



HAL
open science

Étude comparative de corrélats prosodiques de marqueurs discursifs français et anglais selon leur fonction pragmatique

Lou Lee, Denis Jovet, Katarina Bartkova, Yvon Keromnes, Mathilde Dargnat

► To cite this version:

Lou Lee, Denis Jovet, Katarina Bartkova, Yvon Keromnes, Mathilde Dargnat. Étude comparative de corrélats prosodiques de marqueurs discursifs français et anglais selon leur fonction pragmatique. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition), 2020, Nancy, France. pp.335-343. hal-02798556v1

HAL Id: hal-02798556

<https://hal.science/hal-02798556v1>

Submitted on 7 Jun 2020 (v1), last revised 23 Jun 2020 (v3)

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Étude comparative de corrélats prosodiques de marqueurs discursifs français et anglais selon leur fonction pragmatique

Lou Lee^{1,2}, Denis Jovet², Katarina Bartkova¹, Yvon Keromnes¹, Mathilde Dargnat¹

(1) Université de Lorraine, CNRS, ATILF, F-54000 Nancy, France

(2) Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

{lou.lee;katarina.bartkova;yvon.keromnes;mathilde.dargnat}
@univ-lorraine.fr, denis.jovet@loria.fr

RÉSUMÉ

Ce papier présente une étude des caractéristiques prosodiques de marqueurs discursifs en fonction de leur sens pragmatique. L'étude est menée sur trois marqueurs discursifs français (*alors*, *bon*, *donc*) et trois marqueurs anglais (*now*, *so*, *well*) afin de comparer leurs caractéristiques prosodiques dans ces deux langues. Plusieurs paramètres prosodiques ont été calculés sur les marqueurs discursifs, et analysés selon les fonctions pragmatiques de ceux-ci. L'analyse a été effectuée sur plusieurs centaines d'occurrences de marqueurs discursifs extraits de corpus oraux français et anglais. Les résultats montrent que certaines fonctions pragmatiques des marqueurs discursifs amènent leurs propres caractéristiques prosodiques au niveau des pauses et des mouvements de la fréquence fondamentale. On observe également que les fonctions pragmatiques similaires partagent fréquemment des caractéristiques prosodiques similaires à travers les deux langues.

ABSTRACT

Comparative study on prosodic correlates of discourse markers in French and English according to their pragmatic function

The goal of this study is to investigate the prosodic characteristics of discourse markers according to their pragmatic meaning. The paper focusses on three French discourse markers (*alors*, *bon*, *donc*) and three English markers (*now*, *so*, *well*) in order to compare their prosodic characteristics in these two languages. Several prosodic features were calculated for the discourse markers under consideration and analysed according to the pragmatic functions. Hundreds of occurrences of discourse markers were extracted from French and English speech corpora and analysed. Results show that some pragmatic functions of discourse markers bring about their own prosodic behaviour regarding pauses and movements of the fundamental frequency. Moreover, similar pragmatic functions frequently share similar prosodic characteristics.

MOTS-CLÉS : prosodie, pragmatique, marqueurs discursifs, patterns F0

KEYWORDS: prosody, pragmatics, discourse markers, F0 patterns

1 Introduction

Les marqueurs discursifs (MD) ont été de plus en plus étudiés pendant ces dernières décennies (voir notamment Hansen, 1998 ; Schiffrin, 1987 ; Aijmer, 2002). Les marqueurs discursifs fonctionnent au niveau discursif ou situationnel et donc amènent un sens différent selon le contexte ou la situation du discours. Leur fonctionnement peut varier fortement, et pour étudier cela il faut prêter attention à leur rôle pragmatique, au-delà de leur signification littérale, ou sémantique.

La prosodie est fréquemment utilisée dans les études sur le discours ou sur la parole émotionnelle car elle permet une interprétation au-delà du sens sémantique original. Or, la prosodie des marqueurs discursifs (MD) est relativement peu étudiée, en particulier pour la langue française. La prosodie est une information très utile pour détecter les sens pragmatiques et discursifs de ces marqueurs.

Quelques travaux récents ont étudié la corrélation entre les paramètres prosodiques des marqueurs discursifs et leur usage comme MD ou non-MD ou encore leur sens pragmatique en contexte (Horne et al., 2009 ; Wichmann, 2005 ; Cabarrão et al., 2015). Selon ces études, l'information prosodique sert en effet à correctement prédire le statut du mot, c'est-à-dire, s'il est employé comme MD ou non-MD. Non seulement le statut du mot est prédictible, mais aussi le sens pragmatique plus spécifique peut aussi être marqué par la prosodie.

Dans cette étude, nous présentons une analyse prosodique sur trois marqueurs discursifs français (*alors, bon, donc*) et trois marqueurs anglais (*now, so, well*) selon leurs différentes fonctions pragmatiques. Comme nous l'avons observé dans nos études précédentes sur les marqueurs discursifs français (Bartkova et al., 2016 ; Jouvét et al., 2017 ; Lee et al., 2019), les réalisations prosodiques des marqueurs discursifs dépendent de leurs fonctions pragmatiques. La présente étude emmène nos recherches plus loin, en considérant une autre langue, l'anglais.

Nous allons d'abord présenter les données de parole utilisées et l'annotation des fonctions pragmatiques, et ensuite l'analyse prosodique : analyse de la présence de pauses avant ou après les marqueurs discursifs, et mouvements de la fréquence fondamentale (F0).

2 Corpus de parole et annotations pragmatiques

2.1 Présentation des corpus français et anglais

Les occurrences des mots étudiés, fréquemment utilisés comme marqueurs discursifs, ont été extraites aléatoirement de corpus oraux français et anglais. Les corpus français correspondent à quelques centaines d'heures de parole, et sont composés d'enregistrements de parole « préparée » provenant du corpus de l'évaluation ESTER2 (Galliano et al., 2009). Pour l'anglais, le corpus TED-LIUM 3 (Hernandez et al., 2018) a été utilisé ; il correspond à 452 heures de parole, et est composé d'enregistrement de conférences TED. Les transcriptions manuelles de ces corpus ont été alignées automatiquement avec le signal de parole pour fournir la position des mots, et pour permettre l'extraction des données de parole à étudier.

Les marqueurs discursifs étudiés sont les suivants : *alors, bon, donc* pour le français, et *now, so, well* pour l'anglais. Le choix des mots à étudier a été effectué en fonction de leur fréquence d'occurrence dans nos corpus. Le but étant d'avoir des données correspondant à un usage réel, et en quantité suffisante pour que les analyses soient pertinentes et fiables.

Pour les mots français, 200 occurrences ont été aléatoirement extraites du corpus ESTER, pour chaque mot étudié, avec un contexte large (15 mots précédant et 15 mots suivant le mot cible) pour que son sens dans le contexte soit compréhensible. Ces données ont été écoutées, et manuellement annotées, d'abord avec les étiquettes MD et non-MD, puis avec les étiquettes des fonctions pragmatiques choisies pour chaque marqueur discursif. Pour la plupart des mots, environ 70 % des occurrences ont été identifiées comme MD. Le même processus a été appliqué sur les données anglaises, mais seulement 200 occurrences ont été extraites du corpus TED-LIUM, pour chaque mot. De plus, afin que l'analyse prosodique soit la plus pertinente possible, la segmentation phonétique du mot concerné (ainsi que celle du mot précédent et du mot suivant) a été manuellement vérifiée et corrigée, si nécessaire.

2.2 Fonctions pragmatiques des marqueurs discursifs

Le poids sémantique du mot lors de son usage MD est significativement plus léger en comparaison de celui d'un usage non-MD (Maschler & Schiffrin, 2015 ; Hansen, 1998). La présence d'un marqueur discursif amène au discours un sens pragmatique, et non pas un sens littéral ou sémantique. Plusieurs sens pragmatiques peuvent être attribués à un même marqueur discursif, en fonction de son contexte d'utilisation.

Pour chaque mot étudié, les étiquettes des fonctions pragmatiques ont été choisies d'abord en fonction de la littérature (Beeching, 2007 ; Degand & Fagard, 2011 ; Denturck, 2008 ; Lefevre, 2011), puis elles ont été ajustées afin de couvrir toutes les fonctions pragmatiques rencontrées dans nos données de parole. Pour les données étudiées, chaque marqueur discursif est utilisé avec trois à sept fonctions pragmatiques différentes, et ces fonctions peuvent être partagées entre différents marqueurs discursifs (cf. Lee et al., 2019, pour plus d'informations sur le processus d'annotation et l'accord entre annotateurs).

Le tableau 1 présente le nombre d'occurrences correspondant à un usage comme MD, et la fréquence de chacune des fonctions pragmatiques pour chacun des marqueurs discursifs. À noter que la somme des pourcentages n'est pas égale à 100 %, car certains usages peu fréquents, ou des usages complexes (correspondant à la combinaison de plusieurs marqueurs discursifs, comme « ...*alors bon...* ») ne sont pas mentionnés dans le tableau.

TABLE 1. Fonctions pragmatiques des marqueurs discursifs.

	<i>alors</i>	<i>bon</i>	<i>donc</i>	<i>now</i>	<i>so</i>	<i>well</i>
<i>Nb. occurrences</i>	116	106	131	76	109	108
Introduction	45 %	20 %	17 %	33 %	13 %	13 %
Reprise	5 %		22 %	13 %	27 %	9 %
Conclusion	9 %		40 %		39 %	
Parenthèse	16 %	26 %	19 %	54 %	21 %	24 %
Hésitation	9 %	8 %				
Reformulation	2 %	9 %				5 %
Confirmation		17 %				
Clôture		8 %				
Citation						23 %
Auto-réponse						26 %

Dans la suite de cette étude, nous nous concentrons sur les quatre fonctions pragmatiques les plus fréquentes de nos données, et présentes dans les deux langues : **Introduction** (introduire un sujet ou démarrer la parole) ; **Reprise** (reprenre la parole après une rupture de parole) ; **Conclusion** (débuter, i.e. introduire une conclusion) ; et **Parenthèse** (ajouter des informations supplémentaires ou faire un détour de parole). L'exemple (1) ci-dessous correspond à la fonction Introduction et l'exemple (2) à la fonction Parenthèse du même marqueur *so*.

- (1) {need bees and they're disappearing and it's a big problem what can we do here so what i do is honeybee research i got my phd studying honeybee health} – fichier “NoahWilsonRich_2012X”, position [465.395-475.045]
- (2) {swearing and so on and then we <unk> it ended with kiss my ass and so basically he thought he was dealing with something smart and of course you know we} – fichier “SergeyBrin_2004”, position [1122.340-1135.530]

Nous focalisons l'analyse prosodique des marqueurs discursifs sur l'étude de la présence de pauses avant et/ou après le marqueur discursif, et sur le mouvement de la fréquence fondamentale (F0) par rapport au mot précédent et au mot suivant. L'analyse est menée sur les données provenant des corpus ESTER pour le français et TED-LIUM pour l'anglais qui correspondent tous les deux à de la parole préparée, et permettent donc d'effectuer des comparaisons sur le même type de parole.

2.3 Présence de pause avant et/ou après le marqueur discursif

La présence de pauses avant ou après le marqueur, et leur position (avant ou après le mot) dans le contexte immédiat ont été analysées. Pour chaque fonction pragmatique considérée, le tableau 2 précise le nombre d'occurrences des marqueurs discursifs ainsi que la distribution des pauses autour de ces marqueurs. Quatre cas sont possibles pour les pauses : pas de pause, ni avant, ni après le mot ; pause uniquement avant le mot ; pause uniquement après le mot ; et pause avant et après le mot. Pour ne pas alourdir le tableau, nous ne présentons pas les fréquences d'une pause uniquement après le mot qui est un cas très peu fréquent.

TABLE 2. Distribution des pauses autour des marqueurs discursifs selon leur fonction pragmatique. (Les cases barrées indiquent que la fonction pragmatique n'a pas été observée pour ce mot. Les pourcentages ne sont pas indiqués lorsqu'il y a moins de 10 occurrences.)

Fonction pragmatique	Position de pauses	FR - ESTER			EN - TEDLIUM		
		<i>alors</i>	<i>bon</i>	<i>donc</i>	<i>now</i>	<i>so</i>	<i>well</i>
Non-MD	<i>Nb. occurrences</i>	57	72	18	98	54	67
	Ni avant ni après	46%	75%	67%	59%	78%	70%
	Pause avant	37%	6%	6%	8%	15%	3%
	Avant & après	9%	4%	17%	2%	--	--
Introduction	<i>Nb. occurrences</i>	52	21	21	25	14	14
	Ni avant ni après	13%	29%	81%	12%	14%	29%
	Pause avant	69%	67%	10%	52%	50%	50%
	Avant & après	13%	5%	5%	32%	29%	7%
Reprise	<i>Nb. occurrences</i>	6	/	29	10	28	10
	Ni avant ni après	--	/	48%	0%	4%	0%
	Pause avant	--	/	38%	70%	61%	60%
	Avant & après	--	/	14%	30%	32%	40%
Conclusion	<i>Nb. occurrences</i>	10	0	52	/	41	/
	Ni avant ni après	40%	--	52%	/	24%	/
	Pause avant	50%	--	38%	/	61%	/
	Avant & après	10%	--	--	/	12%	/
Parenthèse	<i>Nb. occurrences</i>	19	28	25	41	23	26
	Ni avant ni après	58%	39%	92%	20%	43%	15%
	Pause avant	37%	32%	4%	54%	48%	62%
	Avant & après	5%	21%	--	24%	4%	15%

Pour les occurrences de non-MD, la plupart des cas correspondent à « pas de pause, ni avant ni après » (de 46% à 78 %) alors que celles de MD montrent plus d'occurrences avec des pauses dans le contexte immédiat. Par exemple, pour la fonction pragmatique Introduction, on note une proportion élevée de présence d'une pause avant le mot, tant en français qu'en anglais (avec une exception pour le marqueur *donc*). Pour les fonctions pragmatiques Reprise et Conclusion, en comparaison de la fonction pragmatique Introduction, la présence d'une pause avant le mot est assez similaire pour les marqueurs discursifs français, et un peu plus fréquente pour les marqueurs discursifs anglais. La présence de pause avant est un peu moins fréquente dans le cas de la fonction Parenthèse, surtout pour le français. Cela est sans doute dû à la caractéristique assez informelle de cette fonction correspondant à une interruption du flux de la parole par le locuteur lui-même pour y ajouter quelque chose.

2.4 Mouvement de la fréquence fondamentale (F0)

Les valeurs de F0 ont été calculées afin de permettre l'étude du mouvement de F0 entre le marqueur discursif et son contexte immédiat. En raison des systèmes d'accentuation différents

pour ces deux langues, les positions considérées sont les suivantes. Pour le français, on considère les valeurs de F0 sur la dernière voyelle du mot précédent (w-1), sur la dernière voyelle du marqueur discursif, et sur la dernière voyelle du mot suivant (w+1). Pour l'anglais, on considère les valeurs de F0 sur la voyelle accentuée du mot précédent (w-1), sur la voyelle du marqueur discursif (à noter que les marqueurs discursifs anglais étudiés sont tous monosyllabiques), et sur la voyelle accentuée du mot suivant (w+1).

Les valeurs calculées de F0, exprimées en semi-tons (ST), ont été mesurées par rapport à la valeur médiane de F0 de chaque locuteur (De Looze & Hirst, 2010). Cela permet une meilleure interprétation et compréhension des mouvements de F0 dans l'étendue vocale théorique qui, pour les différents locuteurs, va de la valeur médiane de F0 moins 6 ST (1/2 octave en dessous), jusqu'à la valeur médiane plus 6 ST, voire plus 12 ST (c'est-à-dire de 1/2 à 1 octave au-dessus).

La quantification vectorielle a été utilisée pour effectuer un clustering des mouvements de F0, et extraire des patterns représentatifs correspondant aux centroïdes de chaque cluster. Les patterns représentatifs ont ensuite été interprétés en termes de direction du mouvement de F0 (plateau, montant, descendant) et de niveau de F0 (haut, médian, bas). Après quelques essais empiriques, le choix de 7 clusters est apparu comme un bon compromis, permettant d'éviter des patterns quasi identiques, des écarts-types trop grands caractérisant des clusters non homogènes, ou des clusters trop petits.

Pattern de F0 dans un contexte sans pause

Sept clusters ont été créés en considérant les valeurs de F0 sur le mot précédent (w-1), sur le marqueur discursif (MD), et sur le mot suivant (w+1). Pour classer les directions des mouvements de F0 en trois catégories (plateau, montant, descendant), la différence entre deux valeurs de F0 a été considérée comme suffisamment significative quand elle atteignait au moins 3 ST (Mertens & d'Alessandro, 1995).

TABLE 3. Valeurs de F0 pour les centroïdes (moyenne \pm écart-type), direction du mouvement de F0, et niveau de F0 pour chaque cluster.

Cluster	F0 (w-1)	F0 (MD)	F0 (w+1)	Mouvement de F0	Niveau F0
Cluster 6 (92 items)	-0.64 \pm 1.29	-0.15 \pm 1.32	-0.21 \pm 1.12	plateau – plateau	médian bas
Cluster 5 (84 items)	1.62 \pm 1.42	0.79 \pm 1.40	1.81 \pm 1.78	plateau – plateau	médian haut
Cluster 2 (56 items)	-4.43 \pm 1.49	-3.05 \pm 2.25	-0.46 \pm 2.78	plateau – montant	bas
Cluster 0 (66 items)	-0.64 \pm 3.65	3.11 \pm 2.54	7.52 \pm 2.62	montant – montant	haut
Cluster 3 (26 items)	0.66 \pm 2.98	6.51 \pm 2.50	-0.23 \pm 1.66	montant – descendant	haut
Cluster 1 (60 items)	7.02 \pm 2.41	0.23 \pm 3.29	1.22 \pm 2.14	descendant – plateau	médian
Cluster 4 (50 items)	1.44 \pm 2.20	-1.16 \pm 1.85	-4.49 \pm 1.42	descendant–descendant	médian

TABLE 4. Nombre d'occurrences de chaque cluster selon les fonctions pragmatiques, et pourcentages d'items correspondant en français et en anglais. Les cas moins fréquents ont été grisés.

	Cluster 6			Cluster 5			Cluster 2			Cluster 0			Cluster 3			Cluster 1			Cluster 4		
	nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En
Parenthèse	11	16%	18%	15	24%	18%	11	12%	23%	7	2%	27%	3	4%	5%	2	24%	5%	8	16%	5%
Conclusion	16	28%	58%	3	6%	8%	6	16%	8%	7	16%	17%	1	3%	0%	6	16%	8%	5	16%	0%
Introduction	6	17%	13%	9	27%	13%	6	17%	13%	6	7%	50%	5	13%	13%	4	13%	0%	2	7%	0%

Le tableau 4 indique les patterns prosodiques observés pour 3 fonctions pragmatiques : par exemple, pour la fonction pragmatique Parenthèse, le pattern correspondant au cluster 6 (i.e., pattern « plateau-plateau » de niveau médian bas, d’après le tableau 3) est observé 11 fois ; et cela correspond à 16 % des occurrences de la fonction pragmatique Parenthèse sur les données françaises, et à 18 % des occurrences de cette fonction pragmatique sur les données anglaises. Ce tableau montre qu’une grande proportion de ces trois fonctions pragmatiques est réalisée sous forme « plateau-plateau » de niveau médian (clusters 6 et 5) ; cela correspond à un pourcentage allant de 26 % des occurrences (pour ‘Introduction’ sur les données anglaises) à 66 % des occurrences (pour ‘Conclusion’ sur les données anglaises), tant en français qu’en anglais. On trouve ensuite la forme « plateau-montant » (i.e., cluster 2) pour 8 % à 23 % des occurrences, tant en français qu’en anglais. Les autres patterns sont moins fréquents, et l’usage est plus spécifique à l’une ou l’autre des langues ; excepté la fonction pragmatique Conclusion pour laquelle on observe aussi un usage significatif du pattern « montant-montant » (cluster 0) dans les deux langues.

Pattern de F0 dans un contexte de pause avant

Le même traitement a été mené pour les marqueurs discursifs précédés par une pause. Sept clusters ont été créés, et le tableau 5 indique les valeurs des centroïdes, ainsi que les mouvements de F0, et niveaux de F0 correspondant.

TABLE 5. Valeurs des centroïdes (moyenne ± écart-type), direction du mouvement de F0, et niveau de F0 pour chaque cluster

Cluster	F0 (MD)	F0 (w+1)	Mouvement de F0	Niveau F0
Cluster 6 (138 items)	-0.18 ±1.08	0.02 ±1.06	plateau	médian
Cluster 4 (40 items)	-3.40 ±1.65	-2.53 ±1.79	plateau	bas
Cluster 0 (50 items)	7.26 ±2.03	9.84 ±1.97	montant léger	haut
Cluster 1 (140 items)	3.32 ±1.61	5.13 ±1.42	montant léger	haut médian
Cluster 2 (74 items)	-1.42 ±2.05	3.67 ±1.12	montant	médian
Cluster 3 (68 items)	-0.05 ±2.61	9.34 ±1.59	montant escarpé	médian
Cluster 5 (62 items)	3.97 ±2.28	-0.04 ±1.56	descendant	haut

Les tableaux 5 et 6 sont équivalents aux tableaux 3 et 4, mais fournissent les informations prosodiques dans le cas où une pause précède le marqueur discursif (et sans pause après). Les cas les plus fréquents correspondent aux clusters 1 (i.e., « montant léger » et niveau F0 haut), et 6 (i.e., « plateau » et niveau F0 médian). On observe que pour chaque langue une proportion significative des occurrences (18 % à 41 %) suit chacun de ces patterns. Les autres patterns de mouvements de F0 sont beaucoup moins utilisés.

TABLE 6. Nombre d’occurrences de chaque cluster selon les fonctions pragmatiques, et pourcentages d’items correspondant en français et en anglais. Les cas les moins fréquents ont été grisés

	Cluster 6			Cluster 4			Cluster 0			Cluster 1			Cluster 2			Cluster 3			Cluster 5		
	Nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En	nb	Fr	En
Introduction	17	19%	25%	4	6%	4%	9	6%	21%	20	27%	21%	11	19%	4%	16	17%	25%	3	6%	0%
Parenthèse	16	41%	18%	3	6%	4%	7	0%	14%	18	29%	27%	7	6%	12%	9	6%	16%	6	12%	8%
Conclusion	16	40%	25%	6	16%	8%	3	0%	13%	13	24%	29%	3	4%	8%	2	0%	8%	6	16%	8%

Il est possible de supposer que le type de parole des corpus analysés, ici, parole semi-préparée, joue un rôle dans le choix des formes des patterns prosodiques utilisés par les locuteurs. Aussi, nous prolongerons l'analyse en comparant les patterns utilisés en parole préparée à ceux utilisés en parole spontanée.

3 Conclusion

Le but de l'étude était l'analyse des caractéristiques prosodiques des marqueurs discursifs selon leur fonction pragmatique dans deux langues, le français et l'anglais. Les résultats montrent que certaines fonctions pragmatiques des marqueurs discursifs amènent leurs propres caractéristiques prosodiques au niveau de la présence de pauses avant ou après le mot, et au niveau des mouvements de F0. Notamment, les fonctions pragmatiques similaires partagent des caractéristiques prosodiques similaires dans les deux langues, quant à la distribution des pauses dans le contexte immédiat : les trois fonctions pragmatiques Introduction, Reprise et Conclusion sont le plus fréquemment précédées par une pause avant alors que la fonction Parenthèse est plus fréquemment prononcée sans aucune pause ni avant, ni après. Cependant, en ce qui concerne les mouvements de F0 autour des marqueurs discursifs, la plupart de ces patterns se situaient au niveau médian de F0 du locuteur ayant le plus souvent une forme de plateau, ce qui peut être expliqué par la caractéristique pragmatique même des marqueurs dont la fonction est de signaler une introduction de quelque chose de plus important que le marqueur lui-même.

Remerciements

Ce travail a été partiellement financé par le CPER LCHN (Contrat Plan Etat Région "Langues, Connaissances et Humanités Numériques").

Références

- Aijmer, K. (2002). *English discourse particles: Evidence from a corpus* (Vol. 10). John Benjamins Publishing.
- Bartkova, K., Bastien, A., & Dargnat, M. (2016). How to be a Discourse Particle?.
- Beeching, K. (2007). La co-variation des marqueurs discursifs *bon, c'est-à-dire, enfin, hein, quand même, quoi* et *si vous voulez* : une question d'identité ?. *Langue française*, (2), 78-93.
- Cabarrão, V., Moniz, H., Ferreira, J., Batista, F., Trancoso, I., Mata, A. I., & Curto, S. (2015). Prosodic classification of discourse markers. In *International Congress of Phonetic Sciences (ICPhS 2015)*. International Phonetic Association.
- De Looze, C., & Hirst, D. (2010). L'échelle OME (Octave-MEdiane) : une échelle naturelle pour la mélodie de la parole. *Proceedings of the XXVIIIème Journées d'Étude Sur La Parole (JEP 2010)*, Mons, Belgium.
- Degand, L., & Fagard, B. (2011). *Alors* between discourse and grammar: the role of syntactic position. *Functions of language*, 18(1), 29-56.
- Denturck, E. (2008). Étude des marqueurs discursifs. *L'exemple de quoi*. *Faculteit Taal-en Letterkunde, Sectie*, 2, 2007-2008.

- Galliano, S., Gravier, G., & Chaubard, L. (2009). The ESTER 2 evaluation campaign for the rich transcription of French radio broadcasts. In *Tenth Annual Conference of the International Speech Communication Association*.
- Hansen, M. B. M. (1998). *The function of discourse particles: A study with special reference to spoken standard French* (Vol. 53). John Benjamins Publishing.
- Hernandez, F., Nguyen, V., Ghannay, S., Tomashenko, N., & Estève, Y. (2018). TED-LIUM 3: twice as much data and corpus repartition for experiments on speaker adaptation. In *International Conference on Speech and Computer* (pp. 198-208). Springer, Cham.
- Horne, M., Hansson, P., Bruce, G., Frid, J., & Filipsson, M. (2009). Discourse markers and the segmentation of spontaneous speech—The case of Swedish men ‘but/and/so’. *Lund Working Papers in Linguistics*, 47, 123-139.
- Jouvet, D., Bartkova, K., Dargnat, M., & Lee, L. (2017). Analysis and automatic classification of some discourse particles on a large set of French spoken corpora. In *International Conference on Statistical Language and Speech Processing* (pp. 32-43). Springer, Cham.
- Lee, L., Bartkova, K., Jouvet, D., Dargnat, M., & Keromnes, Y. (2019). Can prosody meet pragmatics? Case of discourse particles in French.
- Lefeuvre, F. (2011). *Bon et quoi à l’oral : marqueurs d’ouverture et de fermeture d’unités syntaxiques à l’oral*. *Linx. Revue des linguistes de l’université Paris X Nanterre*, (64-65), 223-240.
- Maschler, Y., & Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. *The handbook of discourse analysis*, 1, 189-221.
- Mertens, P., & d’Alessandro, C. (1995). Pitch contour stylization using a tonal perception model. In *Proc. 13th International Congress of Phonetic Sciences* (Vol. 4, pp. 228-231).
- Schiffrin, D. (1987). *Discourse markers* (No. 5). Cambridge University Press.
- Wichmann, A. (2005). Prosody and discourse: a diachronic approach. In *Actes de la conférence Interfaces Discours Prosodie (IDP), Aix en Provence* (pp. 1-11).