FDTB1: Repérage des connecteurs de discours en corpus

Jacques Steinlin, Margot Colinet, Laurence Danlos

FDTB1: Repérage des connecteurs de discours en corpus

Jacques Steinlin[1] Margot Colinet[1] Laurence Danlos[1, 2]
(1) ALPAGE, INRIA et Université Paris Diderot, 75013 Paris
(2) IUF
jacques.steinlin@gmail.com, margotcolinet@gmail.com, Laurence.Danlos@inria.fr

Résumé. Cet article présente le repérage manuel des connecteurs de discours dans le corpus FTB (French Treebank) déjà annoté pour la morpho-syntaxe. C'est la première étape de l'annotation discursive complète de ce corpus. Il s'agit de projeter sur le corpus les éléments répertoriés dans LexConn, lexique des connecteurs du français, et de filtrer les occurrences de ces éléments qui n'ont pas un emploi discursif mais par exemple un emploi d'adverbe de manière ou de préposition introduisant un complément sous-catégorisé. Plus de 10 000 connecteurs ont ainsi été repérés.

Abstract.
FDTB1 : Identification of discourse connectives in a French corpus
This paper presents the manual identification of discourse connectives in the corpus FTB (French Treebank) already annotated for morpho-syntax. This is the first step in the full discursive annotation of this corpus. The method consists in projecting on the corpus the items that are listed in LexConn, a lexicon of French connectives, and then filtering the occurrences of these elements that do not have a discursive use. More than 10K connectives have been identified.

Mots-clés : connecteurs de discours, annotation discursive de corpus, grammaire et discours.

Keywords: discourse connectives, discourse annotation, grammar and discourse.

1 Introduction

Le projet FDTB (French Discourse Treebank) s'inscrit dans la lignée du projet PDTB, Penn Discourse Treebank (Prasad *et al.*, 2008) qui a consisté à ajouter manuellement une couche d'annotation discursive sur le PTB (Penn Treebank), corpus composé d'articles du *Wall Street Journal*, déjà annoté en morpho-syntaxe. De même, le projet FDTB consiste à ajouter manuellement une couche d'annotation discursive sur le FTB, French Treebank (Abeillé *et al.*, 2003), corpus composé d'articles du journal *Le Monde* annoté en morpho-syntaxe. L'annotation complète du PDTB ou FDTB consiste *grosso modo* à repérer les connecteurs (« explicites » et « implicites » [1]), et à annoter leurs sens et leurs arguments. Des expériences préliminaires d'annotation du FDTB (Danlos *et al.*, 2012) ont montré qu'il était difficile d'effectuer toutes ces opérations en une seule passe, entre autres du fait que de nombreux items lexicaux (e.g. *et, en gros, ainsi, alors*) sont ambigus entre un emploi comme connecteur de discours et un emploi non discursif. A titre d'illustration, la conjonction de coordination *et* est connecteur en (1-a) et non-connecteur en (1-b). De même, l'adverbial *en gros* est connecteur en (2-a) et non-connecteur en (2-b).

(1) a. Fred a fini d'écrire son article et il est parti en vacances.
b. Fred et Marie sont de très bons amis.

(2) a. Paul s'est cassé le bras et a attrapé la grippe. En gros, il ne va pas bien du tout.
b. Ce film traite en gros du réchauffement climatique.

1. Un connecteur implicite n'est pas réalisé : c'est le connecteur vide entre deux phrases simplement juxtaposées dans une parataxe. A l'inverse, un connecteur explicite est un item lexical non vide.

La détermination du statut discursif de *et* dans les exemples en (1) est triviale, mais ceci est loin d'être toujours le cas, comme le montre la littérature sur *ainsi* (Molinier, 2013; Karssenberg & Lahousse, 2014) ou *alors* (Bras, 2008; Degand & Fagard, 2011). De ce fait, il est apparu qu'il valait mieux effectuer l'annotation du FDTB en commençant par une première étape, appelée FDTB1, qui consiste **uniquement** à repérer tous les connecteurs de discours du corpus. C'est cette étape que nous présentons ici. Signalons que l'annotation du PDTB n'est pas passée par cette première étape : seuls les 100 connecteurs anglais considérés comme les plus fréquents ont été annotés. Il n'est pas clair de savoir comment la fréquence des connecteurs anglais a été déterminée vu l'ambiguïté dont nous venons de parler. Seule une étude telle que celle menée dans le FDTB1 permet de déterminer la fréquence des connecteurs et d'identifier les 100 connecteurs français les plus fréquents (au moins dans un corpus journalistique).

Ce travail repose donc crucialement sur la notion de connecteur de discours qui est définie de manière fonctionnelle : les connecteurs de discours sont des items lexicaux qui permettent d'exprimer explicitement les relations discursives (sémantiques ou rhétoriques) entre deux segments de discours, « élémentaires » ou « complexes » [2]. Les connecteurs de discours du français ont été répertoriés dans LexConn (Roze *et al.*, 2012), un lexique qui recense de la manière la plus exhaustive possible les connecteurs avec leur catégorie syntaxique et la ou les relations de discours qu'ils lexicalisent. Les catégories syntaxiques sont : conjonction de coordination, conjonction de subordination, préposition (introduisant un VP à l'infinitif ou au participe présent) et adverbial (catégorie qui regroupe principalement des adverbes simples et des syntagmes prépositionnels).

Le travail effectué dans le FDTB1 s'appuie sur LexConn tant sur le plan théorique que méthodologique. Sur le plan théorique, les principes qui ont guidé l'élaboration de LexConn ont tous été retenus dans le FDTB1. Un de ces principes est qu'un segment de discours élémentaire doit comporter un syntagme verbal VP (à temps fini ou non). Ce principe a éliminé de LexConn des prépositions comme *à cause de* ou *en raison de* qui ne peuvent introduire que des syntagmes nominaux (SN). Ce principe a aussi été appliqué dans le FDTB1 : les occurrences d'éléments de LexConn qui n'ont pas portée sur un VP dans le corpus ont été éliminées automatiquement. A titre d'illustration, seules les occurrences de la préposition *pour* introduisant un VP à l'infinitif ont été projetées sur le FTB, en excluant celles introduisant un *SN*[3].

Sur le plan méthodologique, nous avons projeté automatiquement sur le FTB les éléments de LexConn respectant le principe ci-dessus, puis effectué des tâches de désambiguïsation pour savoir si ces occurrences étaient effectivement employées comme connecteurs. Les tâches de désambiguïsation sont les suivantes :

- désambiguïsation morpho-syntaxique (Section 3), par exemple pour les homonymes comme *bref* qui peut être un adjectif ou un adverbe connecteur,
- désambiguïsation entre grammaire et discours (Section 4) pour les adverbiaux (comme *ainsi* et *alors*) qui peuvent avoir un emploi comme connecteur et un emploi d'ajout à l'intérieur de leur phrase hôte,
- désambiguïsation entre grammaire et discours (Section 5) pour les prépositions et conjonctions de subordination (comme *pour* et *pour que*) qui peuvent avoir un emploi comme connecteur et un emploi d'introducteur de complément sous-catégorisé par un élément (verbal, nominal, adjectival ou adverbial) de la phrase où ils apparaissent.

Le corpus FDTB1 est librement disponible à l'adresse `https://gforge.inria.fr/frs/?group_id=6145` où se trouve un manuel d'annotation complet (Danlos *et al.*, à paraître). Les résultats quantitatifs de l'annotation sont donnés à la Section 6 qui décrit aussi les utilisations potentielles du corpus. Avant d'expliquer les tâches de désambiguïsation, nous allons préciser la notion de connecteur de discours qui est au cœur du FDTB1.

2 La notion de connecteur de discours

Nous allons préciser la notion de connecteurs de discours (explicites) en l'opposant à celle de connecteur implicite et de « AltLex ».

Lorsqu'une une phrase (typographique) ne contient aucun connecteur explicite, il est souvent considéré qu'elle est reliée à son contexte gauche par un connecteur implicite (voir note 1). Toutefois, il a été souligné dans

2. Un segment de discours est complexe s'il couvre plusieurs segments élémentaires contigus reliés eux-mêmes par des relations discursives.

3. Un tel filtrage bénéficie de l'annotation morpho-syntaxique du FTB et s'effectue automatiquement avec l'outil Tregex (Levy & Andrew, 2006).

divers travaux qu'une phrase sans connecteur explicite peut se voir relier à son contexte gauche par une relation discursive lexicalisée par des items lexicaux n'appartenant pas à la catégorie des connecteurs de discours, qui ont été baptisés AltLex (Alternative Lexicalization) dans le PDTB. Illustrons sur des exemples : en (3-a), le connecteur explicite *parce que* lie les deux propositions avec un sens causal. En (3-c), le lecteur doit inférer que les deux phrases sont reliées par une relation causale : on doit positionner un « connecteur implicite », noté ø. A l'intermédiaire, (3-b) ne comporte pas de connecteur explicite mais le lecteur ne doit faire aucune inférence : le fait que le contenu de la proposition *Fred a mal dormi* explique le contenu de *Fred est de mauvaise humeur* est explicitement indiqué par la séquence *Ceci est dû au fait que* qui se voit attribuer le statut d'AltLex.

(3) a. Fred est de mauvaise humeur parce qu' il a mal dormi.
b. Fred est de mauvaise humeur. Ceci est dû au fait qu' il a mal dormi.
c. Fred est de mauvaise humeur. ø Il a mal dormi.

Le PDTB décrit quelques cas d'AltLex pour l'anglais et les définit par le fait qu'on ne peut pas leur ajouter de connecteur sans produire un effet de redondance. Pour le français, c'est un vaste champ d'étude non exploré (à l'exception des « verbes de discours » (Danlos, 2006)), mais il nous semble qu'une définition reposant sur une absence d'inférence par le lecteur soit préférable à une définition reposant sur un effet de redondance, la redondance n'étant pas exclue de la langue et éventuellement non perçue [4].

La séquence *Ceci est dû au fait que* en (3-b) est compositionnelle et à ce titre ne saurait en aucun cas être considérée comme un connecteur de discours. En effet, un des premiers critères pour déterminer qu'une séquence composée de plusieurs mots est un connecteur est le fait qu'elle ne soit pas compositionnelle (Roze *et al.*, 2012). Considérons l'adverbial *à ce moment-là*. En (4-a), cet adverbial est compositionnel avec un sens de concomittance temporelle où le déterminant *ce . . . là* est anaphorique comme le montre la paraphrase en *au moment où il a commencé à pleuvoir*. A l'inverse, cet adverbial est non compositionnel en (5-a) où *ce . . . là* est non anaphorique. Ceci indique que seul *à ce moment-là* en (5-a) peut prétendre au statut de connecteur (lexicalisant une relation de conséquence entre les deux phrases). Ce statut est confirmé par deux autres faits : d'abord *à ce moment-là* en (5-a) ne peut pas être clivé — (5-b) est inacceptable contrairement à (4-b) — ce qui va de pair avec le fait qu'un connecteur adverbial n'est pas intégré au contenu propositionnel de sa phrase hôte contrairement à un adverbial temporel. Ensuite, *moment* en (5-a) ne peut pas être modifié — (5-c) est inacceptable contrairement à (4-c) — ce qui va de pair avec le figement versus la compositionnalité de la séquence composée.

(4) a. Il a commencé à pleuvoir. A ce moment-là, Pierre est arrivé.
b. Il a commencé à pleuvoir. C'est à ce moment-là que Pierre est arrivé.
c. Il a commencé à pleuvoir. A ce moment-là précis, Pierre est arrivé.

(5) a. Tu as l'air de penser qu'elle n'est pas honnête. A ce moment-là, tu devrais ne rien lui raconter.
b. Tu as l'air de penser qu'elle n'est pas honnête. #C'est à ce moment-là que tu devrais ne rien lui raconter.
c. Tu as l'air de penser qu'elle n'est pas honnête. #A ce moment-là précis, tu devrais ne rien lui raconter. (Roze *et al.*, 2012)

Tous les critères convergent donc pour indiquer que *à ce moment-là* en (4) est un AltLex tandis que c'est un connecteur en (5-a). Toutefois, la situation n'est pas toujours aussi tranchée : il semble exister un continuum entre AltLex et connecteur de discours, continuum qui reflète un processus de grammaticalisation (une étude dans ce sens a été menée par (Rysová & Rysová, 2014) sur le Tchèque).

En résumé, les connecteurs de discours ont pour fonction de lexicaliser les relations discursives entre deux segments de discours. Ils sont non intégrés au contenu propositionnel de leur phrase hôte, et non compositionnels et figés lorsqu'ils sont composés de plusieurs mots [5].

4. Ainsi, la requête sur Google « Cela a ensuite été suivi » avec deux marqueurs (redondants) de la relation de précédence temporelle, à savoir le connecteur *ensuite* et le verbe de discours *suivre*, ramène aux alentours de 22 800 résultats, comme le texte suivant qui n'est pas perçu comme redondant : *L'excitation a commencé vendredi après un très laconique annonce de quatre lignes par la FINMA. Cela a ensuite été suivi de certains reportages à la fois par ... ...*

5. Néanmoins, certaines conjonctions de subordination comportant le complémenteur *que* (*pour que*, *avant que*) peuvent accepter l'insertion d'un adverbial ou d'une incise : *pour, dit-on, que.*

3 Ambiguïtés morpho-syntaxiques

Le premier aspect de la désambiguation dans le FDTB1 consiste, pour chaque occurrence d'item qui peut être connecteur, à décider si elle correspond morpho-syntaxiquement à la catégorie du connecteur recherché. Le premier cas d'ambiguïté est celui des homonymes, par exemple le mot *car* qui peut-être une conjonction de coordination (répertoriée dans LexConn) ou un nom commun. Le second cas correspond à une suite de mots qui a été répertoriée comme connecteur dans LexConn mais qui peut correspondre à d'autres catégories morpho-syntaxiques. Par exemple, la suite de mots *en fait* est répertoriée dans LexConn comme adverbial (de type syntagme prépositionnel), (6-a), mais elle peut correspondre à un pronom suivi d'un verbe comme en (6-b).

(6) a. Fred avait l'air sûr de lui. En fait, il était mort de trouille.
b. La Grand-Place était piétonne. Le maire en fait un parking.

Ces deux cas d'ambiguïté peuvent être levés automatiquement grâce à l'annotation morpho-syntaxique du corpus initial. Le manuel d'annotation du FDTB1 dresse la liste d'une trentaine d'éléments de LexConn qui présentent une ambiguité morpho-syntaxique.

4 Les adverbiaux entre grammaire et discours

Le deuxième aspect de la désambiguïsation consiste à distinguer les occurrences des adverbiaux de LexConn qui ont une fonction discursive de ceux qui ont un rôle sémantique à l'intérieur de leur phrase hôte (avec la fonction syntaxique d'ajout et plus rarement de complément). Dans les termes de (Molinier & Lévrier, 2000), ceci s'approche de la distinction entre « adverbe de phrase » et autre adverbe. Cette désambiguïsation s'appuie sur les critères utilisés dans LexConn (brièvement rappelés à la Section 2). Au cas par cas, pour aider à déterminer si un adverbial potentiellement connecteur est effectivement employé comme connecteur en contexte, il a paru nécessaire de lister un emploi comme connecteur en donnant un aperçu de la ou les relations de discours qu'il lexicalise, et un emploi comme non connecteur en précisant le rôle sémantique à l'intérieur de la phrase hôte. Ce travail prolonge à large échelle celui qui a été effectué par les linguistes sur quelques connecteurs adverbiaux ; il est illustré pour *au contraire* (jamais étudié) et *ainsi* (largement étudié).

Au contraire a été annoté comme connecteur lorsqu'il lexicalise un contraste, (7-a), ou une sorte de reformulation du contexte gauche perçu comme une litote, (7-b) ; 38 occurrences. *Au contraire* n'est pas retenu comme connecteur lorsqu'il renforce une assertion négative, (8) ; 8 occurrences.

(7) a. Selon cette enquête, 15% se prononcent pour un arrêt rapide du programme nucléaire français, 22% sont au contraire favorables à sa poursuite et à la construction de nouvelles centrales.
b. Qu'il y ait aujourd'hui, ou qu'il y ait encore après le prochain comité directeur, plusieurs textes d'orientation en présence n'est pas en soi nuisible. Cela peut être au contraire une preuve de la vitalité du seul parti véritablement démocratique en France [...]

(8) La nouvelle diminution du taux d'escompte de la Banque du Japon n'a nullement déprimé la monnaie japonaise, au contraire.

Ainsi a été annoté comme connecteur lorsqu'il lexicalise une relation de résultat ou d'exemplification, comme en (9-a) sans inversion de l'ordre canonique sujet-verbe ou en (9-b) avec inversion (Molinier, 2013; Karssenberg & Lahousse, 2014) ; 291 occurrences. *Ainsi* n'est pas connecteur lorsqu'il est utilisé comme anaphore de manière, (10-a), ou comme anaphore ou cataphore d'un discours rapporté, (10-b) ; 32 occurrences.

(9) a. La Commission nationale [...] se limite à vérifier si les obligations comptables et financières sont remplies. Ainsi, il n'existe à ce jour aucun contrôle des dépenses des partis .
b. M. Hockey ne mâche pas ses mots. Ainsi a-t-il invité les pays émergents à « se sevrer de la morphine de l'argent facile et à engager des réformes ».

(10) a. Luc s'est comporté ainsi parce qu'il était fatigué.

b. M. Michel Charasse, ministre du budget, a ainsi déclaré au micro de RMC : « C'est une affaire privée, et je ne vois pas pourquoi les pouvoirs publics seraient impliqués là-dedans ».

Dans notre corpus, il y a une centaine d'adverbiaux qui sont ambigus entre grammaire et discours. Le manuel d'annotation du FDTB1 dresse la liste des adverbiaux de LexConn qui sont toujours employés comme connecteurs (au moins dans ce corpus) : ils sont une cinquantaine.

5 Les prépositions et conjonctions entre grammaire et discours

Le troisième aspect de la désambiguïsation du FDTB1 consiste à distinguer les occurrences des prépositions et conjonctions de subordination qui ont une fonction discursive de celles qui sont sous-catégorisées par un élément verbal, nominal, adjectival ou adverbial. Cette désambiguïsation concerne cinq prépositions qui introduisent des infinitives — *pour*, *afin de*, *plutôt que de*, *jusqu'à* et *avant de* — et trois conjonctions de subordination reliées morphologiquement à trois de ces prépositions, à savoir *pour que*, *afin que* et *plutôt que*. Le cas le plus complexe et le plus fréquent est celui de la préposition *pour* suivie d'une infinitive qui a fait l'objet d'une publication, (Colinet *et al.*, 2014), résumée ici dans les grandes lignes. La préposition *pour* peut être connecteur, avec une valeur finale, causale ou temporelle, (11).

(11) a. Côté alliances, DEC, qui s'est associé à Olivetti pour développer notamment des machines Risc - un microprocesseur à jeu d'instructions réduit...
b. L' an dernier, le correspondant du quotidien britannique Financial Times s'est fait expulser pour avoir fait état de « l'évaporation » des énormes bénéfices tirés des exportations de pétrole pendant la guerre du Golfe.
c. De son côté, la construction de logements reprend effectivement, après une forte baisse en 1991, pour remonter à un rythme annuel de 1,3 million de mises en chantier contre 1 million l'année précédente.

La préposition *pour* peut également introduire un complément sous-catégorisé par un verbe (12-a), un nom (12-b), un adjectif (12-c) ou encore un adverbe (12-d) (l'élément sous-catégorisant est souligné dans ces exemples).

(12) a. Le gouvernement n'a pas profité de l'occasion pour trancher.
b. Olivetti a toutes les qualités pour profiter de la nouvelle phase de croissance.
c. 280000 tonnes de céréales seront nécessaires, chaque année, pour nourrir les poules.
d. Ceci est trop rapide pour être durable.

Enfin, *pour* peut introduire une « relative sans mot QU » (Huddleston & Pullum, 2002), (13-a), et des emplois méta-discursifs, (13-b).

(13) a. Un pont pour franchir l'Amazone a été construit en 1745.
b. pour conclure, pour ne citer que lui, pour le dire autrement, ...

Si les « relatives sans mot QU » et les *pour* introducteurs d'expressions métadiscursives sont faciles à identifier, la distinction entre *pour* connecteur de discours et *pour* introduisant un argument sous-catégorisé n'est pas aisée. Il s'agit en effet d'une instance particulièrement délicate du problème général de la distinction entre arguments et modifieurs, pour laquelle une batterie de critères a été mise au point (Colinet *et al.*, 2014). Ces critères ont permis d'annoter manuellement les 1161 occurrences de *pour* introduisant une infinitive dans le FTB : 518 sont des connecteurs de discours (44%), 558 introduisent des compléments sous-catégorisés, 52 introduisent des relatives sans mot QU, et 33 introduisent des expressions métadiscursives. Ce travail a aussi permis de compléter les lexiques syntaxiques dans lesquels la préposition *pour* est largement ignorée comme introducteur de complément sous-catégorisé (Sagot *et al.*, 2014).

6 Conclusion

Les données chiffrées concernant la taille du FTB et le nombre de connecteurs annotés dans le FDTB1 avec leurs catégories sont données dans la Table 1. Les 536 occurrences de *en V-ant* correspondent à des gérondifs, voir *Fred a réconforté Marie en la complimentant sur son travail*, qui ont toutes été considérées comme connecteurs avec la particularité que le connecteur est en fait *en . . . -ant*, c'est-à-dire la préposition *en* et le suffixe-*ant*.

FTB		FDTB1	
articles	1005	adverbiaux	3221
phrases	18535	conj coord	3653
mots	535 000	conj sub	1949
		prép V-inf	1070
		en V-ant	536
		TOTAL	10429

Table 1 – Taille du FTB et nombre de connecteurs dans le FDTB1

Cette annotation a mis en évidence 29 connecteurs non répertoriés dans LexConn dont une nouvelle version (comptant 353 éléments) est disponible sur le site du FDTB1. Près de 70% des éléments de LexConn ont au moins une occurrence dans le FDTB1. Le manuel donne la liste des 100 connecteurs les plus fréquents du corpus. L'accord entre deux annotateurs experts (deux auteurs de cet article) sur un échantillon de 13 articles donne un kappa de 0,70.

Le seul autre corpus du français écrit qui a été annoté pour le discours est le corpus Annodis (Péry-Woodley *et al.*, 2011) qui est 20 fois plus petit que le FTB. Ce corpus a reçu deux annotations : annotation en relations rhétoriques et annotation en structures multi-échelles. La première correspond à l'étude de l'organisation discursive qui est étudiée dans le FDTB, même si les approches sont différentes : l'annotation en relation rhétoriques d'Annodis s'inspire de la SDRT, Segmented Discourse Representation Theory, (Asher & Lascarides, 2003), tandis que, rappelons-le, l'annotation du FDTB1 et dans le futur du FDTB s'inspire du PDTB avec un focus sur les marques lexicales (connecteurs et AltLex) des relations discursives.

Le FDTB1 est donc le premier corpus écrit où les connecteurs du discours du français sont repérés systématiquement. Ce corpus peut être utilisé par les linguistes intéressés par les connecteurs. Il peut aussi être utilisé pour développer des méthodes d'apprentissage afin de repérer automatiquement les connecteurs dans un autre corpus, et ce d'autant plus aisément qu'il repose sur une annotation morpho-syntaxique.

Pour arriver à une annotation discursive complète à partir du FDTB1, trois tâches seront à effectuer :

1. annotation du sens et des arguments des connecteurs explicites repérés dans le FDTB1,
2. identification des AltLex et des connecteurs implicites (éléments définis à la Section 2),
3. annotation du sens et des arguments des éléments identifiés à l'étape 2.

La première et la troisième tâche seront effectuées dans l'esprit du PDTB, avec quelques modifications mineures concernant la hiérarchie des sens de connecteurs et l'annotation de leurs arguments (Danlos *et al.*, 2012). La première tâche est en cours.

Références

Abeillé A., Clément L. & Toussenel F. (2003). Building a treebank for French. In A. Abeillé, Ed., *Treebanks*. Dordretch : Kluwer Academic Publishers.

Asher N. & Lascarides A. (2003). *Logics of Conversation*. Cambridge : Cambridge University Press.

Bras M. (2008). *Entre relations temporelles et relations de discours*. Université de Toulouse le Mirail : HDR.

Colinet M., Danlos L., Dargnat M. & Winterstein G. (2014). Emplois de la préposition *pour* suivie d'une infinitive : description, critères formels et annotation en corpus. In *Actes du Congrès Mondial de Linguistique Franccaise (CMLF, 2014)*, Berlin, Allemagne.

Danlos L. (2006). Discourse verbs and discourse periphrastic links. In *Proceedings of the second workshop on Constraints in Discourse (CID 2006)*, Maynooth, Ireland.

Danlos L., Antolinos-Bassos D., Braud C. & Roze C. (2012). Vers le FDTB : French Discourse Tree Bank. In *Actes de TALN 2012*, Grenoble, France.

Danlos L., Colinet M. & jacques Steinlin (à paraître). FDTB1 : Repérage des connecteurs de discours dans un corpus français. *Revue Discours*.

Degand L. & Fagard B. (2011). *Alors* between discourse and grammar : The role of syntactic position. *Functions of Language*, **18(1)**, 29–56.

Huddleston R. & Pullum G. (2002). *The Cambridge Grammar of the English Language*. Cambridge : Cambridge University Press.

Karssenberg L. & Lahousse K. (2014). *Ainsi* en tête de phrase + inversion : une analyse de corpus. *SHS Web of Conferences*, **8**, 2413–2427.

Levy R. & Andrew G. (2006). Tregex and Tsurgeon : tools for querying and manipulating tree data structures. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Gènes, Italie.

Molinier C. (2013). *Ainsi* : Deux emplois complémentaires d'un adverbe type. *Linguisticae Investigationes*, **36-2**, 311–327.

Molinier C. & Lévrier F. (2000). *Grammaire des adverbes*. Genève : Droz.

Prasad R., Dinesh N., Lee A., Miltsakaki E., Robaldo L., Joshi A. & Webber B. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrackech, Maroc.

Péry-Woodley M.-P., Afantenos S. D., Ho-Dac L.-M. & Asher N. (2011). La ressource Annodis, un corpus enrichi d'annotations discursives. *Revue TAL*, **52(3)**, 71–101.

Roze C., Danlos L. & Muller P. (2012). LexConn : a French Lexicon of Discourse connectives. *Revue Discours*.

Rysová M. & Rysová K. (2014). The centre and periphery of discourse connectives. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation*, p. 452–459, Phuket, Thailand.

Sagot B., Danlos L. & Colinet M. (2014). Sous-catégorisation en *pour* et syntaxe lexicale. In *Traitement Automatique du Langage Naturel 2014*, Marseille, France.