



HAL
open science

Mesures automatiques de parole non-native : exploration pilote d'un corpus d'apprenants japonais de français et différenciation de niveaux

Verdiana De Fino, Lionel Fontan, Julien Pinquier, Corentin Barcat, Isabelle Ferrané, Sylvain Detey

► To cite this version:

Verdiana De Fino, Lionel Fontan, Julien Pinquier, Corentin Barcat, Isabelle Ferrané, et al.. 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 (JEP 2022), AFCP (Association Francophone de la Communication Parlée), Jun 2022, Noirmoutier, France. pp.1-10. <hal-03700378>

HAL Id: hal-03700378

<https://hal.science/hal-03700378v1>

Submitted on 21 Jun 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.



HAL Authorization

Mesures automatiques de parole non-native : exploration pilote d'un corpus d'apprenants japonais de français et différenciation de niveaux

Verdiana De Fino^{1,2} Lionel Fontan² Julien Pinquier¹ Corentin Barcat³
Isabelle Ferrané¹ Sylvain Detey⁴

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

(2) Archean Labs, 20 Pl Prax-Paris, 82000 Montauban, France

(3) Tokyo University of Foreign Studies, 3-11-1 Asahi-cho, Fuchu-shi, 183-8534 Tokyo, Japon

(4) SILS & GSICCS, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-ku, 169-8050 Tokyo, Japon
verdiana.defino@irit.fr, lfontan@archean.tech, julien.pinquier@irit.fr,
corentin.barcat@gmail.com, isabelle.ferrane@irit.fr, detey@waseda.jp

RÉSUMÉ

Cette étude s'intéresse à l'évaluation automatique de la parole de Japonais apprenant le français. Nos données sont des productions orales semi-spontanées, issues du corpus CLIJAF. Nous évaluons différents niveaux linguistiques (prononciation, lexique, syntaxe, discours) à l'aide de mesures automatiques. Notre premier objectif est d'étudier si ces mesures sont en adéquation avec le niveau général des apprenants, tel que défini par le CECRL (A2 à C2). Notre deuxième objectif est d'explorer la possibilité de prédire ledit niveau en utilisant les paramètres linguistiques les plus pertinents, avec un algorithme d'apprentissage automatique non-supervisé. Les résultats obtenus valident la pertinence de nos objectifs et montrent que des paramètres extraits automatiquement tels que la diversité lexicale, la longueur moyenne des tours de parole (niveau syntaxique), la proportion de marqueurs du discours et plusieurs paramètres liés à la fluence peuvent contribuer à la prédiction du niveau des apprenants.

ABSTRACT

Automatic assessment of non-native speech : an exploratory study of a corpus of Japanese learners of French and level differentiation

This study focuses on the automatic evaluation of the speech of Japanese learners of French. Our data are semi-spontaneous oral productions of learners from the CLIJAF corpus. We evaluate different linguistic levels (pronunciation, lexicon, syntax, discourse) using automatic assessment tools. Our first objective is to study whether these measurements are in line with the general level of proficiency of the learners, as defined by the CEFRL (A2 to C2). Our second objective is to explore the possibility of predicting this level using the most relevant linguistic parameters, with an unsupervised machine learning algorithm. The results confirm the relevance of our objectives and show that automatically extracted parameters such as lexical diversity, the average length of speech turns (syntactic level), the proportion of connectives in the discourse and several parameters related to fluency can contribute to the prediction of learners' level.

MOTS-CLÉS : évaluation automatique, parole non-native, parole semi-spontanée, niveaux linguistiques.

KEYWORDS: automatic evaluation, non-native speech, semi-spontaneous speech, linguistic levels.

1 Introduction

L'évaluation fait partie des composants nécessaires du processus d'apprentissage, d'enseignement et de certification. Tous les apprenants de français, ainsi que leurs enseignants, doivent dès lors pouvoir disposer d'outils leur permettant d'évaluer leur production orale, en particulier vis-à-vis des descripteurs communément admis aujourd'hui du Cadre Européen Commun de Référence pour les Langues (CECRL, Conseil de l'Europe, 2001). Les difficultés d'évaluation de l'oral diffèrent de celles de l'écrit, en ce que la première nécessite la prise en compte du plan phonético-phonologique. Or, ce plan n'est pas toujours cohérent avec celui d'autres composants de l'apprentissage, tels que le lexique ou la grammaire. Il est aussi intéressant de noter que, de par sa complexité, l'évaluation automatique de la parole non native, notamment en français, est un domaine encore insuffisamment documenté.

Notre étude s'inscrit ainsi dans le cadre d'un programme de recherche conjoint entre le laboratoire commun ALAIA¹ (*Apprentissage des Langues Assisté par Intelligence Artificielle*) et le projet « *From corpus to target data as steps for automatic assessment of L2 speech : L2 French phonological lexicon of Japanese learners* »², relatif à l'évaluation automatique de la production orale d'apprenants japonais de français. Le premier objectif de cette étude est d'explorer la possibilité de mesurer automatiquement les compétences orales, à différents niveaux linguistiques, d'apprenants japonais de niveaux A2 à C2. L'objectif second est de prédire, à partir de ces mesures automatiques, le niveau global des apprenants. Les analyses que nous employons sont basées d'une part sur le signal acoustique pour mesurer des paramètres phonético-phonologiques (prononciation des phonèmes et fluence phonétique), et d'autre part sur les transcriptions orthographiques de ces productions pour mesurer divers paramètres lexicaux, syntaxiques et discursifs.

Après avoir présenté, dans la section 2, différents outils existants destinés à évaluer objectivement certaines des capacités langagières des apprenants à différents niveaux linguistiques, nous décrivons, en section 3, la méthodologie suivie ainsi que les paramètres étudiés qui pourraient entrer en compte dans l'établissement d'une mesure automatique. La section 4 présente une analyse des résultats ouvrant à la discussion en section 5.

2 Évaluation automatique de productions non-natives

Il existe à ce jour différents outils permettant de mesurer objectivement les productions des apprenants de langue étrangère à différents niveaux linguistiques. Ces mesures peuvent servir d'indicateurs pour déterminer les capacités, et *in fine* les « niveaux » d'habileté, des apprenants. Ces méthodes sont mises en place aussi bien au niveau de la production orale, pour fournir une évaluation de paramètres segmentaux et suprasegmentaux, qu'au niveau de la production écrite pour fournir des évaluations aux niveaux lexical, syntaxique et discursif.

L'évaluation automatique de la prononciation s'effectue généralement de deux manières. La première est basée sur le calcul de paramètres phonémiques et prosodiques, la deuxième sur le calcul des scores de reconnaissance des phonèmes obtenus à l'issue de l'utilisation d'un système de reconnaissance automatique de la parole, basé sur des techniques d'alignement libre ou forcé (Detey *et al.*, 2016). Il en est de même pour l'évaluation de la fluence dans le cadre des langues étrangères, que nous

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

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

définissons comme étant « le degré de fluidité du discours sans pauses ni autres marques de disfluence » (Derwing & Munro, 2015, p. 5, notre traduction). En effet, des systèmes de reconnaissance de la parole sont aussi utilisés afin de mesurer des paramètres temporels, tels que le débit de parole et la longueur moyenne des pauses (Cucchiaroni *et al.*, 2000). Étant donné que l'utilisation de tels systèmes connaît des limites, car ils sont dépendants de la langue pour laquelle leurs modèles ont été entraînés, de nouvelles méthodes ont été récemment développées pour mesurer la fluence de manière plus automatique et indépendamment de la langue cible. Nous pouvons citer par exemple l'algorithme présenté par (Fontan *et al.*, 2020), issu de travaux pilotes portant sur l'évaluation automatique de la fluence phonétique d'apprenants japonais de français en tâche de lecture (Fontan *et al.*, 2018; Detey *et al.*, 2020). Ces travaux s'appuient sur la méthode de segmentation Forward-Backward Divergence Segmentation (André-Obrecht, 1988) basée sur la détection de ruptures dans la trajectoire de l'énergie du signal de parole au cours du temps et permettent, en outre, de calculer des variables à partir de pseudo-syllabes (Farinas & Pellegrino, 2001) et de pauses silencieuses, comme le débit de parole ou encore le pourcentage de parole.

Concernant les mesures des capacités lexicales des apprenants, c'est la dimension désignée par le terme de « richesse » qui est généralement évaluée (Laufer & Nation, 1995; Bonvin & Lambelet, 2019). La richesse lexicale se divise en trois sous-dimensions, communément appelées diversité lexicale, densité lexicale et sophistication lexicale. La diversité lexicale caractérise la taille du vocabulaire dans un texte ou un énoncé, et définit ainsi le nombre de mots différents produits par un locuteur. Différentes méthodes existent pour la mesurer, dont les plus connues sont l'index de Guiraud (Guiraud, 1959) et le Type-Token Ratio, ou TTR (Daller *et al.*, 2003). La densité lexicale, contrairement à la diversité lexicale, se concentre essentiellement sur la production de mots lexicaux, et en calcule la proportion dans un énoncé. Un mot lexical, ou mot à sens plein, est un mot dont la catégorie est un verbe, un nom, un adjectif ou un adverbe. La sophistication lexicale est définie par le nombre de mots lexicaux produits par un locuteur qui relèvent d'une connaissance ou d'une pratique plus avancée de la langue, définie par l'emploi de mots rares ou peu fréquents (« décapotable » à la place de « voiture » par exemple). La notion de fréquence des mots dans la langue, et notamment à l'oral, est donc une information indispensable pour cette mesure. Par ailleurs, il existe une différence d'utilisation de mots spécialisés ou rares selon les niveaux des locuteurs non-natifs (Ovtcharov *et al.*, 2006). Pour rendre compte de ces tendances dans la langue française, nous avons à disposition la base de données Lexique3 (New *et al.*, 2005) qui contient, pour plus de 135 000 entrées lexicales³, diverses informations comme les différentes formes lexicales, les lemmes, et surtout les fréquences d'occurrence. Cette base de données a été établie à partir de corpus de livres et de sous-titres de films, elle est donc aussi bien adaptée pour l'évaluation de l'écrit que de l'oral.

Le terme d'analyse syntaxique, quant à lui, fait souvent référence à l'analyse de la complexité syntaxique d'un énoncé. L'unité la plus courante sur laquelle la complexité syntaxique est objectivement mesurée est la phrase, même si la réalité de la syntaxe de l'oral est plus complexe à traiter que celle de l'écrit (Rossi-Gensane, 2010). Parmi les mesures communément utilisées, on trouve le nombre moyen de mots par phrase, le nombre moyen de propositions coordonnées et subordonnées, et la profondeur moyenne des arbres syntaxiques (Lahuerta Martínez, 2018; Blache, 2010).

Enfin, en ce qui concerne l'évaluation de la compétence discursive dans le cadre de l'apprentissage des langues, il est bien souvent question d'évaluer la cohésion du discours. L'utilisation d'éléments de liaison, ou connecteurs (« ensuite », « de plus », etc.), contribue à la structuration et à l'articulation du discours, et permet de relier les propositions qui le composent (Beacco *et al.*, 2004, p. 171).

3. http://lexique.org/_documentation/Manuel_Lexique.3.pdf

3 Méthodologie

3.1 Procédure générale

Notre étude repose principalement sur l'utilisation d'outils de mesures automatiques appliqués à un corpus de productions orales réalisées par des apprenants japonais de français, dont les niveaux CECRL s'échelonnaient entre A2 et C2. Ces productions ont été enregistrées en situation d'interaction orale (entretiens semi-dirigés) avec un locuteur natif du français. Les mesures obtenues automatiquement à partir du signal audio et de sa transcription sont comparées aux niveaux, afin d'identifier les plus pertinentes. Les mesures automatiques sont enfin combinées pour prédire le niveau CECRL des apprenants.

3.2 Corpus

Les données d'apprenants employées dans notre étude sont issues du corpus CLIJAF⁴, constituées dans le cadre méthodologique général du projet IPFC (Detey & Kawaguchi, 2008; Racine *et al.*, 2012). Plus précisément, nous analysons un sous-ensemble de CLIJAF, correspondant à une tâche de production orale semi-spontanée d'étudiants japonais apprenant le français en université au Japon. Cette tâche consistait en un entretien individuel semi-directif en français, mené par un locuteur natif (enseignant ou doctorant), durant lequel ont été posées 12 questions. Ces entretiens ont été enregistrés, segmentés manuellement en tours de parole et transcrits orthographiquement.

Pour cette étude, et parmi les 12 questions posées, nous avons sélectionné six questions permettant de faire ressortir plus de compétences linguistiques (exemple : « *Quelles sont les principales différences culturelles ou sociales entre la France et le Japon ?* » amène à une réponse plus longue et met plus en avant les compétences linguistiques comparé à « *Quelles langues parles-tu ?* »). Les questions posées sont communes à tous les apprenants, permettant ainsi d'analyser des données liées à un même contexte. Le sous-ensemble du corpus CLIJAF considéré ici contient environ une heure et 20 minutes de données enregistrées par 18 apprenants (15F, dont un A1, quatre B1, huit B2, trois C1 et deux C2), venant de quatre universités japonaises différentes. Les données ont été enregistrées dans les universités respectives des étudiants, soit en studio d'enregistrement, soit dans des salles de classe calmes. Les données audio ont été transcrites orthographiquement en conservant les informations temporelles. Une annotation a été réalisée par un des locuteurs natifs ayant participé aux entretiens à l'aide du logiciel Transcriber⁵ en termes de (1) correction lexicogrammaticale (exemple : production de *« *je suis allé le Tour Eiffel* » avec « *le* » annoté comme « incorrect » et associé à l'annotation corrective « *à la* ») et (2) de divergence phonético-phonologique (exemple : « *pâtisserie* » réalisée avec une battue [r] au lieu d'une uvulaire [ʁ] et l'item lexical ainsi annoté comme contenant une divergence consonantique mais sans transcription phonétique précise à ce stade de l'annotation). Afin de pouvoir être exploitable, le corpus respecte en partie les exigences en la matière (Durand *et al.*, 2014) en incluant les métadonnées des apprenants (profils linguistiques) et en offrant des possibilités d'exploitation *via* un concordancier dédié (Detey *et al.*, 2018).

Les fichiers audio ont été segmentés en tours de parole, le tour de parole étant ici la première réponse produite par l'apprenant en adéquation avec la question posée. Pour l'évaluation syntaxique, nous

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

5. <https://transcriber.fr.softonic.com>

considérons le tour de parole comme unité d'analyse. Nous disposons ainsi de six fichiers audio et texte par apprenant, un par réponse à chaque question considérée (enregistrement et transcription). Pour l'évaluation de la prononciation, de la fluence, du lexique et du discours, nous avons groupé les tours de parole de chaque étudiant afin de constituer une seule et même unité d'analyse et ainsi obtenir une mesure globale par apprenant.

3.3 Évaluation phonético-phonologique

3.3.1 Mesures acoustico-phonétiques et identification phonémique

L'évaluation de la prononciation des apprenants a été effectuée par le biais d'un système de reconnaissance automatique de la parole. Les modèles acoustiques de ce système ont été entraînés sur 340h de données issues de différents corpus audio français, une phase d'augmentation de données ayant été également réalisée (Heba, 2021). Nous avons utilisé ce système *via* la plateforme Paty⁶ (Plateforme de Parole Atypique), développée en partenariat par l'IRIT et le LPL. Cette plateforme permet, suivant les modèles acoustiques et le type de reconnaissance choisis (alignement libre ou forcé), de traiter un fichier audio et de fournir la transcription phonétique ou orthographique. Le résultat de la reconnaissance associe un indice de confiance (IC) à chaque unité décodée, sous la forme d'une probabilité. Un IC moyen est ainsi calculé pour chaque apprenant.

3.3.2 Évaluation de la fluence

Pour l'évaluation de la fluence phonétique, nous avons utilisé l'outil décrit en section 2 (Fontan *et al.*, 2020). À partir des frontières des segments de signal audio détectés et de leur énergie, l'algorithme identifie des pseudo-syllabes et des pauses silencieuses. Ainsi, à partir de ces éléments, et pour chacun des 18 enregistrements (regroupant les six réponses), nous avons calculé quatre paramètres, dont les trois premiers sont fonction de la durée totale du fichier audio traité. Il s'agit du débit de parole (nombre de pseudo-syllabes divisé par la durée de l'enregistrement, DP), du pourcentage de parole (PP), de l'écart-type de la durée des pseudo-syllabes (ETD) et du nombre normalisé de pauses silencieuses (NPS).

Dans ces travaux, ces quatre paramètres avaient été utilisés pour calculer une régression linéaire multiple afin de prédire la fluence phonétique d'étudiants anglophones apprenant le français, pour une tâche de lecture oralisée. Bien que le contexte de nos travaux soit différent des points de vue de la langue maternelle des apprenants et le type de tâche, nous avons utilisé la même équation de régression basée sur ces quatre paramètres. À partir des données extraites des 18 enregistrements à disposition, nous avons obtenu des scores de fluence que nous avons normalisés (NF) pour être compris entre 0 (fluence la plus basse) et 1 (fluence la plus haute).

6. <https://paty.irit.fr/demo/>

3.4 Évaluation lexicale, syntaxique et discursive

3.4.1 Richesse lexicale

Diversité lexicale (IG) : nous avons choisi d'implémenter l'index de Guiraud (IG) pour mesurer la diversité lexicale des productions orales des apprenants, car il est reconnu comme étant plus stable que le TTR pour des énoncés de longueurs variables (van Hout & Vermeer, 2007). Son calcul est donné par la formule $IG = \frac{V}{\sqrt{N}}$, avec V représentant le nombre de mots distincts utilisés, et N le nombre total de mots présents dans l'énoncé traité.

Densité lexicale (DL) : pour calculer le pourcentage de mots lexicaux produits, nous utilisons la formule $DL = \frac{V_{sp}}{N} \times 100$, avec V_{sp} représentant le nombre de mots à sens plein (ou mots lexicaux) distincts utilisés.

Sophistication lexicale (SL) : étant donné que cette étude s'inscrit dans le contexte de la production orale, nous avons choisi d'implémenter notre mesure de sophistication lexicale en utilisant les fréquences des lemmes compris dans la base de données Lexique3, et obtenus grâce aux corpus de sous-titres de films. Plus un lemme est fréquent dans les sous-titres des films, plus il est considéré comme fréquent en français. Pour obtenir SL, nous calculons la proportion de mots lexicaux dont la fréquence de lemme est inférieure à 10, ce qui est considéré comme rare dans la base de données Lexique3. Nous appliquons ainsi la formule $SL = \frac{V_{10}}{N} \times 100$, avec V_{10} représentant le nombre de mots lexicaux ayant une fréquence de lemme inférieure à 10 dans Lexique3.

3.4.2 Complexité syntaxique

Comme énoncé dans la section 3.2, l'unité sur laquelle nous mesurons la complexité syntaxique est le tour de parole. Nous proposons ici de concentrer nos mesures sur la longueur moyenne et la profondeur moyenne de l'arbre syntaxique (arbre de dépendance) d'un tour de parole. Pour chaque apprenant, un score moyen de longueur de tours de parole (LP) est mesuré *via* la formule $LP = \frac{1}{6} \sum_{i=0}^6 len(TP_i)$, avec $len(TP_i)$ la longueur du tour de parole i . Nous incluons aussi les hésitations comme *eah*, *mh* et les amorces inachevées.

Il en est de même pour la profondeur moyenne des arbres syntaxiques (AS), que nous mesurons avec la formule $AS = \frac{1}{6} \sum_{i=0}^6 Prof(A_i)$, avec $Prof(A_i)$ la profondeur moyenne de l'arbre syntaxique du tour de parole i . Il est à noter que la profondeur moyenne d'un arbre est calculée en sommant d'abord les profondeurs des feuilles, puis en divisant cette somme par le nombre de feuilles de l'arbre.

3.4.3 Cohésion du discours

Concernant l'évaluation de la cohésion du discours, nous nous intéressons au nombre de connecteurs du discours employés par chaque étudiant. Nous utilisons la liste LEXCONN (Roze *et al.*, 2012) qui fournit 431 connecteurs du discours pour le français. Pour chaque production des apprenants, nous calculons une proportion de connecteurs du discours (CD) utilisés. Pour se rapprocher des critères utilisés dans le cadre du CECRL concernant la cohésion du discours, nous calculons aussi la diversité des connecteurs (DC) employés, en adaptant la formule de IG décrite plus haut.

4 Résultats

4.1 Validation de nos mesures automatiques

Les résultats des différentes mesures phonético-phonologiques (IC, DP, PP, ETD, NPS, NF), lexicales (IG, DL, SL), syntaxiques (LP, AS) et discursives (CD, DC) ont été soumis au test de Kruskal-Wallis. Les résultats obtenus pour la diversité lexicale ($H(4) = 13.4, p < .01$) et la proportion de connecteurs du discours ($H(4) = 10.53, p < .05$) montrent que le niveau CECRL a un effet significatif sur ces deux mesures linguistiques⁷. L'évolution de ces mesures selon le niveau CECRL est représenté sur les Figures 1 et 2⁸.

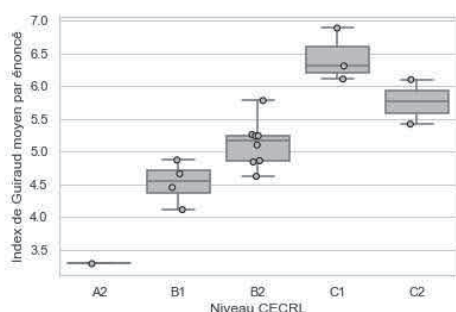


FIGURE 1 – Diversité lexicale (IG)

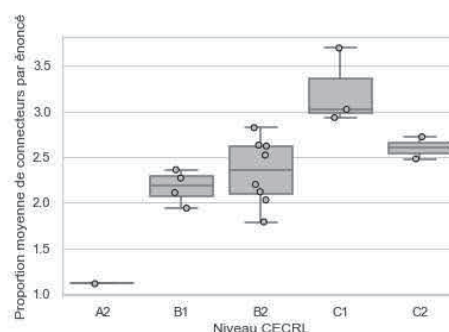


FIGURE 2 – Connecteurs du discours (CD)

Ces résultats sont cohérents avec l'hypothèse selon laquelle plus un apprenant a un niveau élevé, plus il/elle aura tendance à produire un discours diversifié contenant plus de connecteurs du discours. Ces premiers résultats sont encourageants car ils semblent ouvrir sur la possibilité de proposer une mesure qui soit en lien avec le niveau d'habileté des apprenants japonophones, notre objectif étant de pouvoir par la suite le prédire.

4.2 Prédiction du niveau des apprenants

Afin de valider cette hypothèse, nous avons utilisé l'algorithme d'apprentissage automatique non-supervisé *k-means*, dit de « clustering ». Chaque apprenant japonophone est représenté par un jeu de six paramètres, parmi les 13 que nous avons mesurés, extraits de ses interventions : IG, LP, CD, PP, DP et NF. Ce choix a été fait après une phase d'essai de toutes les combinaisons possibles, comprenant au moins cinq paramètres (pour avoir au moins un paramètre par niveau linguistique). Nous avons appliqué l'algorithme de *k-means* pour regrouper les apprenants en cinq groupes, idéalement, un groupe par niveau. Après une analyse en composantes principales pour réduire cette représentation à deux dimensions, nous obtenons la répartition présentée en Figure 3, avec un score de pureté de 0,88.

Nous pouvons observer que les apprenants de niveaux A2 et B1 ont été parfaitement regroupés en

7. Autres résultats : <https://github.com/vdefino31/JEP2022/blob/main/Kruskal-Wallis.png>

8. Pour visualiser l'évolution des autres mesures : <https://github.com/vdefino31/JEP2022>

deux groupes distincts. Il en est de même pour le niveau C1. Nous remarquons une confusion de groupe seulement pour les niveaux B2 et C2.

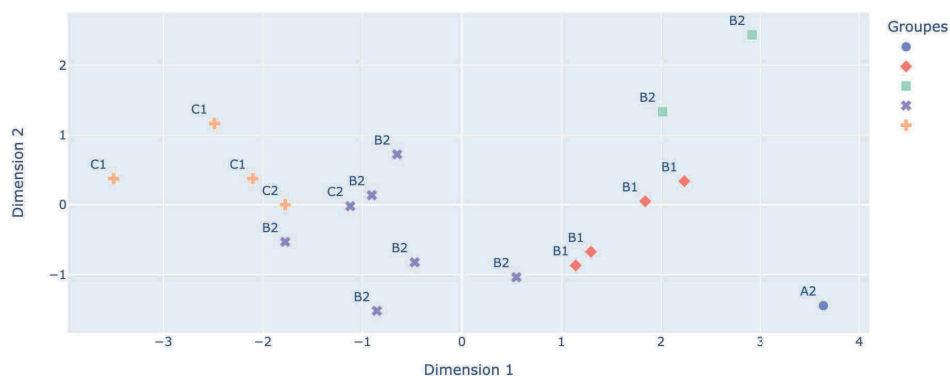


FIGURE 3 – Coordonnées des 18 participants selon les deux composantes principales résultant du partitionnement en *k-means*

5 Discussion et conclusion

Nous obtenons des résultats très satisfaisants concernant les paramètres IG, LP, AS, CD, DP, NPS et IC. Toutefois, les niveaux C1 et C2 ne semblent pas se comporter comme attendu : le groupe C1 obtient en moyenne de meilleurs scores que le C2. Ceci peut refléter une certaine disparité entre l'évaluation d'un niveau global et celle d'habiletés plus spécifiques, telles que la prononciation ou la production orale. Ainsi, dans les informations dont nous disposons, un apprenant de niveau certifié globalement C1 ne va pas systématiquement avoir un niveau C1 en production orale. Nos conclusions doivent donc être modulées par le fait que le volume de données à disposition pour cette première étude est réduit et ne nous permet pas de généraliser. Cependant, les pistes d'investigation ouvertes sont prometteuses. Il en va de même pour les résultats obtenus par l'algorithme de clustering. Les résultats laissent entrevoir la possibilité de prédire le niveau des apprenants, malgré les limites évoquées précédemment. Il est ainsi important de souligner que la combinaison de paramètres utilisée avec l'algorithme du *k-means* est basée sur des paramètres pouvant être évalués automatiquement, sans forcément recourir à une connaissance *a priori* de la langue cible.

À partir d'un ensemble de mesures obtenues automatiquement et relatives à différents niveaux, acoustiques et linguistiques, nous avons montré que certaines évoluent selon le niveau certifié CECL des apprenants japonophones. De plus, l'utilisation de six paramètres automatiques pour prédire le niveau des apprenants, à l'aide d'un algorithme d'apprentissage automatique non supervisé s'est montrée très concluante à l'échelle de notre étude pilote, avec un score de pureté de 0,88. Nos résultats sont d'autant plus prometteurs que les paramètres utilisés ne dépendent pas de la langue cible. Notre méthode pourrait donc se révéler pertinente pour l'évaluation automatique des productions orales réalisées dans d'autres langues cibles. Nous envisageons pour la suite de nos travaux d'enrichir le corpus grâce à l'évaluation, par des enseignants-experts, du niveau en production orale des apprenants afin de pouvoir comparer nos mesures à des évaluations de terrain.

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.
- BONVIN A. & LAMBELET A. (2019). Exploration empirique de la richesse lexicale : la perception humaine. *Linguistik Online*, **100**(77), 65–94.
- CONSEIL DE L'EUROPE (2001). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer (CECR)*. Paris : Didier.
- CUCCHIARINI C., STRIK H. & BOVES L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, **107**, 989–99.
- DALLER M., HOUT R. & TREFFERS-DALLER J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, **24**, 197–222.
- DERWING T. M. & MUNRO M. J. (2015). *Pronunciation Fundamentals : Evidence-based perspectives for L2 teaching and research*. Amsterdam : John Benjamins.
- DETEY S., FONTAN L., LE COZ M. & JMEL S. (2020). Computer-assisted assessment of phonetic fluency in a second language : a longitudinal study of Japanese learners of French. *Speech Communication*, **125**, 69–79.
- 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.
- DETEY S., LE COZ M., FONTAN L., BARCAT C., KAWAGUCHI Y., AKIHIRO H., SUGIYAMA K. & KONDO N. (2018). Annotations minimales multi-niveaux d'un corpus de parole spontanée d'apprenants japonais de FLE et traitement automatique : perspectives didactiques. In *FLORAL-IPFC2018 : contact de langues et (inter)phonologie de corpus*, Paris : MSH.
- DURAND J., GUT U. & KRISTOFFERSEN G. (2014). *The Oxford handbook of corpus phonology*. Oxford : Oxford University Press.
- 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.
- FONTAN L., LE COZ M. & DETEY S. (2018). Automatically Measuring L2 Speech Fluency without the Need of ASR : A Proof-of-concept Study with Japanese Learners of French. In *Proc. Interspeech 2018*, p. 2544–2548.

- GUIRAUD P. (1959). *Problèmes et méthodes de la statistique linguistique*, volume 2. Dodrecht : D. Reidel.
- HEBA A. (2021). *Reconnaissance automatique de la parole à large vocabulaire : des approches hybrides aux approches End-to-End*. Thèse de doctorat, Université Toulouse 3 Paul Sabatier.
- LAHUERTA MARTÍNEZ A. C. (2018). Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. *Assessing Writing*, **35**, 1–11.
- LAUFER 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 Lexique 3. *Behavior Research Methods, Instruments, & Computers*, **36**(3), 516–524.
- OVTCHAROV V., COBB T. & HALTER R. (2006). La richesse lexicale des productions orales : Mesure fiable du niveau de compétence langagière. *Canadian Modern Language Review-revue Canadienne Des Langues Vivantes*, **63**, 107–125.
- 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.
- ROSSI-GENSANE N. (2010). Oralité, syntaxe et discours. In S. DETEY, J. DURAND, B. LAKS & C. LYCHE, Eds., *Les variétés du français parlé dans l'espace francophone : ressources pour l'enseignement*, p. 83–106. Paris : Ophrys.
- ROZE C., DANLOS L. & MULLER P. (2012). LEXCONN : a French lexicon of discourse connectives. *Discours - Revue de linguistique, psycholinguistique et informatique*.
- VAN HOUT R. & VERMEER A. (2007). Comparing measures of lexical richness. In H. DALLER, J. MILTON & J. TREFFERS-DALLER, Eds., *Modelling and assessing vocabulary knowledge*, p. 93–116. Cambridge : Cambridge University Press.