



HAL
open science

Analyse contrastive interlangue des marqueurs discursifs en Français Langue Étrangère et en Français Langue Maternelle

Minerva Rojas

► **To cite this version:**

Minerva Rojas. Analyse contrastive interlangue des marqueurs discursifs en Français Langue Étrangère et en Français Langue Maternelle. 2022. hal-03844436

HAL Id: hal-03844436

<https://hal.science/hal-03844436>

Submitted on 8 Nov 2022

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.

Analyse contrastive interlangue des marqueurs discursifs en Français Langue Étrangère et en Français Langue Maternelle

Minerva Rojas

Université Côte d'Azur – minerva.ROJAS@univ-cotedazur.fr

Abstract

The aim of this paper is to analyse the use of discourse markers (DMs) in oral production by French speakers and by learners of French as a foreign language in immersion. We created a monological and dialogical oral corpus (Corpus SCFLE) to calculate the proportion of DMs used individually and in both groups in relation to the number of words used. The results show that there is a remarkable variability across the two groups and that native speakers use more DMs than learners.

Keywords: discourse markers, corpus analysis, French as a Foreign Language, contrastive interlanguage analysis.

Résumé

Cette contribution a comme but d'analyser l'utilisation des marqueurs discursifs (MD) en production orale par des locuteurs francophones et par des apprenants de français langue étrangère en immersion. Nous avons créé un corpus oral monologique et dialogique (Corpus SCFLE) afin de calculer la proportion des MD utilisés individuellement et dans l'ensemble des deux groupes par rapport au nombre des mots utilisés. Les résultats montrent qu'il existe une variabilité remarquable dans l'ensemble des deux groupes et que les natifs utilisent davantage des MD par rapport aux apprenants.

Mots clés: marqueurs discursifs, étude de corpus, Français Langue Étrangère, analyse contrastive interlangue.

1. Introduction

Les marqueurs discursifs (MD) contribuent à la dynamique communicative ; ils servent à organiser les idées du locuteur et aident l'interlocuteur à interpréter les informations reçues (Martín Zorraquino & Portolés Lázaro, 1999). Pour les apprenants d'une langue étrangère l'utilisation des MD devient aussi essentielle car une absence ou une mauvaise utilisation rend difficile la communication dans la langue cible (Wierzbicka, 2003). Dans la recherche en acquisition des langues (RAL) l'analyse des MD est en relation avec l'étude des stratégies de communication (SC) et de la fluidité énonciative car ils contribuent à garder le flot

de la production orale sans interrompre la communication (Dörnyei & Scott, 1997 ; Dörnyei & Kormos, 1998) et leur absence ou leur excès relèvent des vulnérabilités de la fluidité énonciative (Segalowitz, 2010). Ainsi, les MD seraient un élément linguistique permettant de distinguer les apprenants des natifs ainsi que les apprenants les plus performants de ceux qui ne le sont pas (Hasselgren, 2002).

2. Cadre théorique

Notre intérêt par les MD s'ancre dans l'analyse des stratégies de communication (SC) en langue étrangère. Dans ce domaine, les MD, aussi nommés remplisseurs (*fillers*), serviraient à garder le canal de communication ouvert ; ils ne seraient pas reliés au sens du message mais à la gestion de la production pouvant aussi avoir des rôles discursifs (Dörnyei et Scott, 1997 : 193). La définition proposée dans ce cadre-là rejoint celles proposées par des spécialistes en MD. D'après Chanet (2004) les MD constituent des unités non référentielles et leur absence dans la production ne suppose aucune perte d'information essentielle à la compréhension du message. Leur utilisation a plutôt une fonction sémantico-pragmatique (Rodríguez Somolinos, 2011), et le sens exprimé diffère de celui associé aux catégories grammaticales classiques qu'ils recouvrent, comme les noms, les verbes ou les adjectifs (Dostie et Pusch, 2007).

Pour leur classification, nous avons combiné la proposition de Chanet (2004) et celle de Dörnyei et Scott (1997), en identifiant en tant que MD i) les connecteurs (*donc, alors, puis, etc*) ; ii) les particules recouvrant des catégories traditionnelles distinctes mais fonctionnant comme des MD à l'oral (*ben, voilà, du coup, genre*) ; iii) les mots phatiques (*tu vois, tu sais, hein, oui, ouais*), exclus par Chanet (2004) mais pris en compte par Dörnyei et Scott (1997). Enfin, il faut signaler que les pauses vocalisées *ehh* ou *ah*, et les marqueurs régulateurs fournissant une rétro-alimentation communicative (*hm, d'accord, oui*) ont été exclus de cette classification.

Malgré l'intérêt croissant pour les MD en didactique des langues (Delahaie, 2011 ; Tran, Tutin et Cavalla, 2018), il y a très peu d'études s'intéressant à l'utilisation des MD par des apprenants de français langue étrangère (FLE) (Deng, 2016 ; Sugiyama, 2017 ; Vladimirska, Gridina et Turlă-Pastare 2021) et à la comparaison entre apprenants et locuteurs natifs (Delahaie, 2009 ; Sugiyama, 2017). Par ailleurs nous n'avons pas trouvé de références concernant l'étude de leur emploi à long terme chez des apprenants de FLE. Pour toutes ces raisons, nous nous intéressons à l'analyse longitudinale de l'utilisation des MD en FLE et à l'analyse contrastive interlangue¹ (Granger, 2015) entre locuteurs francophones et apprenants de FLE. Le but est de savoir si leur utilisation diffère

¹ À la différence de l'analyse contrastive classique (analyse de deux langues différentes) l'analyse contrastive interlangue consiste à analyser la même langue, dans notre cas le français, en différents moments de son apprentissage (Granger, 2015).

quantitative et qualitativement entre le français langue maternelle et le FLE et comment leur utilisation évolue quantitativement chez les apprenants.

3. Méthodologie

Pour l'analyse des MD nous avons exploré le corpus oral annoté de français Corpus SCFLE composé de 47.025 mots. Ce corpus a été constitué dans le cadre d'une recherche doctorale entre les années 2015-2018 à l'Université Savoie Mont Blanc (Rojas Madrazo, 2020).

3.1. Participants et tâches

Les participants (N=26) sont divisés en deux groupes. L'échantillon d'apprenants (N=12) est constitué par tous les étudiants de L.E.A. de la mention anglais-FLE² promotion 2015-2018 de l'Université Savoie Mont Blanc. Le groupe des locuteurs francophones (N=14) est composé par des étudiants de la même licence, L.E.A., mention espagnol-anglais inscrits dans la même université ; l'échantillonnage des participants a été non-aléatoire basé sur le volontariat.

Tous les participants ont effectué les mêmes tâches, dont quatre en monologue et quatre en interaction. Les tâches en monologue consistent à regarder une courte séquence vidéo et à raconter, sans temps de préparation, la scène présentée dans la vidéo ; ce type de tâches (*story retell*) est souvent utilisé dans la RAL. Quant aux tâches en interaction, elles se déroulent en binômes et consistent à parler des thèmes de communication reliés à ceux proposés par le CECRL (2001) tels que les applications mobiles, la cuisine, la télévision et les moyens de transport, déclenchées par un texte introductoire et des questions sur le thème concerné.

Le recueil des productions des apprenants a été organisé en quatre collectes de données étalées entre la première année et la troisième année de licence : novembre 2015 (t1), mars 2016 (t2), novembre 2016 (t3) et novembre 2017 (t4). Dans chacune des collectes de données, les apprenants ont effectué une tâche en monologue et une tâche en interaction. En ce qui concerne le groupe de natifs, les tâches ont été effectuées entre novembre et décembre 2015 en plusieurs séances selon la disponibilité des étudiants.

3.2. Transcription et annotation du corpus

Pour la création du corpus nous avons utilisé la suite *EXMARaLDA* (Schmidt et Wörner, 2009) qui compte sur trois logiciels indépendants : *Partitur Editor*, *Coma* et *Exakt*. Dans le *Partitur Editor*, utilisé pour transcrire les productions orales enregistrées, trois couches ou pistes de transcription ont été créées dont la principale consacrée à la transcription orthographique des productions orales suivant les normes de transcription CHAT (MacWhinney, 2000). Ainsi, les MD

² Les étudiants apprenant le FLE sont des étudiants étrangers installés en France pour effectuer leurs études supérieures en Langues Étrangères Appliquées (L.E.A.)

sont signalés dans la piste principale précédés par deux virgules (, ,*ben*).

3.3. Analyses

À partir de la collecte des données brutes (nombre de MD utilisés par chacun des locuteurs et nombre des mots de chaque production), nous avons calculé un taux individuel d'utilisation de MD. Le taux est calculé en divisant le total des MD utilisés par le nombre des mots de la production observée, multiplié par 100. Ensuite, les résultats ont été soumis à des statistiques descriptives pour mesurer les tendances centrales et la variabilité de chaque groupe (francophones et apprenants). Le test U de Mann-Whitney pour des échantillons non appariés a été lancé afin de savoir si les éventuelles différences entre les groupes sont significatives statistiquement. En outre, les dix MD les plus utilisés ont été extraits afin d'analyser leur utilisation en calculant leur taux par rapport au nombre total des mots des corpus³ et comparer les choix des deux groupes dans chacune des tâches.

Concernant l'évolution du taux des MD chez les apprenants, nous avons lancé deux tests : le test Friedman pour 4 échantillons appariés pour analyser les différences entre les quatre collectes de données ; et le test Wilcoxon pour 2 échantillons appariés pour analyser les différences entre la première et la quatrième collecte de données. Enfin, les taux individuels des MD ont été soumis à des calculs de corrélation de Spearman (r_s) pour examiner si la relation entre le taux des MD et le temps est directe ou inversement proportionnelle et si elle est significative statistiquement (par des raisons d'espace nous ne rapporterons pas ces résultats).

4. Résultats

4.1. Taux de MD pour 100 mots. Différences entre les deux groupes

Tableau 1. Taux de MD dans les deux groupes. Résultats test U de Mann-Whitney.

	apprenants n=12					natifs n=14					U- MW	W	Z	Sig. Asint. (2-q)
	M	Mdn	SD	Min	Max	M	Mdn	SD	Min	Max				
Taux MD t1 mon.	4.29	3.320	2.86	1.43	11.69	4.90	4.414	2.75	1.53	10.00	67.00	145.00	-.87	.382
Taux MD t2 mon.	3.31	1.894	3.15	1.25	10.48	5.60	5.175	2.67	1.19	10.83	42.00	120.00	-2.16	.031*
Taux MD t3 mon.	3.96	3.267	2.28	1.58	10.23	5.64	5.861	2.11	2.08	8.70	43.00	121.00	-2.11	.035*
Taux MD t4 mon.	3.38	3.465	1.63	.91	7.41	4.59	3.986	2.09	2.21	9.21	47.50	125.50	-1.88	.060
Taux MD t1 int.	3.86	3.715	1.76	1.14	7.33	5.28	4.905	3.23	.89	11.25	64.00	142.00	-1.03	.304
Taux MD t2 int.	4.14	3.537	1.66	1.71	6.88	5.14	4.508	2.25	2.10	10.78	58.00	136.00	-1.34	.181
Taux MD t3 int.	3.33	2.750	1.87	.92	6.74	6.22	4.135	4.46	2.83	17.65	45.00	123.00	-2.01	.045*
Taux MD t4 int.	4.34	3.985	2.09	1.18	7.29	5.10	4.898	2.24	1.05	10.60	70.00	148.00	-.72	.471

² Le nombre de mots du corpus SCFLE est distribué ainsi : le corpus des apprenants contient 22.397 mots (monologue = 6.785 ; interaction = 15.612) ; le corpus des natifs comprend 24.628 (monologue = 8.767 ; interaction = 15.861). Le temps total transcrit et annoté du corpus SCFLE est de 284 minutes dont 97,096 minutes de monologue (53,831 minutes pour les apprenants ; 43,264 minutes pour les natifs) et 186,775 minutes d'interaction (110,425 minutes pour les apprenants ; 76,350 pour les natifs).

D'après les résultats descriptifs (tableau 1), tant dans le groupe de natifs que dans le groupe d'apprenants, il existe une variabilité nette entre les participants, constatée par les valeurs de l'écart-type (*SD*) et les valeurs minimales (*Min*) et maximales (*Max*). Malgré la variabilité des résultats, les locuteurs francophones utilisent davantage les MD ; les valeurs de la moyenne (*M*) et de la médiane (*Mdn*) obtenues par les locuteurs francophones restent au-dessus des résultats du groupe d'apprenants tout au long de l'étude dans les deux types de tâche (monologue et interaction). Les résultats du test U de Mann-Whitney montrent que l'utilisation des MD en monologue est significativement différente entre les deux groupes dans les collectes de données intermédiaires ((t1) $z = -0.87$, $p > 0.05$; (t2) $z = -2.16$, $p < 0.05$; (t3) $z = -2.11$, $p < 0.05$; (t4) $z = -1.88$, $p > 0.05$). Dans la tâche en interaction, il n'existe pas de différences significatives sauf dans la deuxième collecte de données ((t1) $z = -1.03$, $p > 0.05$; (t2) $z = -1.34$, $p > 0.05$; (t3) $z = -2.01$, $p < 0.05$; (t4) $z = -0.72$, $p > 0.05$).

En ce qui concerne les taux des MD extraits, les différences constatées portent sur certains choix (figure 1) ; les natifs utilisent davantage *du coup* et *enfin*

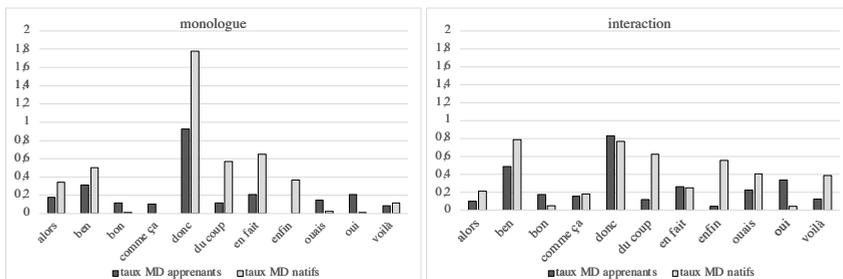


Figure 1. Différences apprenants - francophones. Taux des MD extraits pour 100 mots.

(monologue et interaction), *en fait* (monologue), *genre* et *voilà* (interaction). En revanche, les apprenants préfèrent *ouais*, *oui*, *bon*, *comme ça* (monologue) et *oui* (interaction). Des différences plus équilibrées surviennent dans les cas du MD *alors* (interaction et monologue), *en fait*, *comme ça* et *donc* (interaction). Il faut enfin signaler que *donc* est le MD privilégié par les deux groupes, mais il est plus utilisé par les locuteurs francophones en production monologique.

4.2. Taux de MD. Évolution du groupe d'apprenants

Les résultats descriptifs du taux des MD (tableau 1) montrent une diminution des valeurs de la moyenne d'un point entre le début et la fin de l'étude dans la tâche monologique. Cette diminution est constatée également dans les valeurs de l'écart-type, ce qui indiquerait que les différences entre les apprenants se resserrent à la fin de l'étude. D'ailleurs, l'étendu entre les valeurs minimales et les valeurs maximales diminue du début à la fin de l'étude à cause de la

diminution des valeurs maximales lors de la dernière collecte de données (t4). Dans la tâche en interaction, les résultats restent proches des valeurs de la tâche en monologue, mais les valeurs de la moyenne et de l'écart type augmentent légèrement du début à la fin. La variabilité entre les apprenants est moins marquée en interaction mais elle est aussi notable. D'après les résultats du test Friedman, il n'existe pas de différence significative entre le taux de MD des 4 collectes de données ni dans la tâche en monologue ($Q = 3.60, p > 0.05$) ni dans la tâche en interaction ($Q = 3.10, p > 0.05$) ; ce qui est corroboré par les résultats du test Wilcoxon tant dans la tâche en monologue ($Z = -.71, p > 0.05$) que dans la tâche en interaction ($z = -.78, p > 0.05$).

Tableau 2. Résultats test Friedman ; résultats test Wilcoxon.

Test Friedman Taux MD (t1 - t2 - t3 - t4)	N	Q	Sig. Asint. (2-queues)
Taux MD monologue	12	3.60	.308
Taux MD interaction	12	3.10	.376
Test Wilcoxon Taux MD (t1-t4)	N	Z	Sig. Asint. (2-queues)
Taux MD monologue	12	-.71	.480
Taux MD interaction	12	-.78	.433

5. Conclusion

Dans cette étude nous avons constaté que les natifs utilisent une proportion plus élevée de MD en monologue et en interaction. Pourtant les différences ne sont significatives que dans les collectes des données intermédiaires ce qui nous fait suggérer qu'elles ne sont pas en relation avec le temps d'immersion des apprenants. Nous avons également remarqué que les différences entre les apprenants et les natifs relèvent plutôt du choix des MD utilisés ; chaque groupe a des MD privilégiés et les natifs en utilisent un répertoire plus large. Quant à l'évolution du groupe d'apprenants, les résultats montrent que l'utilisation de MD varie très peu et les changements survenus ne sont pas significatifs statistiquement. Pourtant il existe une grande variabilité entre les apprenants dans la proportion de MD utilisés du début à la fin de l'étude. Il serait donc pertinent d'analyser les trajectoires individuelles et les choix des apprenants pour comprendre au mieux l'utilisation de MD à long terme et les différences entre les sujets.

Bibliographie

Castele A.V. (2011) El uso de los corpora en el análisis de los marcadores discursivos por aprendientes de ELE. In C. Hernández González, A. Carrasco Santana, E. Alvarez Ramos (dir.) *La Red y sus aplicaciones en la enseñanza-aprendizaje del español como*

- lengua extranjera*, pp. 527-534. ASELE & Universidad de Valladolid.
- Chanet C. (2004). Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie. *Recherches sur le français parlé*, 18: 83-107.
- Delahaie J. (2009). Oui, voilà ou d'accord ? Enseigner les marqueurs d'accord en classe de FLE. *Synergies pays scandinaves*, 4: 17-34.
- Delahaie J. (2011). Les marqueurs discursifs, un objet d'enseignement pertinent pour les étudiants Erasmus ?. *Éla. Études de linguistique appliquée*, 162: 153-163.
- Deng D. (2016). Oui, voilà : analyse des deux marqueurs discursifs utilisés par les locuteurs du français d'origine chinoise en France. *Cahiers AFLS* 20(1): 43-69.
- Dörnyei Z. and Scott M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language learning*, 47(1): 173-210.
- Dörnyei Z. and Kormos J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in second language acquisition*, 20(3): 349-385.
- Dostie G. and Pusch C. D. (2007). Présentation. Les marqueurs discursifs. Sens et variation. *Langue française*, 2: 3-12.
- Granger S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1): 7-24.
- Hasselgren A. (2002). Learner corpora and language testing: Smallwords as markers of learner fluency. Dans S. Granger, J. Hung & S. Petch-Tyson. *Computer learner corpora, second language acquisition and foreign language teaching*, pp. 143-174. John Benjamins.
- Martín Zorraquino M.A. and Portolés Lázaro J. (1999). Los marcadores del discurso. *Gramática Descriptiva de la Lengua Española. Tercera parte. Entre la oración y el discurso*. Espasa Calpe, pp. 4051-4213.
- Rojas Madrazo M. (2020) *Stratégies de communication, fluidité et lexique en production orale. Étude longitudinale d'apprenants en immersion*. Thèse de doctorat. Université Savoie Mont Blanc.
- Segalowitz N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Schmidt T. and Wörner K. (2009). EXMARaLDA—Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19(4): 565-582.
- Sugiyama K. (2017). Différences dans l'utilisation des marqueurs discursifs : analyse comparative entre apprenants et Français natifs. In H. Tyne, *Le Français en contextes*. [En ligne]. Presses universitaires de Perpignan. doi :10.4000/books.pupvd.2880
- Tran T.T.H., Tutin A. and Cavalla C. (2016). Pour un enseignement systématique des marqueurs discursifs à l'aide de corpus en classe de FLE: l'exemple des marqueurs de reformulation. *Linguistik Online*, 78(4): 113-128.
- Vladimirska E., Gridina J. and Turlă-Pastare D. (2021). Les marqueurs discursifs du français : un regard croisé sur une catégorie controversée. *Kalbotyra*, 74: 268–285.
- Wierzbicka A. (2003) *Cross-cultural pragmatics: the semantics of human interaction*, Mouton de Gruyter.