



Corpus de parole non-native et prédiction automatique du niveau de performance en expression orale : application à CLIJAF

Verdiana De Fino^{1,2}, Lionel Fontan², Sylvain Detey³, Isabelle Ferrané¹, Julien Pinquier¹

(1) *IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France*

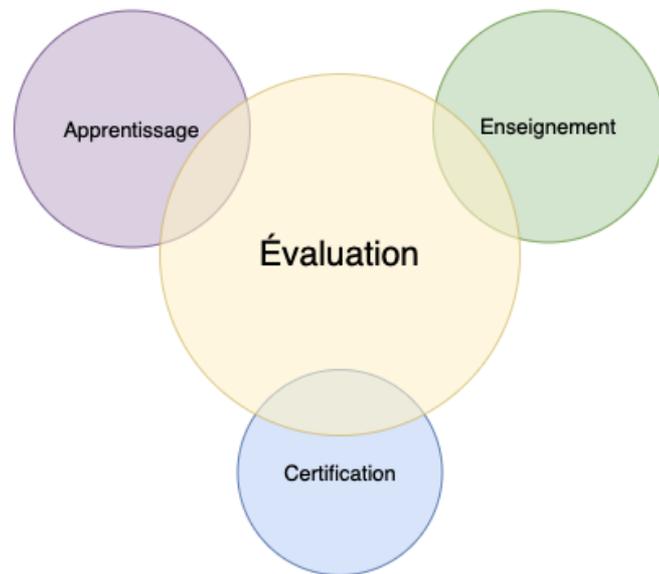
(2) *Archean Labs, Montauban, France*

(3) *SLS & GSCCS, Waseda University, Tokyo, Japon*

Contexte

- Apprenants et enseignants de français doivent pouvoir disposer d'outils d'évaluation de l'oral
- Cohérence avec les descripteurs du CECRL
- Difficultés d'évaluation
 - oral ≠ écrit
 - aspects phonético-phonologiques de l'oral

¹CECRL = Cadre Européen Commun de Référence pour les Langues



Cadre de l'étude



- LabCom ANRALAIA² (Apprentissage des langues Assisté par Intelligence Artificielle)
- Projet KAKEN « From corpus to target data as steps for automatic assessment of L2 speech : L2 French phonological lexicon of Japanese learners »³

Objectifs

1. Explorer la possibilité de mesurer automatiquement le niveau CECRL d'apprenants japonais
2. Prédire le niveau CECRL des apprenants
3. Prédire le niveau CECRL en production orale des apprenants

²<https://www.irit.fr/SAMOVA/site/projects/current/labcom-alaia/>

³Detey, S. (dir.) (2020-2024). JSPS : Grant-in-Aid for Scientific Research (B) 20H01291

Évaluation automatique de productions non natives



Différents outils existent pour mesurer objectivement les productions des apprenants de langue étrangère à différents niveaux linguistiques

- Production orale : évaluation de paramètres phonétiques segmentaux et suprasegmentaux
- Production écrite : évaluation de paramètres lexicaux, syntaxiques et discursifs

Évaluation automatique de productions non natives



Prononciation des phonèmes

- Calcul des scores de reconnaissance des phonèmes (utilisation d'un système de reconnaissance automatique de la parole (RAP), alignement libre ou forcé) (Detey et al., 2016)

Fluence : « degré de fluidité du discours sans pauses ni autres marques de disfluence » (*Derwing & Munro, 2015*)

- Calcul de différentes variables (débit de parole, pourcentage de parole) à partir de pseudo-syllabes et de pauses silencieuses (Fontan et al., 2020; Farinas & Pellegrino, 2001). Méthode de segmentation basée sur la détection de ruptures dans la trajectoire de l'énergie du signal de parole au cours du temps (André-Obrecht, 1988)

Évaluation automatique de productions non natives



Capacité lexicale : richesse lexicale (Laufer & Nation, 1995)

- Diversité lexicale → proportion de mots différents produits
Index de Guiraud (Guiraud, 1959)
- Densité lexicale → proportion de mots lexicaux produits
- Sophistication lexicale → proportion de mots lexicaux rares/ peu fréquents produits
Ressource utilisée : Base de données Lexique3 (New et al., 2005)

Évaluation automatique de productions non natives



Capacité syntaxique : complexité syntaxique (Lahuerta Martínez, 2018; Blache, 2010)

- Calcul de la longueur moyenne des phrases (nombre de mots)
- Calcul de la profondeur moyenne des arbres syntaxiques

Capacité discursive : cohésion du discours (Beacco et al., 2004) et planification

- Proportion de marqueurs (connecteurs) du discours
- Diversité des marqueurs du discours
- Proportion de mots d'hésitation
- Proportion de faux départs

Corpus - CECRL global



Sous-ensemble corpus CLIJAF⁴

→ cadre méthodologique général du projet IPFC (Detey & Kawaguchi, 2008; Racine et al., 2012)

Tâche de production orale semi-spontanée d'étudiants japonais apprenant le français à l'université au Japon

→ Entretien individuel semi-directif en français, mené par un locuteur natif

→ Entretiens enregistrés, segmentés manuellement en tours de parole et transcrits orthographiquement

→ unité d'analyse : réponse de l'apprenant

→ 1h20 de données

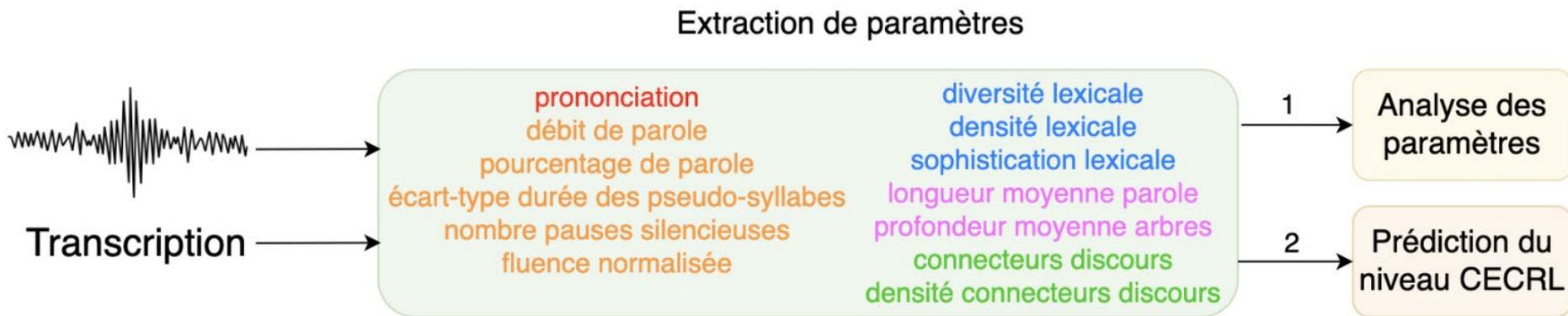
→ 18 apprenants (3H, 15F)

→ niveau CECRL d'après certification TCF/ DELF/ DALF/ Futsuken

Niveau	Nombre
A2	1
B1	4
B2	8
C1	3
C2	2

⁴Detey, S. (dir.) (2011-2019). A longitudinal interphonological corpus of Japanese learners of French. JSPS : Grant-in-Aid for Scientific Research (B) 23320121 & 15h03227

Schéma récapitulatif

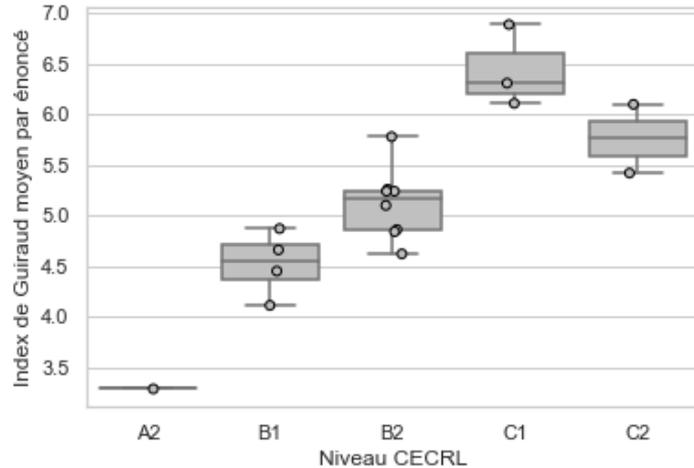


1 : Explorer la possibilité de mesurer automatiquement le niveau CECRL d'apprenants japonais

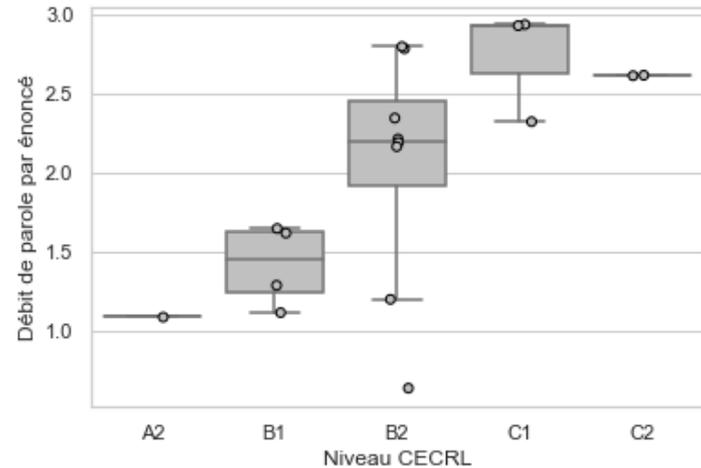
2 : Prédire le niveau CECRL des apprenants

Résultats - Validation de nos mesures automatiques

Évolution de la **diversité lexicale** selon le niveau CECRL



Évolution du **débit de parole** selon le niveau CECRL



Résultats - Prédiction du niveau des apprenants



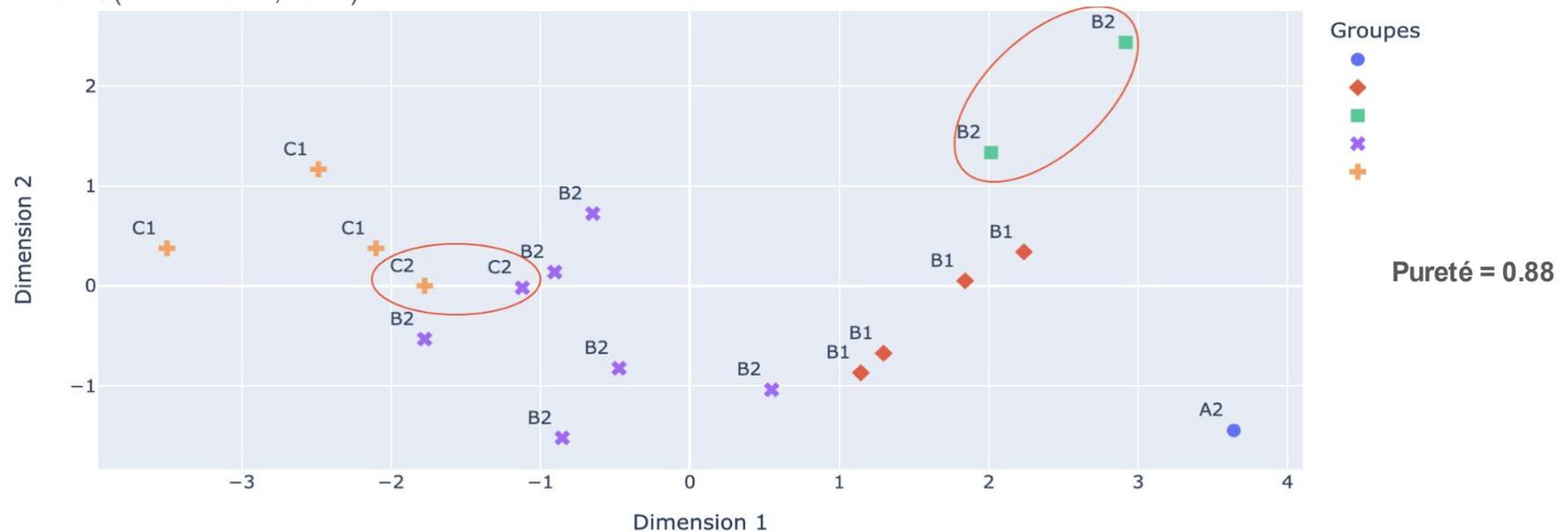
Algorithme d'apprentissage automatique non supervisé *k-means*

Apprenants japonophones représentés par un jeu de six paramètres

- **pourcentage de parole**
- **débit de parole**
- **fluence normalisée**
- **diversité lexicale**
- **longueur moyenne des réponses**
- **proportion de connecteurs du discours**

Résultats - Prédiction du niveau des apprenants

Coordonnées des 18 participants selon les deux composantes principales résultant du partitionnement en *k-means* (De Fino & al., 2022)



Conclusion - Prédiction du niveau des apprenants



- Peu de données à disposition
 - Certaines mesures acoustiques et linguistiques sont corrélées au niveau certifié CECRL
 - Une classification automatique non supervisée est possible
 - Le groupe des apprenants C1 obtient de meilleurs scores que le groupe C2
→ disparité entre l'évaluation de l'oral et de l'écrit ?
- ⇒ Enrichissement du corpus grâce à l'évaluation, par des enseignants-experts, du niveau CECRL en production orale des apprenants

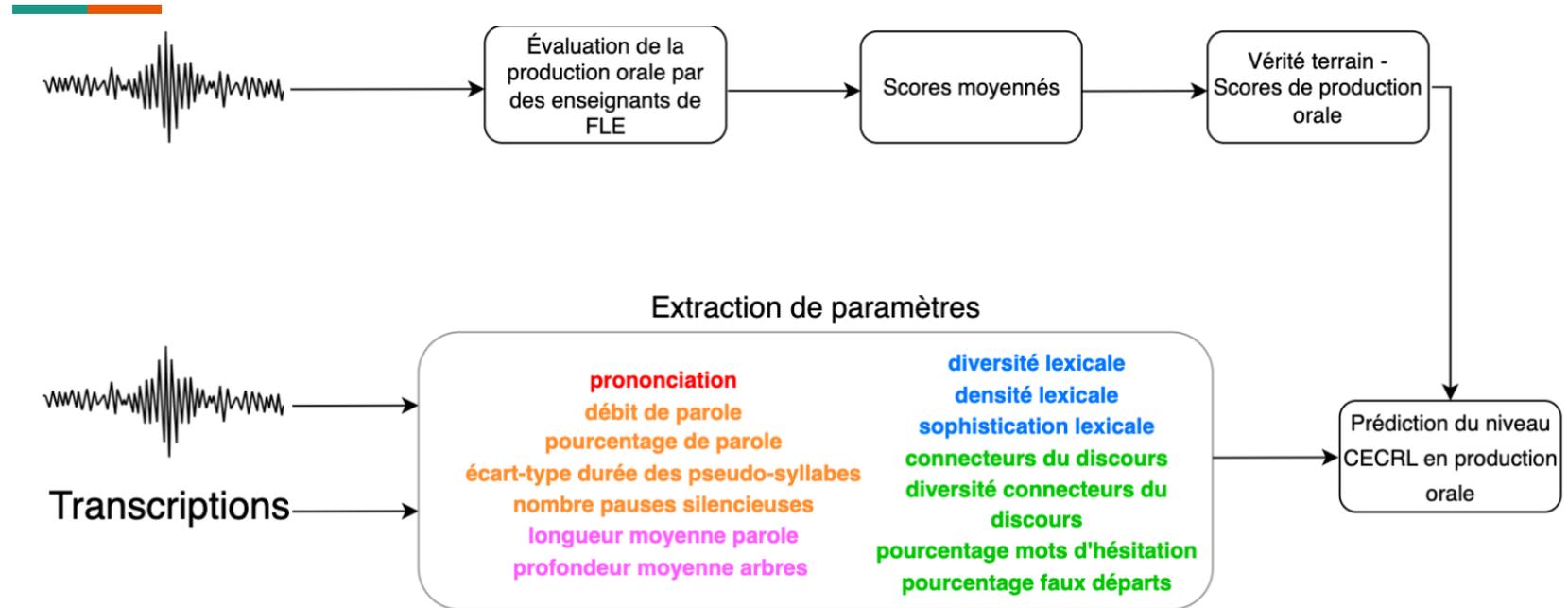
Évaluation experte du niveau en production orale

- Trois enseignants (2M, 1F) de français langue étrangère (FLE)
→ examinateurs officiels du DELF/DALF
- Évaluation du niveau en production orale de 38 apprenants
- Enregistrements présentés une fois à chaque enseignant de FLE, dans un ordre aléatoire, en utilisant le logiciel Prodigy⁵
- Niveaux CECRL traduits en échelle numérique discrète (A1 = 1, A2 = 2, ..., C2 = 6)
- Fort accord inter-annotateur
→ scores moyennés

Pair of teachers	ρ	<i>p</i> -value
1, 2	0.78	<.001
1, 3	0.79	<.001
2, 3	0.70	<.001

⁵<https://prodi.gy/>

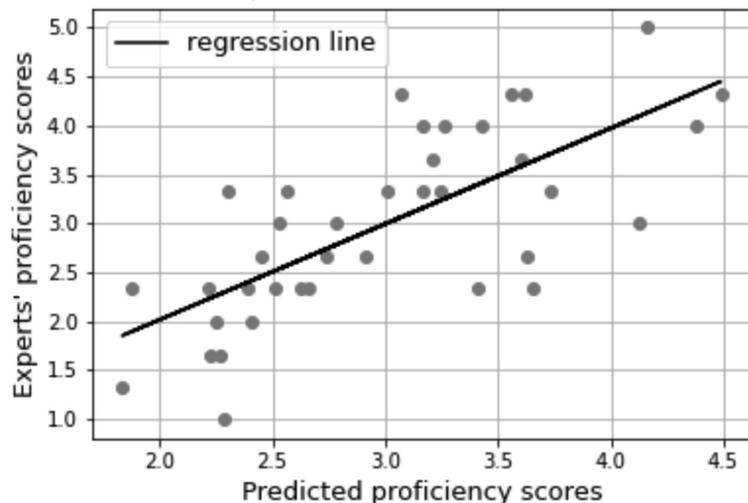
Schéma récapitulatif



3 : Prédire le niveau CECRL en production orale des apprenants

Résultats - Prédiction du niveau en production orale

- Régression Lasso stratégie leave-one-out nested cross validation ($\alpha = 0.1$)
- Erreur absolue moyenne : MAE = 0.53
- Corrélation entre nos prédictions et les scores donnés par les enseignants : $r = 0.71$ (De Fino & al., 2022)



Teacher	r	p -value
1	0.60	< .001
2	0.59	< .001
3	0.73	< .001

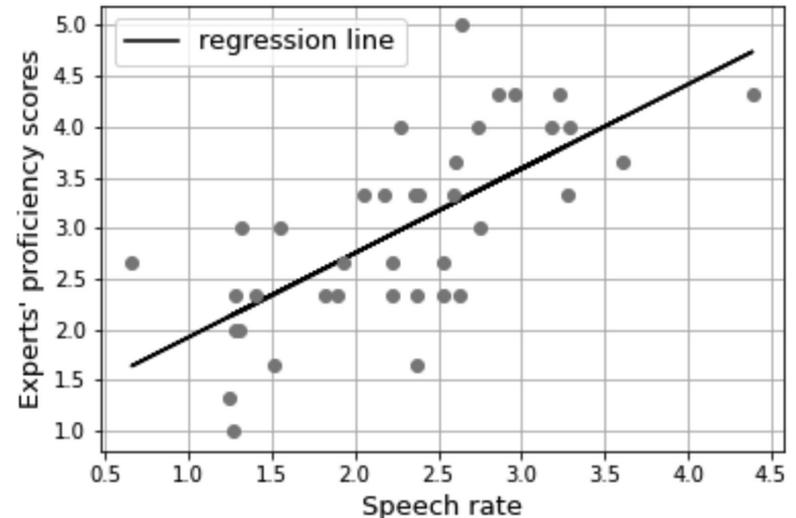
Corrélation (Pearson) entre nos prédictions et les scores donnés en fonction des enseignants

Résultats - Prédiction du niveau en production orale

- Cinq paramètres sélectionnés par la régression Lasso
- Le **débit de parole** contribue le plus pour la prédiction du niveau en production orale d'apprenants japonais de français

Feature	Std. coefficient
Speech rate	0.160
Proportion of false starts	-0.108
Proportion of hesitation words	0.037
Mean length of answers (words)	0.019
Lexical density	-0.007

Coefficients standardisés associés aux cinq paramètres sélectionnés par la régression Lasso



Discussion et conclusion



- Prédiction du niveau en production orale d'apprenants japonais de français satisfaisante en utilisant cinq paramètres mesurés automatiquement ($MAE = 0.53$, $r = 0.71$)
 - d'autres paramètres ont été écartés par la régression Lasso parce qu'ils étaient déjà corrélés avec certains paramètres sélectionnés ?
 - d'autres types de modèles pourraient potentiellement améliorer la prédiction ?
- Corrélation $r = 0.73$ entre nos prédictions et les scores donnés par l'enseignant 3
 - la prédiction automatique pourrait être considérée comme un quatrième évaluateur
- Certaines mesures sont basées sur les transcriptions orthographiques (planification du discours)
 - ⇒ Remplacer cette étape manuelle par l'utilisation d'un système RAP

Références



- André-Obrecht R. (1988). A new statistical approach for the automatic segmentation of continuous speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(1), 29–40.
- Beacco J. C., Bouquet S. & Porquier R. (2004). *Niveau B2 pour le français*. Paris : Didier.
- Blache P. (2010). Un modèle de caractérisation de la complexité syntaxique. In *Actes de la 17e conférence sur le Traitement Automatique des Langues Naturelles*, p. 81–90, Montréal, Canada : ATALA.
- De Fino, V., Fontan, L., Pinquier, J., Barcat, C., Ferrané, I., Detey, S. (2022). Mesures automatiques de parole non-native : exploration pilote d'un corpus d'apprenants japonais de français et différenciation de niveaux. *34èmes Journées d'Études sur la Parole (JEF2022)*, Noirmoutier (France).
- De Fino, V., Fontan, L., Pinquier, J., Ferrané, I., Detey, S. (2022). Prediction of L2 speech proficiency based on multi-level linguistic features. *Interspeech2022*, Sep 2022, Incheon (South Korea).
- Derwing T. M. & Munro M. J. (2015). *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam : John Benjamins.
- Detey S., Fontan L. & Pellegrini T. (2016). Traitement de la prononciation en langue étrangère : approches didactiques, méthodes automatiques et enjeux pour l'apprentissage. *Revue TAL*, 57(3), 15–39.
- Detey S. & Kawaguchi Y. (2008). Interphonologie du Français Contemporain (IPFC) : récolte automatisée des données et apprenants japonais. In *Journées PFC : Phonologie du français contemporain : variation, interfaces, cognition*, Paris : MSH.
- Farinas J. & Pellegrino F. (2001). Automatic rhythm modeling for language identification. In *Proc. 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, p. 2539–2542.
- Fontan L., Le Coz M. & Alazard C. (2020). Using the forward-backward divergence segmentation algorithm and a neural network to predict L2 speech fluency. In *Proc. 10th International Conference on Speech Prosody 2020*, p. 925–929.
- Guiraud P. (1959). *Problèmes et méthodes de la statistique linguistique*, volume 2. Dodrecht : D. Reidel.
- Lahuerta Martínez A. C. (2018). Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. *Assessing Writing*, 35, 1–11.
- Lauffer B. & Nation P. (1995). Vocabulary size and use : Lexical richness in L2 written production. *Applied linguistics*, 16(3), 307–322.
- New B., Pallier C. & Ferrand L. (2005). Manuel de Lexique3. *Behavior Research Methods, Instruments, & Computers*, 36(3), 516–524.
- Racine I., Zay F., Detey S. & Kawaguchi Y. (2012). Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2 : l'exemple du projet "Interphonologie du français contemporain" (IPFC). In A. Kamber & C. Skupien, Eds., *Recherches récentes en FLE*, P. 1–19. Bern : Peter Lang.