



HAL
open science

Les marqueurs discursifs : étude comparative de leur utilisation et fréquence chez les locuteurs francophones et dans les manuels de français langue étrangère

Simona Ruggia, Minerva Rojas

► To cite this version:

Simona Ruggia, Minerva Rojas. Les marqueurs discursifs : étude comparative de leur utilisation et fréquence chez les locuteurs francophones et dans les manuels de français langue étrangère. 2022. hal-03844456

HAL Id: hal-03844456

<https://hal.science/hal-03844456>

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.

Les marqueurs discursifs : étude comparative de leur utilisation et fréquence chez les locuteurs francophones et dans les manuels de français langue étrangère

Simona Ruggia, Minerva Rojas

¹Université Côte d'Azur – Simona.RUGGIA@univ-cotedazur.fr
Minerva.ROJAS@univ-cotedazur.fr

Abstract

This study examines the use of several discourse markers (DMs) in oral interactions between French native speakers and those used in textbooks for teaching French as a foreign language (FFL). Two corpora (corpus Manuels FLE and corpus SCFLE) are examined to explore possible differences in the frequency of occurrence of a selection of DMs and to calculate a rate of marker use. The results show that the distribution of DMs differs between the two corpora and that their proportion is higher in spontaneous French than in textbook interactions.

Keywords: discourse markers, oral interactions, French native speakers, French as a foreign language textbooks.

Résumé

Cette étude se penche sur l'utilisation de certains marqueurs discursifs (MD) lors d'interactions orales entre locuteurs natifs de langue française et de celles exploitées dans les manuels destinés à enseigner le français langue étrangère (FLE). À travers l'analyse de deux corpus (corpus Manuels FLE et corpus SCFLE), nous explorons les éventuelles différences en repérant la fréquence d'occurrence d'une série de MD et en calculant leur taux d'utilisation. Les résultats montrent que la distribution des MD diffère entre les deux corpus et que leur proportion est plus élevée chez les locuteurs natifs.

Mots clés : marqueurs discursifs, interactions orales, locuteurs francophones, manuels de français langue étrangère.

1. Introduction

Les marqueurs discursifs (MD) du français parlé ont suscité un intérêt remarquable à partir des années 1970 en donnant lieu à de nombreuses descriptions, classifications et terminologies issues de diverses approches linguistiques. Il convient donc de préciser que nous utiliserons le terme de MD

« dans un sens large, en tant que classe hétérogène » (Rodriguez Somolinos, 2011 : 6) comprenant des unités lexicales normalement utilisées à l'oral (Dostie et Pusch, 2007), qui ne sont pas nécessairement des unités référentielles car leur absence ne suppose pas une perte d'information (Chamet, 2004) mais qui fournissent des pistes à l'interlocuteur sur la façon dont le locuteur construit son discours et connecte les contenus (Bazzanella, 2006).

Dans le cadre de notre étude, nous nous intéressons aux MD d'un point de vue linguistico-didactique, afin de comparer leur utilisation et fréquence dans deux corpus numériques composés d'interactions orales de nature différente. D'une part, un corpus constitué d'interactions (dialogues et polylogues) proposées sous forme d'enregistrements et de textes dans plusieurs manuels de français langue étrangère (FLE) couvrant tous les niveaux¹ du *Cadre Européen Commun de Référence pour les langues* (CECRL) (Conseil de l'Europe, 2001). D'autre part, un corpus comprenant des interactions (dialogues) de locuteurs francophones. Pour ce faire, nous allons nous pencher sur l'emploi des MD à travers l'analyse quantitative et qualitative de deux corpus, afin de vérifier si les modèles présentés dans les manuels de FLE reproduisent les caractéristiques d'une interaction orale spontanée ; si les MD les plus utilisés dans les manuels sont ou non les plus employés dans les interactions spontanées ; et si la fréquence d'utilisation des MD est significativement différente dans les manuels par rapport au français parlé spontané.

2. Cadre théorique

Afin de décrire les emplois de certains MD présents dans nos deux corpus, nous nous appuyerons sur la taxonomie esquissée par Waltereit (2007 : 95). Ainsi, nous allons distinguer les MD qui « existent sous la forme d'un seul mot » (*Ib.* : 94) comme *mais, donc, alors*, etc. que nous appellerons simples en les opposant aux marqueurs « complexes » qui sont constitués de « deux [ou plusieurs] mots contigus » (*Ib.* : 94). Ces derniers correspondent aussi bien aux MD phrasèmes (Dostie, 2004) constitués de mots qui ne sont pas des marqueurs comme *tu sais, tu vois, tu sais pas quoi*, qu'aux MD composés d'autres MD comme *bon ben, enfin bref, et puis ben*, etc. N'ayant pas la place ici pour détailler le rôle des MD lors des interactions verbales, nous nous contenterons d'affirmer qu'ils fournissent un cadre interprétatif dans lequel le message doit être compris.

3. Méthodologie

L'exploration et l'analyse des corpus a été effectuée grâce à une lecture contrôlée et assistée par l'analyse statistique des données textuelles, la logométrie

¹ Le CECRL propose une classification de la compétence à communiquer langagièrement en 6 niveaux : A1, A2, B1, B2, C1 et C2. (Conseil de l'Europe, 2001: 23-38).

(Mayaffre, 2010) qui a « une dimension métrique et réflexive par rapport aux observables du texte » et « un pouvoir descriptif qui a une valeur quasi probatoire » (Ruggia, 2021 : 88).

Les corpus ont été constitués et étudiés à l'aide de deux outils différents. Le corpus Manuels FLE avec la plateforme Hyperbase web² qui permet de réaliser des bases hypertextuelles avec les textes qu'on lui fournit et combine des fonctions documentaires et statistiques pour décrire, caractériser, classer et interpréter les textes. La tokenisation, l'étiquetage morpho-syntaxique et la lemmatisation du corpus ont été effectués avec Spacy. Le corpus SCFLE a été constitué et analysé à l'aide de la suite de logiciels EXMARaLDA (Schmidt & Wörner, 2009) et CLAN (MacWhinney, 2000) ; la recherche de concordances sur EXMARaLDA-Exakt est facilitée par l'annotation de surface des MD et l'utilisation des expressions régulières (ReGex).

3.1. Les corpus

Les deux corpus numériques sont échantillonnés et constitués de transcriptions d'interactions orales en langue française de nature et de taille différentes.

Le corpus Manuels FLE³, dont l'étendue est de 446.350 mots, contient les dialogues et les polylogues de divers ensembles pédagogiques de FLE⁴. Ces interactions, qui fournissent un modèle de la langue cible aux apprenants allophones, sont organisés en 6 classes correspondant aux 6 niveaux du CECRL. Le choix du corpus répond à plusieurs critères : les textes sont issus de manuels qui s'inscrivent dans l'approche actionnelle du CECRL, manuels de méthodes couvrant plusieurs niveaux, manuels de méthodes récentes, ouvrages conçus pour la préparation aux diplômes du DELF et DALF⁵. Ce dernier critère a été ajouté puisque la plupart des méthodes actuellement sur le marché ne couvrent pas tous les niveaux.

Le corpus SCFLE⁶, qui comporte 15.861 mots, est constitué de dialogues entre

² Hyperbase a vu le jour en 1989 sous forme de logiciel grâce à l'œuvre d'E. Brunet au sein du laboratoire « Bases, Corpus, Langage » et n'a pas cessé d'évoluer. Depuis 2015, une version web a été développée par L. Vanni : <http://hyperbase.unice.fr/>.

³ Pour cette étude nous n'avons exploité qu'une partie du corpus des manuels FLE accessible sur Hyperbase web qui contient aussi des monologues et dont la taille est de 614.025 mots. Ce corpus a été établi dans le cadre de nos recherches sur l'évaluation de(s) niveau(x) d'un texte en fonction des échelles du CECRL grâce au deeplearning (Ruggia et Vanni, 2022).

⁴ Par manque de place, nous renvoyons à Ruggia (2021) pour les références des manuels utilisés.

⁵ Diplôme d'Études de Langue Française, Diplôme Approfondi de Langue Française

⁶ Le corpus SCFLE a été réalisé dans le cadre de notre thèse de doctorat (Rojas Madrazo, 2020) afin d'analyser le développement de l'apprentissage du FLE chez des apprenants en immersion à travers d'analyses longitudinales et contrastives. Il comporte au total 47.025 mots, dont 15.861 correspondant aux interactions des natifs ; 15.612 correspondant aux interactions entre apprenants ; 8.767 correspondant à des monologues des natifs et 6.785 correspondant aux monologues des apprenants.

locuteurs natifs, étudiants de deuxième année de licence LEA (10 femmes et 4 hommes), dont l'échantillonnage a été effectué sur la base du volontariat. Les interactions portent sur des thèmes de communication proposés dans le CECRL (Conseil de l'Europe, 2001 : 45), tels que les loisirs, la nourriture et le transport. Les échanges ont été déclenchés après la lecture d'un texte bref en relation avec la thématique sélectionnée et de quelques questions.

3.2. *Exploration du corpus et extraction des MD*

Nous avons effectué une première exploration des deux corpus, consacrée au repérage des MD simples les plus fréquents et les moins fréquents dans chaque corpus. Cela nous a permis de distinguer 15 MD qui cumulent au moins 5 occurrences dans un de deux corpus. A partir de ces premiers résultats, nous nous sommes penchées sur les MD complexes comprenant au moins un des MD simples les plus fréquents, à savoir *ben*, *bon* et *du coup*. Nous avons également retenu ceux qui cumulent plus de 5 occurrences dans, au moins, un des deux corpus analysés, suivant le même critère que pour les MD simples. Il faut signaler que dans les cas où deux de ces trois marqueurs apparaissent ensemble ils sont regroupés sous le MD qui apparaît en premier. Ainsi :

(1) *ben du coup* est classé comme MD complexe de *ben*.

(2) *du coup ben* est classé comme MD complexe de *du coup*.

3.3. *Analyses*

Étant donné que la taille des corpus est visiblement différente, nous avons normalisé les données en calculant le taux d'utilisation des MD par rapport au nombre total des mots des corpus, autrement dit en divisant le nombre total d'occurrences du MD observé par le nombre des mots du corpus où il apparaît, multiplié ensuite par 100. Cette opération permet d'une part, d'obtenir la distribution statistique des MD dans les deux corpus et d'autre part, d'observer quelle est la proportion d'utilisation d'un MD par rapport aux autres MD analysés.

Le taux a donc été calculé pour les 15 MD simples retenus et pour les 15 MD complexes contenant *bon*, *ben* et *du coup*. Nous avons également déterminé le taux total de ces mêmes MD dans chaque corpus ; ce dernier montrera quelle est la proportion de la totalité de MD analysés.

Enfin, les résultats ont été soumis au test U-Mann Whitney pour deux échantillons non appariés, afin d'observer si la différence entre le taux d'utilisation des MD est significative. L'hypothèse nulle est qu'il n'y a pas de différences significatives entre les deux corpus.

4. Résultats

En ce qui concerne la distribution des fréquences, on constate des différences : dans le corpus Manuels FLE, les trois MD simples les plus utilisés parmi ceux qui ont été retenus sont *mais*, *donc* et *alors* (tableau 1), alors que dans le corpus SCFLE ce sont *mais*, *ben* et *donc* (tableau 1). Il est donc à noter que, le marqueur *mais* est le plus employé dans les deux corpus.

Les différences corroborées concernent aussi les MD simples les moins fréquents : *quoi*, *disons* et *genre* dans le corpus Manuels FLE et *tu vois*, *hein* et *disons* dans le corpus SCFLE. Là aussi on retrouve un même marqueur : *disons*, très rarement utilisé dans les deux corpus.

Tableau 1. Nombre d'occurrences et taux des MD dans les deux corpus.

MD simples	Corpus Manuels		Corpus SCFLE		MD complexes	Corpus Manuels		Corpus SCFLE	
	occurr.	taux	occurr.	taux		occurr.	taux	occurr.	taux
N° de mots	446350		15861		N° de mots	446350		15861	
MD s. analysés	6733	1,508	861	5,428	MD c. analysés	187	0,042	59	3,720
<i>mais</i>	2671	0,598	168	1,059	<i>ah bon</i>	89	0,020	1	0,006
<i>donc</i>	1473	0,330	122	0,769	<i>mais bon</i>	38	0,009	9	0,057
<i>alors</i>	1156	0,259	34	0,214	<i>bon ben</i>	13	0,003	0	0,000
<i>là</i>	801	0,179	11	0,069	<i>bon mais</i>	6	0,001	0	0,000
<i>bon</i>	687	0,154	10	0,063	<i>bon alors</i>	5	0,001	0	0,000
<i>puis</i>	584	0,131	47	0,296	<i>ah ben</i>	15	0,003	5	0,032
<i>ben</i>	470	0,105	125	0,788	<i>ben alors</i>	6	0,001	0	0,000
<i>voilà</i>	369	0,083	62	0,391	<i>ben du coup</i>	4	0,001	5	0,032
<i>enfin</i>	240	0,054	88	0,555	<i>et ben voilà</i>	2	0,000	1	0,006
<i>hein</i>	174	0,039	5	0,032	<i>et puis ben</i>	2	0,000	2	0,013
<i>tu vois</i>	65	0,015	5	0,032	<i>donc du coup</i>	5	0,0011	7	0,0441
<i>du coup</i>	59	0,013	99	0,624	<i>et donc du coup</i>	2	0,0004	0	0,0000
<i>quoi</i>	44	0,010	39	0,246	<i>du coup ben</i>	0	0,000	9	0,0567
<i>disons</i>	19	0,004	1	0,006	<i>du coup ah</i>	0	0,000	9	0,0567
<i>genre</i>	3	0,001	15	0,095	<i>du coup euh</i>	0	0,000	11	0,0694

Les résultats du taux d'utilisation de chaque marqueur (tableau 1) illustrent également les différences entre les deux corpus, puisque le taux est plus élevé dans le corpus SCFLE que dans le corpus Manuels FLE sauf pour les marqueurs *là*, *alors* et *bon*. Les différences les plus remarquables se trouvent entre l'utilisation des MD *ben*, *du coup* et *enfin* qui dépassent respectivement 0,5 points. Les différences les moins importantes concernent les MD *hein* et *tu vois*, très peu présents dans les deux corpus : *hein* est légèrement plus utilisé dans le corpus Manuels FLE (0,039 vs 0,032) alors que *tu vois* est plus fréquent dans le corpus SCFLE (0,032 vs 0,015). Ces disparités sont nettement visibles grâce au taux total : 1,5 MD tous les 100 mots (Manuels FLE) et 5,4 MD tous les 100 mots (SCFLE).

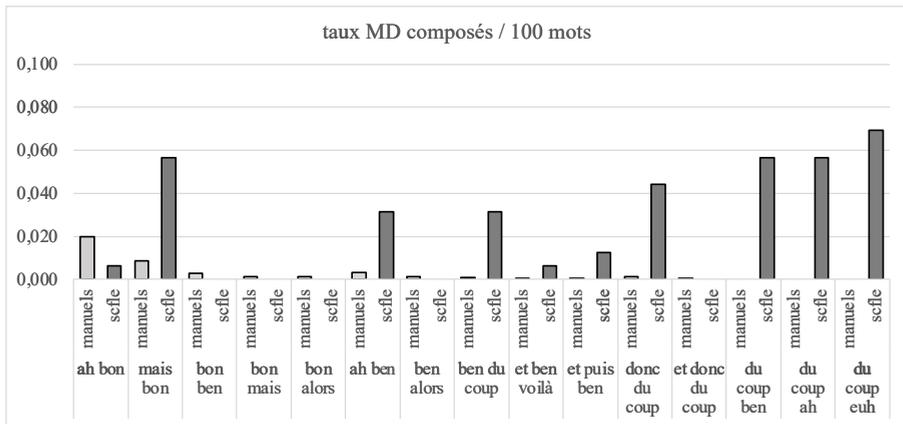


Figure 1. Taux MD complexes corpus Manuels FLE et corpus SCFLE.

L'écart entre les deux corpus est aussi marqué par les choix préférentiels des MD complexes (Figure 1). Un éventail plus large ressort du corpus Manuels FLE, ce qui pourrait être dû à la taille du corpus, cependant le taux d'utilisation est encore beaucoup plus élevé (3,720 vs 0,042) dans le corpus SCFLE. En outre, le type de MD complexes les plus fréquents dans chaque corpus varie, ainsi dans les manuels les locuteurs privilégient les MD contenant *bon*, à savoir *ah bon*, *mais bon* et *bon ben*, tandis que ce sont les MD composés par *du coup* : *du coup euh*, *du coup ah* et *du coup ben* les préférés par les locuteurs natifs. Enfin, les MD complexes *et ben voilà*, *et puis ben*, *ben bon* sont les moins utilisés dans les deux corpus. Malgré les différences trouvées, les résultats du test U Mann Whitney montrent qu'elles ne sont pas significatives statistiquement ni dans le cas des MD simples ($z = -.772$; $p > 0.05$) ni dans le cas des MD complexes ($z = -.628$; $p > 0.05$).

5. Conclusion

L'étude exploratoire des corpus Méthodes FLE et SCFLE a démontré que la fréquence d'utilisation, ainsi que les choix préférentiels des MD simples et complexes sont plus élevés lors des interactions orales des locuteurs natifs français que lors de celles proposées comme modèles discursifs dans les manuels aux apprenants de FLE.

Ces premiers résultats, qui se doivent d'être corroborés par l'étude d'autres grands corpus, nous amènent néanmoins à nous interroger sur la typologie et les emplois des marqueurs discursifs dans les interactions orales présentes dans les manuels FLE. Ainsi, cette étude n'est à considérer que comme une première étape d'une recherche qui vise à analyser les marqueurs discursifs aussi bien d'un point de vue quantitatif que qualitatif, afin de voir si les modèles discursifs

destinés aux apprenants de FLE reflètent les caractéristiques du français oral, notamment du français spontané.

Bibliographie

- Bazzanella C. (2006). Discourse markers in Italian: Towards a «compositional» meaning. In Fischer, K. (ed.) *Approaches to discourse markers*. Elsevier.
- Chanet C. (2004). Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie. *Recherches sur le français parlé*, 18: 83-107.
- Conseil de l'Europe. (2001). *Cadre Européen Commun de Référence pour les Langues*. Unité de politiques linguistiques. Disponible sur : <https://rm.coe.int/cadre-europeen-commun-de-reference-pour-les-langues-apprendre-enseigne/1680a4e270>
- Dostie G. (2004). *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*. De Boeck, Duculot.
- Dostie G., et Pusch C. D. (2007). Présentation. Les marqueurs discursifs. Sens et variation. *Langue française*, 2(154): 3-12.
- MacWhinney B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs*. Lawrence Erlbaum Associates Publishers.
- Mayaffre D. (2010). *Vers une herméneutique matérielle numérique. Corpus textuels, logométrie et langage politique*. Habilitation à Diriger des Recherches. Université Nice Sophia Antipolis.
- Rodriguez Somolinos A. (2011). Présentation. Les marqueurs du discours – Approches contrastives. *Langages*, 4(184): 3-13. Disponible sur : <http:// Cairn.info/revue-langages-2011-4page-3.htm>
- 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.
- Ruggia S. (2021). La lecture contrôlée et assistée par l'analyse statistique des données textuelles : comment et pourquoi interroger un corpus numérique ?, *Le français dans le monde, Recherches et Applications, Langues et pratiques numériques : nouveaux repères et nouvelles littératies en didactique des langues ?*, 69: 84-100.
- Ruggia S. et Vanni L. (2022). DeepFLE : la plateforme pour évaluer le niveau d'un texte selon le CECRL, *Dialogues et Cultures*, 66 : 235-254.
- Schmidt T., et Wörner K. (2009). EXMARaLDA—Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19(4): 565-582.
- Waltereit R. (2007). A propos de la genèse diachronique des combinaisons de marqueurs. L'exemple de bon ben et enfin bref. *Langue française*, 2(154): 94-109. Disponible sur : <http://www.cairn.info/revue-langue-francaise-2007-2-page-94.htm>