

# Vers la compréhension des difficultés de lecture en L2 à travers des paramètres acoustiques et de mouvement des yeux

Sofiya Kobylyanskaya<sup>1,2</sup>, Ioana Vasilescu<sup>1,2</sup>, Laurence Devillers<sup>1,2,3</sup> et Olivier Augereau<sup>4</sup>

<sup>1</sup> LISN-CNRS, Orsay, France

<sup>2</sup> Université Paris-Saclay, Orsay, France

<sup>3</sup> Sorbonne-Université, Paris, France

<sup>4</sup> Lab-STICC, ENIB, Brest, France

**Résumé.** L'étude porte sur l'analyse de l'acquisition de l'anglais langue étrangère L2 à travers deux modalités : la parole et le mouvement des yeux lors d'une expérience de lecture à voix haute. Les expériences ont été menées en France et au Japon ce qui permet de mettre en évidence les influences de la L1 et de la culture sur les performances en L2. Nous présentons les premières tendances observées sur les deux axes d'analyse et proposons des directions pour de futures recherches.

**Mots-clés :** L2 , multimodalité , analyse acoustique , mouvement des yeux.

## 1 Introduction

LeCycl (Learning Cyclotron) est un projet ANR trilatéral entre la France, le Japon et l'Allemagne dont le but est de renforcer le processus d'acquisition et de transfert des connaissances, et plus particulièrement de l'anglais langue étrangère (L2). L'objectif de l'étude (au sens large) est de trouver des corrélations entre les performances en anglais L2, des habitudes de l'enseignement spécifiques à la culture du pays de l'apprenant (la France vs le Japon) et des mesures objectives provenant de l'analyse acoustique et du mouvement des yeux. Afin d'atteindre cet objectif, dans un premier temps, nous nous focalisons sur l'analyse des zones textuelles difficiles lors de la lecture à voix haute. Nous considérons qu'une zone difficile est un extrait de texte d'au moins un mot pouvant contenir un mot inconnu marqué par les participants et/ou dont la lecture se manifeste par des disfluences (pauses, hésitations, reprises). Ensuite, en déterminant les difficultés de lecture, nous envisageons d'implémenter des nudges [13] adaptés aux particularités individuelles des apprenants en leur suggérant des stratégies d'amélioration de leur technique de lecture. Dans cet objectif, trois protocoles expérimentaux (P1,P2,P3) [Table 1] de lecture à voix haute ont été mis en place en France[9] (P1) et au Japon[12] (P2, P3). Dans cet article, nous nous focalisons davantage sur les parties réalisées par l'équipe française (P1, P3 partie 1), comparons les protocoles, décrivons les mesures utilisées et les étapes d'analyse.

**TABLE 1.** Protocoles expérimentaux

	<b>Protocol 1</b>	<b>Protocol 2</b>	<b>Protocol 3 (joint)</b>
<b>Lieu</b>	France	Japon	Japon
<b>Équipe</b>	LISN-CNRS	OMU	LISN-CNRS+OMU
<b># participants</b>	50	20	20
<b>L1</b>	français	japonais	japonais
<b>Tâche</b>	3 txt, QCM, mots inconnus	mangas, mots inc.	P1+P2+manga sans img
<b>Équipement</b>	micro, eye-tracker	micro, eye-tracker	micro, eye-tr.
<b>Durée</b>	30 min	3h	1h : txt+manga sans img 2h : mangas avec img

## 2 Contexte et hypothèses

De nombreuses études abordent la question de l'acquisition d'une L2 via la modélisation de la parole ou des mesures du mouvement des yeux. Nous proposons de fusionner les deux méthodes, l'apport consistant en des expérimentations qui s'appuient sur les deux modalités tout en prenant en compte le facteur culturel. D'un point de vue linguistique, la réalisation des phonèmes de la L2 dépend de celles de la L1 et s'explique par le phénomène du "crible phonologique" [4]. D'un point de vue du mouvement des yeux, on peut estimer la compréhension du texte lu en L2 [1] mais on obtient peu d'informations sur les indices effectifs. Plusieurs études montrent la complémentarité de ces modalités, par exemple, en mettant en avant la corrélation entre la planification de la parole et le mouvement des yeux ainsi qu'entre le temps de l'acquisition de la forme phonologique du mot et sa longueur [7].

Le but de notre étude est d'estimer les difficultés de la lecture à voix haute des textes en L2 en nous basant sur ces métriques et de trouver des corrélations entre les résultats vs L1, l'environnement de l'apprentissage [10], les habitudes de l'enseignement spécifiques à la culture du pays de l'apprenant [5], le type de support d'enseignement [11]. Concrètement, nous vérifions si les différences du système linguistique du français et du japonais font émerger des passages textuels difficiles différents pour les représentants des deux cultures : par exemple, des mots contenant des "th" ou des "h" pour les francophones et les mots contenant des clusters consonantiques pour les japonophones. En plus, nous souhaitons étudier la stratégie dont les participants abordent les difficultés de prononciation ou de compréhension en fonction de leur système éducatif : la lecture syllabique, la recherche du sens dans le contexte gauche/droite/en haut/en bas, des pauses silencieuses/remplies (hésitations), des répétitions, le changement du degré de l'expressivité de la lecture etc. Notre étude porte également sur l'impact du support d'apprentissage sur les performances de lecture et de compréhension. Par exemple, l'utilisation des images et des mangas dans l'enseignement a été efficace dans plusieurs cas [11,14]. Nous explorons plus particulièrement l'impact

du texte vs texte+images (manga), le dernier étant propre à la culture japonaise y compris en situation d'apprentissage.

### 3 Méthodologie

**Public.** L'expérience a été menée auprès d'un public adulte francophone et japonophone natif fait majoritairement d'étudiants âgés de 18 à 30 ans (85% en France et 100% au Japon) et ayant le niveau A1-C2 en anglais. Au moment de la rédaction de l'article, nous avons recueilli des données auprès de 50 participants en France (P1) et 20 au Japon (P3).

**Équipement.** Pour l'enregistrement des participants, nous utilisons le microphone AKG Perception Wireless 45 Sports Set Band-A 500-865 MHz, l'eye-tracker Tobii Nano Pro et le système d'enregistrement du mouvement des yeux et de la voix Eye Got It [2], comme dans [9].

**Support** Deux types de support sont mis en place : des textes en anglais de différents niveaux [9] (P1) et des mangas en anglais traduits du japonais [12] (P2). Le P3 est une fusion des P1 et P2 avec l'utilisation de deux supports comme dans P1 et P2 et d'un extrait de manga sous format texte. Cela permet de comparer les stratégies de lecture et le niveau de compréhension de textes accompagnés ou non d'images. Le choix d'un extrait de manga pour P3 a été basé sur des critères linguistiques. Du point de vue syntaxique, les mangas étudiés contiennent majoritairement des phrases courtes, sans subordonnées relatives ni temps verbaux complexes, car ils représentent le discours direct interactif qui se caractérise par des structures grammaticales simples. Ainsi, les critères suivants ont été pris en compte : la présence du lexique du niveau avancé (C1) permettant d'avoir des supports de niveaux comparables ; la présence du champ lexical (en l'occurrence de la nourriture) afin de faire référence au texte du niveau avancé du P1 ; la présence des clusters consonantiques afin de créer des conditions phonologiques complexes pour les participants japonophones dont la langue maternelle ne comprend que des syllabes type consonne-voyelle (CV).

### 4 Méthode d'analyse

L'objectif courant de l'étude est d'identifier des zones textuelles qui soulèvent des défis linguistiques de sources diverses : phonologique, lexicales, syntaxiques et de mettre en évidence des caractéristiques acoustiques ou physiologiques de leur réalisation par rapport aux zones non marquées comme difficiles (par exemple, le débit ralenti, le nombre de saccades/fixations augmenté/diminué). Nous présentons ensuite les démarches d'analyse des deux modalités : la parole et le mouvement des yeux qui sont étudiées séparément à cette étape de travail.

**Parole.** Tout d'abord, le texte a été aligné avec le son à l'aide du logiciel Web-Maus [8]. Le texte a été manuellement enrichi en disfluences (hésitations, répétitions etc.) afin de correspondre à la production du locuteur par rapport au

texte original et de permettre un alignement fidèle à la prononciation. Une fois le texte aligné, des fenêtres d'analyse de 5 secondes ont été extraites avec les intervalles d'une seconde. La durée de la fenêtre est suffisante pour l'extraction des 88 caractéristiques acoustiques avec OpenSMILE eGeMAPS [3] et 14 caractéristiques prosodiques myprosody<sup>5</sup> et d'observer leur variation d'un segment à l'autre. Les mêmes caractéristiques ont été extraites sur le texte entier afin de trouver des valeurs moyennes qui nous servent de référence. Une autre référence est la lecture d'un locuteur natif de l'anglais américain.

**Mouvement des yeux.** Les caractéristiques du mouvement des yeux ont été extraites également au niveau de la lecture du texte entier et au niveau des extraits de 5 secondes ce qui permet de faire la corrélation avec les caractéristiques acoustiques. De la même manière, les valeurs moyennes du participant et d'un locuteur natif servent de référence. Les valeurs suivantes ont été extraites : le nombre de saccades/fixations, la durée minimale/moyenne/maximale des saccades/fixations, la déviation standard des durées des saccades/fixations.

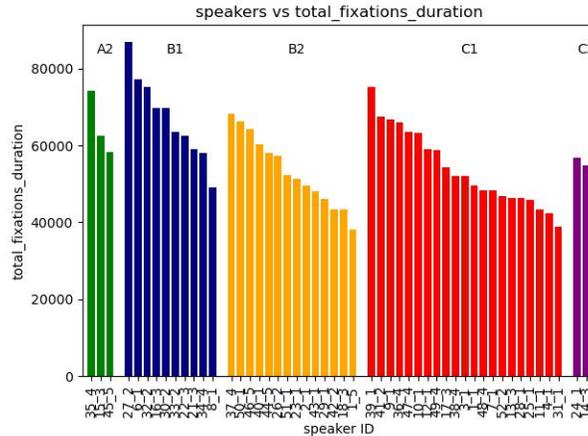
## 5 Résultats préliminaires

Cette section est dédiée aux analyses des premières tendances qui émergent pour ce qui est des deux modalités, parole et mouvement des yeux. Pour l'instant nous avons mis en oeuvre ces mesures sur les données collectées en France (50 participants) et nous évaluerons par la suite les différences culturelles. Nous considérons des valeurs moyennes d'un ensemble de valeurs de mouvement des yeux et de parole en vue de corrélations entre les caractéristiques extraites, le niveau L2 des participants et le niveau des textes lus. En nous basant sur ces observations, nous pourrions ensuite nous focaliser sur la recherche des extraits de textes difficiles et les caractéristiques de leur réalisation.

**Mouvement des yeux.** Au niveau du mouvement des yeux, on remarque que le nombre et la durée totale des fixations diminue avec l'augmentation du niveau de la L2 [Fig.1]. Les participants des niveaux B1 et B2 effectuent plus de saccades et elles sont plus longues que celles des participants des niveaux C1 et C2. Par contre, les saccades sont moins nombreuses et plus courtes pour les participants du niveau A2. Cette tendance nécessite d'être vérifiée auprès d'un plus grand nombre de participants, car seulement 3 participants de notre corpus ont le niveau A2. Les participants des niveaux B1 et B2 utilisent davantage de saccades de régression en lisant les textes des niveaux intermédiaire et avancé. L'hypothèse peut être formulée que les débutants, confrontés à des problèmes de compréhension, se focalisent plutôt sur la prononciation que sur la compréhension, alors que les participants plus avancés comprennent le texte et ont besoin de faire moins de retours en arrière. Concernant les participants intermédiaires, nous supposons que leur niveau leur permet d'avoir une bonne compréhension des textes, mais nécessite des retours en arrière occasionnels afin de clarifier le contexte ou la structure syntaxique.

---

5. <https://github.com/Shahabks/myprosody>



**FIG. 1.** Durée totale de fixations par participant et par niveau pour le texte avancé. Axe x - l’ID du participant, Axe y - la durée des fixations

**Parole.** Du point de vue de la parole, une attention particulière est portée aux variations de la fréquence fondamentale (F0), au débit de la parole et d’articulation, la durée de la parole, le nombre des pauses et de syllabes, car ces caractéristiques sont parmi d’autres qui sont distinctives de la fluence en L2 [6]. On observe une baisse de la F0 avec la progression de la L2 qui peut être due à des raisons linguistiques telles que l’impact de la complexité du texte ou extralinguistiques, telles que les émotions (la montée du stress), le genre, les habitudes culturelles. Dans la plupart des cas (sauf pour A2), le nombre de pauses et de syllabes diminue avec l’avancement du niveau L2. On suppose que les participants ayant des niveaux de langue plus élevés font moins de disfluences lors de la lecture (moins de répétitions et d’hésitations) et donc, moins de syllabes sont détectées par le système automatique. L’intensité de la voix et le nombre de segments voisés (où la F0 est non nulle) augmentent avec le niveau L2 sauf pour le C2 pour lequel nous n’avons que 2 participants. Cela peut refléter la montée du niveau de confiance avec l’avancement de la L2, mais aussi des particularités personnelles non liées à la maîtrise de la langue étrangère.

## 6 Conclusions et discussion

L’article présente les résultats préliminaires d’une analyse multimodale de l’acquisition de la L2 et propose des pistes pour de futures analyses des difficultés de lecture liées à des particularités linguistiques, culturelles et personnelles. Les premières tendances observées confirment l’intérêt d’une analyse jointe des paramètres acoustiques et du mouvement des yeux pour l’étude de l’apprentissage de la L2. En effet, nous observons des corrélations des paramètres étudiés avec le niveau L2 des participants et le niveau de difficulté des textes lus. Ces ten-

dances nécessitent une validation statistique plus approfondie tout comme des données supplémentaires pour confirmer les tendances. À cet effet, nous avons prévu d'augmenter le nombre de participants ayant les niveaux A2 et C2 pour équilibrer la base de données. Les futures étapes du travail seront dédiées à l'extraction des mêmes caractéristiques sur le corpus collecté au Japon et la recherche des tendances, au calcul des paramètres distinctifs des zones textuelles selon un critère de difficulté permettant de corrélérer et de prédire les contextes problématiques et d'adopter des stratégies de remédiation adaptées.

## Références

1. Augereau, O., Fujiyoshi, H., Kise, K. : Towards an automated estimation of english skill via toeic score based on reading analysis. pp. 1285–1290 (12 2016)
2. El Baha, M., Augereau, O., Kobylanskaya, S., Vasilescu, I., Devillers, L. : Eye got it : a system for automatic calculation of the eye-voice span (2022)
3. Eyben, F., Wöllmer, M., Schuller, B. : Opensmile : The munich versatile and fast open-source audio feature extractor. Association for Computing Machinery, New York, NY, USA (2010)
4. Flege, J. : Second language speech learning : Theory, findings and problems, pp. 229–273 (01 1995)
5. Grainger, P. : The impact of cultural background on the choice of language learning strategies in the jfl context. *System* **40**(4), 483–493 (2012)
6. Hsieh, C.N., Zechner, K., Xi, X. : Features Measuring Fluency and Pronunciation, chap. 6, pp. 101–122 (11 2019)
7. Huettig, F., Rommers, J., Meyer, A. : Using the visual world paradigm to study language processing : A review and critical evaluation. *Acta psychologica* **137**, 151–171 (06 2011). <https://doi.org/10.1016/j.actpsy.2010.11.003>
8. Kisler, T., Reichel, U., Schiel, F. : Multilingual processing of speech via web services. *Computer Speech & Language* **45**, 326–347 (2017)
9. Kobylanskaya, S. : Speech and eye tracking features for l2 acquisition : A multimodal experiment. In : Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds.) *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium*. pp. 47–52. Springer International Publishing, Cham (2022)
10. Nikitina, L. : Creating an authentic learning environment in the foreign language classroom. *International Journal of Instruction* **4**(1) (2011)
11. Ranker, J. : Using comic books as read-alouds : Insights on reading instruction from an english as a second language classroom. *The Reading Teacher International Reading Association* **61**, 296–305 (12 2007)
12. Takaïke, T., Iwata, M., Kise, K. : Estimation of unknown words using speech and eye gaze when reading aloud comics. In : *MANPU 2022 The 5th International workshop on comics analysis, processing and understanding* (2022)
13. Thaler, R., Sunstein, C. : *Nudge : Improving decisions about health, wealth, and happiness*. Yale University Press (2008)
14. Themelis, C., Sime, J.A. : *Comics for inclusive, technology-enhanced language learning* (2020)