et | j'apporte mon total soutien à la prise de position claire et ferme prise par notre recteur (acaVJ1r)
mesdames | messieurs | nous ne sommes pas au Liban (acaVJ1r)

MDU splittées : 1 UR = n UIM

tout professeur est supposé | être également un chercheur (acaDL1r)
mais ici comme là-bas l'université n'est pas une colline inspirée | détachée du monde (acaVJ1r)

5. LES UNITÉS MINIMALES DU DISCOURS DANS LES DISCOURS ACADÉMIQUES

Nous avons choisi d'analyser 300 secondes de quatre discours académiques, prononcés lors de la même rentrée académique. La répartition en nombre de mots, nombre d'UR, nombre de mots par UR, nombre d'UIM, nombre de mots par UIM et durée moyenne des UIM est la suivante :

⁷ Nous n'entendons pas par *canoniques* le fait que ce type soit le plus représenté dans les textes, ou qu'il représente une norme, mais qu'il s'agit du cas « non marqué ».

	nombre de mots	nombre d'UR	nombre de mots par UR	nombre d'UIM	nombre de mots par UIM	durée moyenne d'une UIM (en sec.)
acaCB1s	845	29	29,4	123	6,87	2,44
acaVJ1r	922	42	21,7	139	6,63	2,16
acaDL1r	837	44	19,8	166	5,04	1,81
acaDD1r	925	37	25,4	120	7,71	2,5
moyenne	882	38	24	137	6,56	2,19

On le voit, les unités syntaxiques sont nettement moins nombreuses que les unités intonatives majeures. Aucun des textes n'échappe à cette règle, et les tests montrent qu'il y a une relation assez sûre dans la répartition des données de ces quatre textes. Pour le dire autrement, ils présentent de fortes caractéristiques convergentes du point de vue des paramètres relevés dans ce tableau.

Les unités de rection sont dans ces données particulièrement longues, et le nombre moyen de mots graphiques⁸ par UR est donc très élevé. Ceci s'explique par le fait que le genre est celui du *written to be spoken*. Le locuteur est en fait un lecteur, qui oralise un texte écrit (même si ce texte est conçu pour être lu) et le discours est planifié comme l'est un texte écrit.

De ce point de vue-là, nos données sont à rapprocher de celles du journal parlé à la radio, où le nombre moyen de mots par unité de rection est de 17, alors qu'il est seulement de 11 dans des interviews et que ce chiffre tombe à 7 dans la conversation spontanée (cf. Degand *et al.* soumis). Par contre, en ce qui concerne la durée des UIM, les discours académiques sont à rapprocher de l'interview radiophonique puisqu'ils obtiennent une moyenne équivalente à celle-ci, comme le montre le tableau ci-dessous :

	Radio News	Interview	Conversation
nombre de mots par UR	16.75	11.07	6.82
durée moyenne d'une UIM (en sec.)	3.23	2.17	1.56

⁸ Un mot graphique est une suite de lettres séparée par un séparateur (blanc, apostrophe, trait d'union).

On voit donc déjà à ce stade de l'analyse que les discours académiques présentent entre eux de fortes convergences, et des différences notables avec les trois autres genres analysés⁹.

Lorsque nous confrontons les segmentations prosodique et syntaxique, afin de dégager les unités minimales du discours, nous obtenons les résultats suivants :

	canoniques	splittées	condensées
acaCB1s	20,69 %	79,31 %	0 %
acaVJ1r	9,52 %	78,57 %	11,90 %
acaDL1r	18,18 %	77,27 %	4,55 %
acaDD1r	29,73 %	70,27 %	0 %
moyenne	19,53 %	76,36 %	4,11 %

De manière assez frappante, la répartition des différents types de MDU converge dans les quatre textes. En effet, en moyenne, un peu plus de trois quart des unités minimales du discours sont des unités splittées, c'est-à-dire qu'une unité de rection se répartit sur plusieurs unités intonatives majeures. Cela est vrai pour les quatre textes pris en considération, acaDD1r affichant néanmoins un score légèrement inférieur.

Viennent ensuite les MDU canoniques, dans lesquelles on a une correspondance entre unité de syntaxe et unité prosodique. Ces cas concernent une MDU sur cinq dans le corpus. Les cas de MDU condensées sont quant à eux très rares, puisqu'en moyenne, on n'en a que 4 % dans les données, deux textes n'en attestant aucune.

Cette répartition décroissante des MDU, de la splittée à la condensée, se retrouve dans trois des quatre textes. Seul acaVJ1r déroge à cette tendance, les condensées étant plus fréquentes que les canoniques. Les cas rencontrés sont les passages dans lesquels le locuteur s'adresse directement au public, lors d'apostrophes. Dans l'exemple ci-dessous, les deux unités de rection averbales et l'unité de rection complète sont regroupées dans une seule unité intonative :

mesdames | messieurs | l'Université est au service de la société qui l'entoure

Notons que ce type d'apostrophe n'implique pas nécessairement un tel regroupement prosodique des unités de rection averbales. En effet, celles-ci peuvent correspondre

⁹ Ces fortes convergences à l'intérieur des quatre discours académiques sont confirmées lorsque l'on analyse la longueur moyenne des séquences ou encore le nombre de séquences par unité de rection. Ces paramètres séparent également significativement le genre « discours académique » des trois autres genres.

chacune à une unité intonatives majeure autonome et constituer des MDU canoniques, comme c'est le cas à six reprises dans le discours de acaDL1r :

mesdames | messieurs | chers collègues | chers étudiants
mesdames | messieurs

On ne peut donc pas dire que nos choix de codage, qui créent une unité de rection averbale autour de chaque nom plutôt que de regrouper tous les noms cooccurrents, vont à l'encontre d'une tendance générale qui serait corroborée par la prosodie, et que ce choix de codage a pour conséquence de créer artificiellement des MDU condensées. Les réalisations différentes de regroupement que l'on trouve dans les deux textes montrent qu'il s'agit bien d'options, de stratégies différentes choisies par les locuteurs.

6. COMPARAISON DES MDU SELON LE GENRE DISCURSIF

En ce qui concerne l'hypothèse liée au genre, comparons les résultats obtenus pour les discours académiques avec ceux de trois autres types de discours étudiés du point de vue de la structuration des unités minimales du discours est (cf. Degand *et al.*, soumis) :

	canoniques	splittées	condensées
journal radio	33,33 %	39,21 %	27,45 %
interview	35,29 %	39,21 %	23,52 %
conversation spontanée	35,29 %	5,88 %	47,05 %
moyenne	34,06 %	28,10 %	32,68 %

On le voit, la répartition des trois types de MDU est très différente de celle que l'on trouve dans les discours de la rentrée académique. Les trois genres divergent par ailleurs entre eux, notamment par une mobilisation différentes des MDU splittées et condensées. Les deux genres les plus formels (le journal et l'interview radiophoniques), ont recours aux MDU splittées, qui se produisent très peu dans la conversation spontanée. Dans celle-ci, c'est presque une MDU sur deux qui est condensée, cette stratégie étant la moins fréquente dans les deux autres genres.

Enfin, on notera que les trois genres ne se différencient pas par la proportion de MDU canoniques, qui représentent approximativement un tiers des cas du corpus.

7. LA STRATÉGIE DES MDU SPLITTÉES

On l'a vu, la stratégie la plus souvent mise en œuvre dans les discours académiques est celle qui consiste à répartir une unité de dépendance syntaxique sur plusieurs unités intonatives majeures. Cette structuration des unités minimales du discours concerne trois cas sur quatre dans nos données.

Il nous semble que trois raisons principales peuvent expliquer ce type de MDU.

La première est en lien avec la gestion de l'information : une UR est segmentée parce qu'elle contient plus d'une idée. Dans l'exemple suivant, le locuteur préfère délivrer son message en trois unités intonatives séparées, qui correspondent à trois unités d'information :

au-delà du savoir et du savoir-faire / nous attachons beaucoup d'importance au savoir être / et aux qualités humaines qu'il suppose (acaCB1r)

Certaines segmentations tiennent au besoin de mise en évidence, d'emphase, comme c'est le cas du groupe sujet de l'unité de rection suivante, détaché dans une unité intonative majeure :

nous étudiants / avons pourtant le sentiment que tout n'est pas perdu (acaDD1r)

Enfin, les MDU splittées s'expliquent également par les disfluences et les achoppements dans la planification des énoncés, où le piétinement sur l'axe syntagmatique se marque par une frontière d'UIM. Si l'on ne rencontre pas dans les discours académiques des MDU qui seraient splittées pour cette raison¹⁰, celles-ci ne sont pourtant pas absentes des genres formels, comme le montre cet exemple repris dans une interview :

je ne pense pas corriger autour de moi les / les prononciations réputées vicieuses de / de ceux qui m'entourent [interview]

Évidemment, si les raisons que l'on vient d'évoquer permettent d'expliquer les stratégies de regroupement entre unité syntaxique et unité prosodique, la forte proportion de MDU splittées tient aussi au type de discours particulier que sont les discours académiques. Ces textes appartiennent au genre de l'écrit oralisé, et les caractéristiques syntaxiques sont celle de l'écrit : des unités de rection particulièrement longues, qui se prêtent donc bien à un découpage en plusieurs unités intonatives majeures.

8. CONCLUSIONS

¹⁰ Parce que le discours est très préparé et que les rares cas d'achoppement ne donnent pas lieu à une fin d'UIM.

Les quatre discours de la rentrée académique que nous avons analysés présentent, des points de vue syntaxique et prosodique, des caractéristiques convergentes qui nous permettent de les regrouper et de les identifier comme un genre discursif, distinct des autres genres analysés que sont le journal, l'interview radiophonique et la conversation spontanée. L'analyse des unités minimales du discours permet elle aussi de distinguer ces quatre genres, puisque ceux-ci mobilisent préférentiellement des unités dont la structuration interne n'est pas la même. Les discours académiques sont à cet égard très particuliers, puisque trois quarts des MDU appartiennent à la catégorie des splittées, c'est-à-dire qu'une unité de dépendance syntaxique se répartit sur plusieurs unités intonatives majeures. L'une des raisons en est le type particulier de mode de production de ces textes : préalablement écrits pour être lus, ils ont de fortes caractéristiques de l'écrit, et les unités de la syntaxe sont particulièrement longues. Notre étude mériterait donc d'être complétée par d'autres données, qui relèveraient du même mode de production, tels que les discours politiques, les conférences, etc.

Références citées

- BILGER, M. et CAMPIONE, E. (2002), « Propositions pour un étiquetage en "séquences fonctionnelles" », *Recherches sur le français parlé 17*, Université de Provence, pp. 117-136.
- BLANCHE-BENVENISTE, Cl., BILGER, M., ROUGET, Chr. et van den EYNDE, K. (1990), *Le français parlé : études grammaticales*, Paris, Éditions du CNRS.
- DEGAND, L. et SIMON, A. C. (2005), « Minimal Discourse Units : Can we define them, and why should we ? », in AURNAGUE, M., BRAS, M., LE DRAOULEC, A. et VIEU L. (Éds), *Proceedings of SEM-05. Connectors, discourse framing and discourse structure: from corpus-based and experimental analyses to discourse theories*, Biarritz, 14-15 novembre 2005, pp. 65-74.
- DEGAND, L., DISTER, A. et SIMON, A. C. (2007), « Segmenting spoken discourse : How prosody and syntax work hand in hand when defining minimal discourse units in spoken French », communication présentée lors du colloque Ipra (International Pragmatics Association) 2007, Göteborg, juillet 2007.
- DEGAND, L., DISTER, A. et SIMON, A. C. (soumis), « Minimal discourse units in spoken French: Method and application », *Discours Journal*.
- DISTER, A., DEGAND, L. et SIMON, A. C. (2008a), *Guide de codage syntaxique du projet MDU*.
- DISTER, A., DEGAND, L. et SIMON, A. C. (2008b), « Approches syntaxiques en français parlé : vers la structuration en unités minimales du discours », *Actes du 27^e colloque international sur le lexique et la grammaires comparés*, L'Aquila, septembre 2008.
- GROSZ, B. J. et SIDNER, C. L. (1986), « Attention, intentions, and the structure of discourse », *Computational Linguistics 12*, pp. 175-204.
- MANN, W. C. et THOMPSON, S. A. (1988), « Rhetorical structure theory : Toward a functional theory of text organization », *Text 8*, pp. 243-281.
- POLANYI, L. (1988), « A formal model of the structure of discourse », *Journal of Pragmatics 12*, pp. 601-638.
- ROULET, E., FILLIETTAZ, L. et GROBET, A. (2001), *Un modèle et un instrument d'analyse de l'organisation du discours*, Bern, Peter Lang.

SELTING, M. (2000), « The construction of units in conversational talk ». *Language in Society* 29, pp. 477-517.