

LES CORPUS COMME INPUT ET COMME OUTPUT : L'EXEMPLE DES MARQUEURS BON ET BIEN

This article presents some results of the DiCoi project, which aims to create teaching material from oral corpora of French and to study the development of the interactional competence of learners of L2 French in vocational training for manual professions. We describe the main lines of a teaching intervention on the discourse markers *bon* and *bien* and the production of these markers in the longitudinal (two years) L2 French learner corpus DiCoi, including free interactions between peers. The results show that the use of *bien* is more productive than that of *bon*, as well as an increase in the frequency and diversity of the functions of the two markers among learners. The discussion comes back to the use of corpora as input and output.

● Anita Thomas
| Université de Fribourg
France Rousset
| Université de Genève



Anita Thomas est professeure au département de plurilinguisme et didactique des langues étrangères à l'Université de Fribourg.



France Rousset est doctorante à l'Université de Genève et a été collaboratrice scientifique dans le projet DiCoi.

1. Introduction

L'objectif de cet article est double. Premièrement, il présentera une démarche didactique sur les marqueurs discursifs (MD) *bon* et *bien* développée à partir de corpus de français parlé puis testée dans dix classes de jeunes en formation professionnelle de métiers manuels en Suisse romande. Deuxièmement, il présentera la production de ces deux marqueurs par seize apprenant-e-s ayant le français langue seconde (L2) à partir d'une analyse des données du corpus longitudinal d'interactions libres récolté lors des interventions en classe. Ces deux types de corpus, comme input et comme output, sont au cœur du projet de recherche appliquée DiCoi, dans le cadre duquel l'étude a été menée.

Dans cet article, nous traiterons les questions suivantes: 1) comment enseigner les marqueurs discursifs – *bon* et *bien* – polysémiques et polyfonctionnels avec des corpus oraux de français? 2) quel est le développement longitudinal de la

production de ces deux MD par les apprenant-e-s L2 dans des interactions libres? Pour ce faire, nous aborderons l'utilisation des corpus en didactique du FLE avant de présenter les grandes lignes du projet DiCoi. Nous nous concentrerons ensuite sur la didactisation des MD *bon* et *bien* dont nous présenterons les principaux emplois. Nous montrerons ensuite comment les apprenant-e-s utilisent ces deux MD au cours de deux années puis nous terminerons sur une discussion.

2. Utilisation des corpus comme ressource didactique

L'utilisation des corpus de langue parlée en didactique des langues étrangères s'est intensifiée au cours de ces dernières années en raison de leur mise à disposition grâce à internet. Ces bases de données comprenant à la fois des enregistrements audio et leur transcription permettent d'exposer les apprenant-e-s à la langue telle qu'elle est réellement utilisée (Boulton & Tyne, 2014). Parmi

les corpus oraux de français existants, nous pouvons citer le corpus FLEURON (<https://fleuron.atilf.fr>) qui comprend des interactions principalement administratives en milieu universitaire, mais aussi quotidiennes et le corpus CLAPI (<http://clapi.icar.cnrs.fr>) qui est composé de différents types d'interactions allant du privé au professionnel. Ce corpus a donné lieu à deux plateformes didactisées – CLAPI-FLE (<http://clapi.ish-lyon.cnrs.fr/FLE/accueil.php>) et CORAIL (<http://clapi.icar.cnrs.fr/Corail/index.html>) – toutes deux proposent des activités ciblant plusieurs phénomènes interactionnels fréquents. Finalement, le corpus OFROM (<https://ofrom.unine.ch>) comprend différents types d'interactions faisant intervenir des locuteur-trice-s de Suisse romande. Nous avons utilisé ces corpus et ressources didactisées pour construire notre propre matériel didactique.

3. Le projet de recherche DiCoi

Le projet de recherche appliquée DiCoi (digitalisation – corpus – interaction, <https://centre-plurilinguisme.ch/fr/recherche/Dicoi>) a été financé par le Centre de compétence sur le plurilinguisme (2021–2024). Il a pour objectif de tester l'apport des corpus de français parlé comme ressource didactique dans l'enseignement de la compétence d'interaction en français L2 et de décrire le développement longitudinal sur deux ans d'apprenant-e-s en formation professionnelle ayant un niveau intermédiaire (B1 – B2) de français (Thomas & Rousset, 2023). Le projet a été mené dans dix classes mixtes d'élèves ayant le français depuis la naissance (37 élèves francophones) ou comme L2 (30 élèves apprenant-e-s). Ces élèves suivent un apprentissage de base (attestation de formation professionnelle, AFP) pour des métiers manuels dans les secteurs de la restauration, la cuisine, la construction, la confection, la menuiserie, l'automobile ou encore du sanitaire.

3.1 Matériel didactique

Les interventions didactiques à partir des corpus ont été réparties sur deux ans. Nous avons réalisé huit interventions et onze exercices sous format numérique, que nous avons testés de manière empirique dans les dix classes. Les exercices ont été construits à l'aide du logiciel H5P (<https://h5p.org>) qui permet de présenter les exercices sous forme de livre numé-

rique. Nous avons ainsi construit des vidéos comprenant des inputs théoriques, des activités de compréhension orale et des analyses accompagnées d'exercices interactifs de différents types (vrai/faux, choix multiples, *drag and drop*, etc.). Les exercices ont été construits de sorte que les élèves puissent aller à leur rythme, réécouter certains extraits, afficher ou non les transcriptions ou encore refaire les activités. Tous les extraits proviennent de corpus, pour la plupart tirés des bases de données susmentionnées, et ont été intégrés soigneusement en tenant compte des difficultés des élèves, en particulier des apprenant-e-s L2 (vocabulaire et syntaxe accessible, élocution, input oral et écrit en parallèle, etc.). Le matériel didactique complet (y compris le déroulement des interventions en classe) est disponible en libre accès sur la plateforme <https://dicoi.ch>.

Les thématiques abordées dans le matériel didactique visent principalement la compétence d'interaction, par exemple les marqueurs discursifs *genre*, *trop*, *juste*, *bon* et *bien*, mais aussi des situations langagières professionnelles comme les interactions au travail ou l'entretien d'embauche.

Les séquences didactiques ont eu pour but de familiariser les élèves à l'utilisation des corpus et de leur permettre de gagner en indépendance à ce niveau-là en développant des compétences transversales notamment en informatique.

3.2 Corpus longitudinal d'interactions libres

Dans le but de documenter le développement longitudinal des apprenant-e-s du français L2, nous avons réalisé un enregistrement audio des élèves au moyen d'un dictaphone au début de chacune des huit interventions. Nous leur avons demandé de discuter librement deux par deux durant environ dix minutes sur les sujets de leur choix, tout en leur suggérant quelques thèmes, par exemple leur travail ou l'actualité¹. Les données ont ensuite été transcrites à l'aide du logiciel CLAN (<https://dali.talkbank.org/clang>). Ce corpus sera disponible sur le corpus SWIKO (<https://ifm-swiko.unifr.ch>, accès sur demande) qui rassemble les données d'apprenant-e-s de nombreux projets qui ont été menés à l'institut de plurilinguisme (voir par exemple Karges et al., 2020).

¹ Dans la mesure du possible, nous avons essayé de garder les mêmes paires d'apprenant-e-s, mais selon les classes nous avons été obligés de changer les binômes (absences, nombre impair, pas d'autre apprenant-e, etc.).

A gauche du mot	Mot recherché	A droite du mot
L2 : accueil des chercheurs et euh	bon	alors ce type de service euh en l'occurrence carte de séjour il faut savoir qu'avant euh les étudiants doivent aller L1 : à la préfecture L2 : à l'hôtel de police voilà à Lobau L1 : oui L2 : boulevard Lobau
L2 : et	bon	
A : donc c'est	bon	
E :	bon	
A : voilà E :	bon	
E : ouais ça va j'ai compris A :	bon	
E :	bon	
	bon	
	bon	
	bon	

Figure 1
Extrait du concordancier de FLEURON pour *bon*

Étape 4
Complétez au fur et à mesure le tableau ci-dessous selon le modèle (= extrait 1).

	voisins de <i>bon</i>	fonction (= à quoi ça sert ?)
1	et euh bon	gagner du temps quand on réfléchit à ce qu'on va dire ensuite, souvent avec <i>euh</i>
2		
3		

Figure 2
Tableau à remplir par les élèves à partir du concordancier



Figure 3
Fréquence d'utilisation de *bon* et *bien* par les apprenant-e-s

4. *bon* et *bien* : deux marqueurs polysémiques et polyfonctionnels

Plusieurs études se sont intéressées aux fonctions de ces deux MD tant à l'écrit qu'à l'oral en interaction. Peltier et Ranson (2020) proposent un état des lieux des analyses antérieures sur *bon* en établissant une nouvelle catégorisation des fonctions. Cette classification recense neuf fonctions textuelles (par exemple introduction d'un nouveau thème, résultat, formulation, ...) et deux fonctions touchant à l'attitude (contraste et résignation). Il en ressort aussi que *bon* est souvent accompagné d'autres MD. On retrouve notamment des collocations comme *mais bon* connotant souvent un contraste, *euh bon* dans des situations de reformulation / hésitation, *bon alors* pour introduire un nouveau thème, ou encore *ah bon* dans des acceptations. Finalement, *bon* permet à la fois d'initier et de clore une séquence (Lefevre, 2011).

Quant à *bien*, de nombreuses études se sont penchées sur ses emplois (voir par exemple Moline, 2012 pour une vue d'ensemble). En effet, *bien* peut avoir une valeur d'intensité, de (demande de) confirmation, d'approximation ou encore d'atténuation (Mosegaard Hansen, 1998 Parmi les fonctions recensées de *bien* en tant que MD dans les interactions, on relève celles d'introduire un nouveau thème, d'introduire une réponse, souvent avec *eh bien*, de valider le tour précédent ou de clore une séquence, souvent avec *très bien*.

5. *bon* et *bien* : séquence didactique

A partir des études antérieures, plusieurs choix didactiques ont été faits. Nous avons mis en place une séquence didactique d'une durée d'une heure réalisée lors de la troisième intervention en classe. Ainsi, les élèves possédaient déjà quelques connaissances sur les corpus, puisqu'ils avaient travaillé dessus durant les deux premières interventions. Dans cette optique, nous avons décidé de suivre l'approche du *data-driven learning* (André, 2019; Johns, 1991) et de faire travailler les élèves directement sur les données du corpus FLEURON (Figure 1). Nous avons porté notre choix sur ce corpus car il propose un concordancier permettant d'accéder directement aux interactions et

il était déjà connu des élèves lors d'inputs précédents.

Après avoir rappelé son utilisation, nous avons entamé l'analyse du MD *bien* en plénum afin d'activer la démarche d'analyse des interactions. En étant semi-guidés, les élèves ont ensuite travaillé en autonomie afin de relever et d'identifier les différentes fonctions de *bon* dans le corpus FLEURON en les reportant sur une feuille (Figure 2). Nous avons ensuite procédé à une mise en commun en plénum ouvrant une discussion sur leurs pratiques langagières quotidiennes.

A la fin de l'intervention, nous avons récolté les feedbacks des élèves sur l'activité. Il en est ressorti que la majorité des élèves a apprécié de travailler sur ce corpus qu'ils ont jugé facile d'utilisation et utile. Néanmoins, la tâche d'identification des fonctions a été perçue comme étant difficile tout en ayant permis une prise de conscience de la polysémie et de la polyfonctionnalité de ces deux MD utilisés et entendus au quotidien.

6. Développement longitudinal de *bon* et *bien* par les apprenant·e·s L2

Rappelons que les enregistrements ont eu lieu au début de chaque intervention pour une durée de 10 minutes sous la forme d'interactions libres entre paires. Ces interactions ont été transcrites dans CLAN, permettant de réaliser différents types d'analyses. Dans cet article, nous nous focalisons sur les données longitudinales de 16 apprenant·e·s. Ces apprenant·e·s ont été sélectionnés parce qu'ils ont participé à tous les enregistrements (ou en ont manqué maximum un) et que les enregistrements étaient transcrits au moment de l'analyse. Comme le montre le tableau 1, les apprenant·e·s ont des L1 non-européennes et la plupart sont des hommes.

Tout d'abord, nous avons effectué une recherche concernant la fréquence d'utilisation des deux MD par les seize apprenant·e·s. Les résultats sont présentés dans la Figure 3. L'axe horizontal correspond aux huit enregistrements et l'axe vertical indique le nombre d'occurrences pour chaque MD. A noter que A121 était absent à DiCoi7 et A124 à DiCoi6, A125 à DiCoi8.

Code	Classe	L1	Genre	Années en Suisse (2021)	Code	Classe	L1	Genre	Années en Suisse (2021)
A101	01	Tigrinya	F	7	A134	05	Tigrinya	H	4
A114	02	Tigrinya	H	4	A135	05	Tigrinya	H	3
A115	02	Tigrinya	H	6	A136	05	Tigrinya	H	7
A121	03	Tigrinya	H	7	A139	05	Ouzbek	H	6
A122	03	Tigrinya	H	5	A152	06	Tibétain	F	11
A124	04	Tigrinya	H	6	A160	07	Tigrinya	H	8
A125	04	Tigrinya	H	5	A161	07	Tigrinya	H	8
A128	04	Tigrinya	H	4	A163	07	Tigrinya	H	8

Tableau 1

Apprenant·e·s de l'étude

Au total, nous avons relevé 89 occurrences pour *bon* et 511 pour *bien*. Dans FLEURON *bien* était également plus fréquent que *bon* (226 *bon* et 425 *bien*). Ainsi la plus faible fréquence de *bon* dans les données correspond à ce que l'on trouve en L1. On note tout de même que A128 et A139 n'ont par exemple jamais utilisé *bon* dans leurs discussions sur la période des deux années. Toutes les productions des deux marqueurs sont correctes.

Rappelons que l'intervention sur *bon* et *bien* a eu lieu lors de l'intervention 3. Nous observons l'utilisation de ces MD par certain·e·s apprenant·e·s déjà avant l'input. C'est par exemple le cas pour A114, A121, A124 ou A135. Nous n'observons pas de forte augmentation de l'utilisation des MD lors de l'enregistrement 4, après l'input. Bien qu'il semble y avoir un développement pour A125 et A161 au fil du temps, tendant vers une utilisation plus accrue de ces deux MD, force est de constater que la fréquence ne permet pas de discerner de réels effets de l'input, ni d'observer un développement dans le temps. En outre, la variation au niveau de la fréquence d'utilisation peut être influencée par divers facteurs, en particulier le sujet de discussion et l'engagement des participant·e·s.

6.1 Développement des fonctions de *bon*

Ceci nous a amené à procéder à l'annotation des fonctions et structures dans les transcriptions comme dans les exemples (1) et (2). Dans l'extrait (1), A125 clôt la séquence en cours. Il utilise *voilà* pour indiquer qu'ils se dirigent vers la fin d'une séquence, suivi d'une pause et de *bon merci beaucoup* venant la clore. Dans l'extrait (2), une séquence sur la danse se clôt par

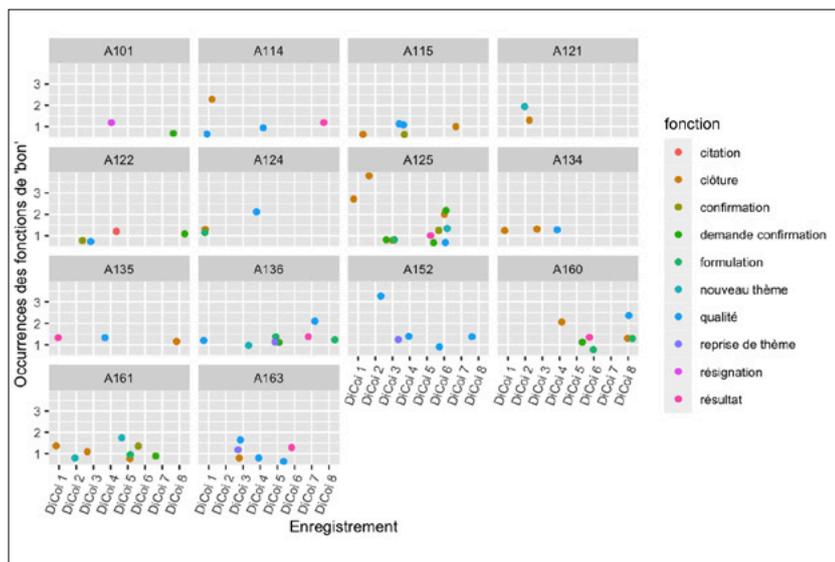


Figure 4
Fonctions de *bon* produites par les apprenant-e-s

ok ok produit par A161. Après une courte pause, A161 propose un nouveau thème de discussion et l'introduit grâce à *bon*.

Le résultat de l'analyse des fonctions de *bon* est présenté dans la Figure 4. Pour mémoire, les apprenants A128 et A139 n'ont pas produit d'occurrences de *bon*. On constate que les fonctions de clôture et de qualité sont présentes dès le premier enregistrement chez plusieurs locuteurs comme A114, A125 ou encore A161. Néanmoins, on observe un développement de l'utilisation de *bon* à partir de l'enregistrement 4 pour A160. De plus, A125, A136, A160 et A161 tendent à utiliser dans un nombre croissant de fonctions au cours du temps (fonctions de clôture, formulation, résultat, reprise de thème, nouveau thème ou encore de confirmation / demande de confirmation).

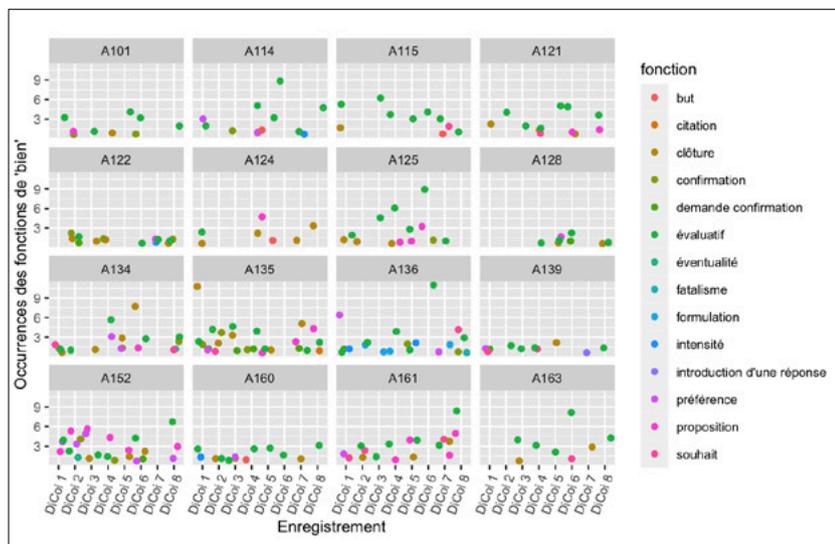


Figure 5
Fonctions de *bien* produites par les apprenant-e-s

6.2 Développement des fonctions de bien

Nous avons procédé de manière similaire pour l'analyse de la production de *bien*. Deux exemples du codage des fonctions de *bien* sont présentés en (3) et (4).

Dans l'extrait (3), A134 demande confirmation avec *français*, validée par A139 avec *ouais*. A134 valide ensuite le tour de A139 et se dirige vers la fin de la séquence sur l'importance du français en utilisant *c'est bien*, ce que nous avons codé comme une fonction de clôture. Dans l'extrait (4), il s'agit de l'expression fixe *ou bien* en fin de tour de parole.

Le résultat de l'analyse des fonctions de *bien* est présenté dans la Figure 5. On observe que *bien* est utilisé dans plus de fonctions différentes comparé à *bon* et ce par l'ensemble des participant-e-s. Les fonctions présentes dès les premiers enregistrements sont surtout celles d'évaluation et de clôture. Bien que certain-e-s apprenant-e-s produisent *bien* dans une variété de fonctions (évaluation, souhait, reformulation, confirmation/demande de confirmation, préférence) déjà avant notre intervention 3, en particulier A135, A136, A152 et A161, on voit que la diversification des fonctions se fait dès le quatrième enregistrement. Même si ce développement n'est pas forcément lié à notre intervention didactique, on ne peut exclure qu'elle a contribué à un développement de la production de *bien* en rendant saillant les diverses utilisations.

(1) Enregistrement 1: A124 et 125

*A125: aussi sur internet (..) xxx voilà (..) **bon** (..) merci beaucoup .
%com: MD_bon_clôture
*A124: merci au'voir .

(2) Enregistrement 5: A160 et 161

*A161: mais tu dances quand même à la maison ou pas .
*A160: nan &=rires je connais pas danse .
*A161: &=rires ok tu dances pas à la maison .
*A160: +< hm hm .
*A161: ok ok (..) bon t' as fait quoi d'autre pendant le week-end .
%com: MD_bon_nouveau_thème
*A160: euh je rien sauf sortir discuter avec des amis et euh (..) hm: un peu discuté avec mon frère .

(3) Enregistrement 1 : A134 et A139

- *A139: quelle langue que (.) j'aimerais parler?
 - *A134: tu veux: tu aimerais (.) t' &-aim- t' aimerais bien apprendre .
 - *A139: ouais j' aimerais bien apprendre (.) ouais .
 - *A134: +< hm (.) .
 - *A139: ben celui que le français pour l' instant ouais c' est l'important (.) ici c' est le français .
 - *A134: français ouais .
 - *A139: +< ouais (.) .
 - *A139: pour mon formation aussi (.) ouais (.) hm (.) &-qu- ouais .
 - *A134: français xxx (.) .
 - *A139: ouais .
 - *A134: ouais c'est bien .
- %com : MD_bien_clôture**

(4) Enregistrement 5 : A124 et A125

- *A124: nan nan nan jamais jamais moi j' apprends toujours j' apprends j'apprends .
 - *A125: jusqu' à mort?
 - *A124: jusqu' à mourir ouais .
 - *A125: ça va ou bien .
- %com : MD_bien_expression**

Les graphiques de la Figure 5 visualisent en outre un développement dans le temps pour plusieurs apprenant·e-s, A124, A134, A163 et plus faiblement A128. Les apprenant·e-s A135, A136 et A161 utilisent beaucoup *bien* et dans différentes fonctions.

7. Discussion

Dans cette contribution nous avons présenté un exemple d'utilisation des corpus comme ressource pour un public d'apprenant·e-s L2 du français qui se forme à des métiers manuels. Bien que l'utilisation des corpus dans une approche de type *data-driven learning* exige une certaine réflexion métalinguistique, il semble que ce genre de tâche soit faisable avec un public non académique. Ce résultat confirme les études antérieures menées par André (par exemple André, 2019).

Quant à l'effet de notre intervention, les données récoltées ne nous permettent pas d'établir un lien direct entre notre travail sur *bon* et *bien* dans les classes et la production de ces deux MD par les seize apprenant·e-s de cette étude. Néanmoins, les données longitudinales montrent un certain développement au niveau de la diversité des fonctions utilisées et ce même pour *bon* qui est plus rare dans les données. En ce sens l'étude suggère qu'une approche sur corpus pour un phénomène interactionnel fréquent en

français permet d'augmenter les processus attentionnels.

Le recours à des enregistrements d'interaction libres comme méthode pour l'observation du développement longitudinal d'apprenant·e-s comprend à la fois des forces et des faiblesses. La force d'un corpus d'interactions libres réside dans l'authenticité des données récoltées. Les apprenant·e-s parlent de leur quotidien et de ce qui les intéresse. La faiblesse de la méthode réside dans l'absence de contrôle des connaissances, comme on l'aurait par le biais de tests d'élicitation ciblés.

Contrairement à des apprenant·e-s de classes de langues à l'école obligatoire ou post-obligatoire, les apprenant·e-s du projet DiCoi évoluent dans un environnement riche au niveau de l'input du français parlé. Une entrée dans la langue par l'oral, sans passer par les manuels traditionnels fortement influencés par les règles de l'écrit, pourrait expliquer l'utilisation relativement productive de MD dès les premiers enregistrements. Les données de ce corpus décrivent donc principalement un développement informel du français L2.

Remerciements

Le projet DiCoi a été financé par le Centre scientifique de compétence sur le plurilinguisme (Fribourg, Suisse). Il a été

réalisé en partenariat didactique avec Centres de Formation Professionnelle de l'État de Fribourg (CD-CFP). Nous remercions les enseignant·e-s et les élèves ayant accepté de participer au projet ainsi que les assistant·e-s de recherche pour leur soutien lors de la récolte des données.

Références

- André, V.** (2019). Pourquoi faire de la sociolinguistique des interactions verbales avec des enseignants et des apprenants de Français Langue Étrangère ? *Linx. Revue des linguistes de l'université Paris X Nanterre*, 79. <https://doi.org/10.4000/linx.3694>
- Boulton, A., & Tyne, H.** (2014). *Des documents authentiques aux corpus : Démarches pour l'apprentissage des langues*. Didier.
- Johns, T.** (1991). *Should you be persuaded: Two samples of data-driven learning materials*. *English Language Research Journal*, 4, 1–16.
- Karges, K., Studer, T., & Wiedenkiller, E.** (2020). Textmerkmale als Indikatoren von Schreibkompetenz. *Bulletin Suisse de Linguistique Appliquée*, 117–140.
- Lefevre, 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. <https://doi.org/10.4000/linx.1417>
- Moline, E.** (2012). Aperçu des emplois de bien en français contemporain. *Travaux de linguistique*, 65(2), 7–26. <https://doi.org/10.3917/tl.065.0007>
- Mosegaard Hansen, M.-B.** (1998). *The Function of Discourse Particles : A Study with Special Reference to Spoken Standard French*. John Benjamins Publishing.
- Peltier, J. P. G., & Ranson, D. L.** (2020). Le marqueur discursif bon : Ses fonctions et sa position dans le français parlé. *SHS Web of Conferences*, 78, 01006. <https://doi.org/10.1051/shsconf/20207801006>
- Thomas, A., & Rousset, F.** (2023). Utilisation des corpus pour l'enseignement de l'interaction en formation professionnelle de métiers manuels : Exemple d'un exercice numérique sur « genre ». *Corpus*, 24. <https://doi.org/10.4000/corpus.7899>