



**HAL**  
open science

## Suivi longitudinal de la fluence en lecture par évaluation automatique de la parole

Erika Godde, Gérard Bailly, Marie-Line Bosse

### ► To cite this version:

Erika Godde, Gérard Bailly, Marie-Line Bosse. Suivi longitudinal de la fluence en lecture par évaluation automatique de la parole. EIAH 2021 - 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Jun 2021, Fribourg (Virtual), Suisse. pp.70-81. hal-03292753

**HAL Id: hal-03292753**

**<https://hal.archives-ouvertes.fr/hal-03292753>**

Submitted on 23 Jul 2021

**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.

# Suivi longitudinal de la fluence en lecture par évaluation automatique de la parole

Erika Godde<sup>1,2</sup>, Gérard Bailly<sup>1</sup>, Marie-Line Bosse<sup>2</sup>

<sup>1</sup> Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France  
erika.godde@gipsa-lab.fr gerard.bailly@gipsa-lab.fr

<sup>2</sup> Université Grenoble Alpes, CNRS, LPNC, 38000 Grenoble, France  
marie-line.bosse@univ-grenoble-alpes.fr

**Résumé.** La fluence en lecture à haute voix est couramment évaluée en mesurant le nombre de mots correctement lus par minute. Cette évaluation se fait au détriment de l'expressivité et du phrasé, pourtant nécessaire à la compréhension, mais complexes à évaluer car très subjectifs. Nous proposons ici d'évaluer automatiquement les quatre dimensions de la fluence en lecture : décodage, vitesse, phrasé et expressivité. Cet outil s'appuie sur une analyse automatique des paramètres acoustiques, notamment par comparaison à de multiples lecteurs experts, et sur une échelle d'évaluation subjective adaptée pour le français. Nous avons pu montrer que cet outil est plus précis et consistant que des juges humains. Il permet de plus de s'affranchir de la variabilité inter-juges. Il est donc particulièrement adapté à un suivi longitudinal des enfants. Cet outil a été utilisé pour évaluer les lectures de 49 enfants suivis du CE1 au CM1. Ces analyses ont permis de mettre en évidence les profils de développement de chaque dimension de la fluence en lecture.

**Mots-clés :** Fluence en lecture, prosodie en lecture, évaluation automatique, expressivité

**Abstract.** Reading fluency is usually assessed with the number of words correctly read in a minute. Expressivity and phrasing are then omitted, probably because these dimensions are more subjective and complex to assess. We propose here a new tool to assess the four dimensions of reading fluency : pace, smoothness, phrasing and expressivity. This tool use an automatic analysis of acoustic features and multiple expert references to predict the score to a subjective scale. We show that this tool is valid, precise and more consistent than human raters. This tool is then well adapted to multiple assessments. We used it to assess the reading fluency of 49 young readers three times in CE1, CE2 and CM1. These analyses enable us to characterize the development patterns of the four dimensions of reading fluency.

**Keywords:** Reading fluency, reading prosody, automatic assessment, expressivity

## 1 Introduction

La fluence en lecture est couramment évaluée en mesurant le nombre de mots correctement lus par minute (NMCLM). Cependant, cette méthode se concentre sur la mesure du décodage et de la vitesse uniquement, au détriment de la prosodie, expressivité et phrasé. Cette mesure restrictive entretient ainsi la confusion entre lecteur rapide et bon lecteur, créant des lecteurs rapides mais qui ne comprennent pas forcément le texte lu [1]. Cette prévalence du NMCLM dans l'évaluation de la fluence est en partie due à la difficulté d'évaluation des dimensions prosodiques, notamment leurs conséquences sur le phrasé et l'expressivité des lectures. En effet, s'il est aisé d'évaluer la justesse de la prononciation des mots, il y a de nombreuses façons licites de lire un texte avec expressivité.

La prosodie est pourtant une compétence dont l'importance dans l'acquisition de la lecture est de plus en plus reconnue [2][3]. Le lien entre compréhension et prosodie a été montré dans plusieurs langues et à tous les niveaux scolaires [4][5]. Pourtant le développement de la prosodie en lecture a été assez peu étudié [4]. Les études, menées essentiellement en anglais et en espagnol avec des enfants de primaire, montrent que le développement de la prosodie en lecture débute dès les premières années, en parallèle des compétences en décodage et automaticité [6][7]. Le phrasé se développe rapidement du CP au CE2. L'expressivité se met en place plus lentement que les autres dimensions de la fluence et son développement continue au-delà de l'école primaire. Le schéma semble similaire en français. Une étude transversale a montré le développement rapide des dimensions du CE1 au CM2 [8]. Il n'existe à notre connaissance pas d'étude longitudinale en français. Ce type de données permettrait de montrer les étapes de développement de la fluence en lecture et de concevoir des outils d'entraînement adaptés à ces étapes de développement.

Pour mener ces études longitudinales, il est nécessaire d'évaluer la fluence, et notamment ses dimensions prosodiques, de façon fiable à de multiples reprises, pour pouvoir mesurer des variations fines. On trouve deux types d'outils dans la littérature. Tout d'abord les outils d'évaluation subjective. Il s'agit d'échelle de notation sur lesquelles des juges s'appuient pour donner un score à diverses compétences. La plus utilisée dans la littérature est la *Multidimensional Fluency Scale* (MFS) [2]. Elle a été traduite et adaptée pour le français [8]. Cette échelle, que nous appelons *Échelle MultiDimensionnelle de Fluence* (EMDF) en français, permet d'évaluer les quatre dimensions de la fluence : expressivité, phrasé, décodage et vitesse. Les juges donnent un score de 1 à 4 à chacune de ces dimensions. C'est un outil valide, donnant de bons coefficients d'accord entre juges quand le protocole d'utilisation et d'entraînement des juges est strictement respecté [9]. La variabilité inter-juges reste cependant problématique quand on réalise plusieurs évaluations espacées dans le temps. Le deuxième type de mesure de la fluence en lecture est la mesure de paramètres acoustiques [6][7]. Ces mesures sont précises et fiables dans le temps. Elles ne permettent cependant pas de caractériser aisément chaque dimension de la fluence. Dans la littérature, les analyses acoustiques se concentrent essentiellement sur les marqueurs rythmiques, tels que les pauses. Les paramètres mélodiques et d'intensité sont peu exploités [4]. Quand ils le sont, c'est en comparaison avec une lecture

experte unique. Ces analyses se heurtent donc à la problématique de la variété des productions acceptables

Il existe donc un vrai besoin d'outils d'évaluation de la fluence incluant la prosodie permettant un suivi large et fiable des compétences en lecture dans le temps pour la recherche et l'enseignement. Un premier outil proposé par Bolanos et al. [10] utilise à la fois des paramètres verbaux (NMCLM, taux de phonation, taux d'articulation et nombre de répétitions) et acoustiques (localisation de l'accentuation, durée des mots et des syllabes, pauses silencieuses et non silencieuses et lien avec la ponctuation) pour prédire des scores d'enfant de CP au CE2 à une échelle subjective unidimensionnelle. Pour cela, les auteurs utilisent un apprentissage supervisé. Les lectures sont également évaluées à l'aide de cette échelle par deux juges experts. Les auteurs obtiennent de bons résultats avec un coefficient de corrélation entre notes calculées/jugées de 0.86. Ils concluent cependant sur la difficulté de l'évaluation de l'expressivité, dû à un manque d'objectifs simples à atteindre. On peut également noter qu'ils utilisent une échelle unidimensionnelle et ne séparent donc pas les dimensions de prosodie et d'automatisme. La multiplicité des productions acceptables est bien au cœur de la problématique de cette évaluation. Hirst et al. [11] proposent ainsi d'évaluer la prosodie des productions de synthétiseurs vocaux en utilisant une comparaison avec plusieurs références naturelles. Après plusieurs alignements (durée des segments, F0 et intensité), ils utilisent alors comme mesure de la qualité prosodique la valeur moyenne quadratique de la distance entre la production de synthèse et la production naturelle la plus proche. Cette approche ne permet cependant pas de prendre en compte la globalité des productions possibles en se limitant à peu de dimensions et à une seule production cible. Nous avons ici choisi d'utiliser une méthode déjà éprouvée en analyse du regard [12] : la mise à l'échelle multidimensionnelle. Les références sont projetées comme des points dans un espace multidimensionnel. La distance des productions à évaluer à chaque référence est ensuite calculée pour placer ces productions dans l'espace multidimensionnel et obtenir ses coordonnées, caractéristiques de la performance par rapport aux références. Cette méthode utilisée pour l'analyse du regard social peut être transposée pour la lecture.

Dans un premier temps nous présenterons et montrerons la validité d'un nouvel outil automatique de mesure de la fluence en lecture, reposant sur les paradigmes présentés ci-dessus. Cet outil permet de s'affranchir de la variabilité inter évaluateur et de mesurer plus finement chaque dimension de la fluence. Dans un deuxième temps, nous utiliserons cet outil dans le cadre d'une étude longitudinale visant à caractériser le développement des dimensions de la fluence chez les jeunes lecteurs français du CE1 au CM1.

## **2 Méthodologie**

### **2.1 Participants**

Les enfants participants à l'étude sont scolarisés dans deux écoles et un collège de la région grenobloise, soit 14 classes. Ces établissements ne sont pas situés en Réseau

d'Éducation Prioritaire et accueillent un public de milieu socio-économique diversifié. Nous avons obtenu l'aval de la Directrice Académique des Service de l'Éducation Nationale de l'Isère, des Inspectrices de l'Éducation Nationale des circonscriptions concernées, des directrices des écoles et du principal du collège, pour enregistrer les enfants dans les établissements pendant le temps scolaire. Tous les enregistrements ont été réalisés avec l'autorisation écrite des parents et l'accord verbal des enfants.

**Corpus transversal (étalonnage de l'outil automatique).** Les données ont été récoltées auprès d'enfants du CE1 à la 5e en fin d'année scolaire, en mai et juin. Le corpus comprend 646 lectures recueillies auprès de 323 enfants normo-lecteurs dont au moins une langue maternelle est le français. Vingt adultes lecteurs experts ont également été enregistrés. Deux textes ont été lus : un texte narratif et un dialogue.

**Corpus longitudinal.** Pour ce corpus 49 enfants ont été enregistrés en CE1, CE2 et CM1 sur la lecture des deux textes utilisés dans le corpus transversal.

Le Tableau 1 présente les caractéristiques des participants.

Tableau 1. Caractéristiques des participants au corpus transversal et longitudinal. Age moyen (SD) en année:mois(mois).

transversale	CE1	CE2	CM1	CM2	6 <sup>e</sup>	5 <sup>e</sup>	adultes
sujets	61	47	54	63	33	37	20
filles	27	25	26	23	15	21	10
age	7:11 (3.8)	8:11 (3.4)	9:11 (5.1)	10:11 (4.5)	11:11 (5.4)	12:8 (5.4)	29:6 (85)
longitudinale		CE1	CE2	CM1			
sujets		49	49	49			
filles		23	23	23			
Age		7:11(3.8)	8:7(3.8)	9:1(3.8)			

## 2.2 Évaluation subjective

Les 744 lectures du corpus transversal ont été évaluées subjectivement en utilisant l'EMDF [8]. Dans cette étude, vingt-sept juges, 13 enseignants, 4 chercheurs et 10 étudiants, ont évalué 744 lectures au total. Nous disposons de 6853 évaluations au total.

## 2.3 Évaluation objective

L'évaluation objective des lectures se basent sur des paramètres verbaux et non-verbaux. Pour ces analyses acoustiques, l'enregistrement audio est aligné automatiquement avec le texte en utilisant un dictionnaire mis à jour au fil des nouvelles lectures contenant toutes les formes des mots rencontrés, correctes et incorrectes.

**Caractérisation des paramètres verbaux.** Les paramètres verbaux suivants sont extraits de chaque lecture du corpus transversal :

- nombre de mots incorrects par minute (NMI)
- nombre de mots lus par minute (NMLM)
- nombre de répétition par minute (NR)
- nombre de pauses intra-mots par minute (NPIM)
- nombre de noyaux vocaliques par minute (NV)

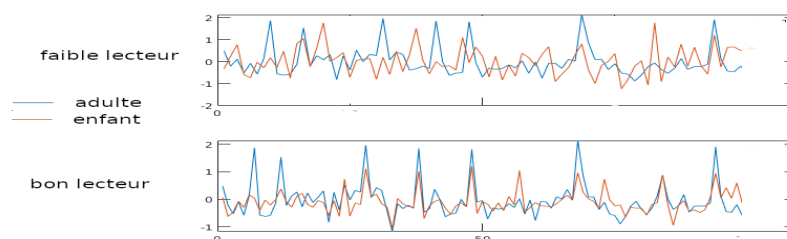


Fig. 1. Alignement de lecture d'enfants (rouge) sur une lecture de référence adulte (bleu) pour l'allongement syllabique. Les pics d'allongements sont dus aux pauses.

**Caractérisation des schémas prosodiques.** Pour prendre en compte la multiplicité des productions acceptables en termes de prosodie, nous avons choisi de comparer les lectures d'enfants à plusieurs productions de lecteurs experts : les lectures de références sont projetées comme des points dans un espace tridimensionnel. Deux performances sont prises en compte : les schémas mélodiques (basés sur les variations de F0) et les schémas rythmiques (basés sur l'allongement des syllabes). Les lectures des enfants sont ensuite projetées dans l'espace de référence par calcul de leur distance à chaque lecture de référence. Pour cela, les lectures sont alignées (seuls les mots correctement lus sont considérés) et la distance cumulée entre les motifs mélodiques (vs. rythmiques) des syllabes de l'énoncé est calculée (la moyenne des motifs est soustraite (ce sont des échelles logarithmiques) afin de s'affranchir du registre vs. de la vitesse de lecture). La matrice des distances entre la lecture de l'enfant aux lectures adultes et alors projetée dans l'espace de référence. La Fig. 1 présente ces alignements pour un bon lecteur et un lecteur moyen. Finalement les coordonnées des lectures d'enfants dans ces espaces sont utilisées pour caractériser les performances prosodiques des enfants : F0x pour l'espace mélodique et COEx pour l'espace rythmique. La Fig. 2 présente les projections dans le premier plan factoriel pour chacun des schémas prosodiques. On peut voir sur la Fig. 2 les clusters correspondant à différents profils de lecteurs, évalués avec l'EMDF. On constate que chaque profil est placé comme attendu par rapport aux lecteurs de référence : on retrouve bien le groupe des bons lecteurs proches des lecteurs de références et les groupe des faibles lecteurs plus éloignés. Ces faibles lecteurs sont également plus proches de ref0), lecteur fictif avec une prosodie « plate » (sans variation de F0 ni de durée syllabique).

## 2.4 Liens entre paramètres objectifs et évaluations subjectives.

Nous avons réalisé des régressions linéaires multiples pour déterminer les relations entre scores subjectifs et paramètres verbaux, puis paramètres verbaux et prosodiques (cf. Tableau 2).

On peut constater que les corrélations entre scores subjectifs et paramètres acoustiques augmentent sensiblement avec l'introduction des paramètres prosodiques dans les régressions, particulièrement pour la dimension expressivité par exemple (.62 à .79 pour le texte 1) et pour la dimension phrasé (.77 à .83 pour le texte 1). L'introduction des paramètres prosodiques apporte donc une information supplémentaire pertinente pour la modélisation des scores subjectifs de fluence. On constate également que les paramètres les plus significatifs pour l'expressivité sont les coefficients du schémas mélodiques et pour le phrasé les coefficients du schéma rythmique. Ces résultats conforte le choix de ces schémas pour la caractérisation de la qualité prosodique des lectures. Finalement, les corrélations sont élevées (entre .8 et .9) ce qui confirme la qualité des modèles présentés en table 2.

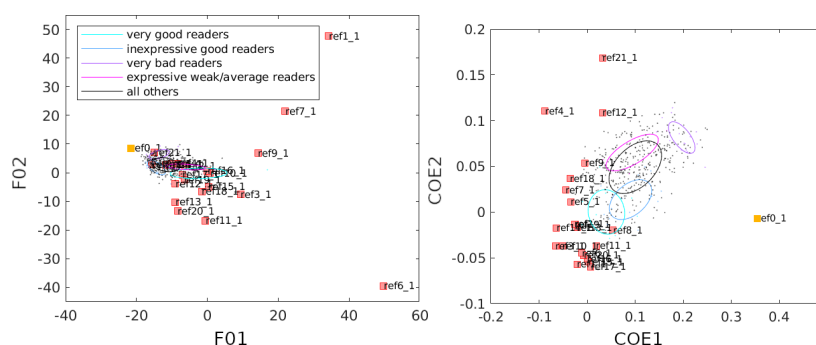


Fig. 2. Projection des lectures d'enfants (points noirs) dans les espaces des lectures de référence (points rouges). Les structures mélodiques et rythmiques sont projetées respectivement à gauche et à droite. Notez la plus grande dispersion des structures rythmiques réalisées par les enfants, assez éloignées de la réalisation monotone (ref0\_1).

## 2.5 Dépendance au texte.

L'évaluation subjective a été réalisée sur les deux textes enregistrés. On constate que les modèles obtenus (cf. 2) pour les deux textes sont proches pour les paramètres verbaux, les différences proviennent principalement des paramètres prosodiques. On peut expliquer cela par la nature différente des textes. Le texte 2 est un dialogue, induisant plus d'expressivité et de rythme qu'un texte narratif comme le texte 1. C'est bien ce qui apparaît dans les modèles présentés dans le 2. Ainsi les modèles peuvent donc différer d'un texte à l'autre mais restent cohérents dans leur structure globale combinant paramètres verbaux et prosodiques.

Tableau 2. Prédicteurs principaux des scores aux évaluations subjectives et corrélation (pseudo R2 de McFaden) pour les régresseurs avec paramètres verbaux et paramètres verbaux et prosodiques. Seuls les régresseurs significatifs sont présentés dans l'ordre d'importance de leur contribution au modèle.

	Paramètres verbaux		Paramètres verbaux + prosodiques	
<b>Texte 1</b>				
expressivité	.62	1 + NV	.79	1 + F01 + COE1 + NMLM
phrasé	.77	1 + NMI + NV	.83	1 + NMLM + COE1
décodage	.75	1 + NMI + NMLM	.76	1 + NMLM + NMI
vitesse	.82	1 + NMLM	.84	1 + NMLM
<b>Texte 2</b>				
expressivité	.65	1 + NV	.83	1 + COE1 + NMLM + F02 + F01
phrasé	.81	1 + NMI + NV	.91	1 + COE1 + NV + COE3 + NMI + F03
décodage	.86	1 + NMI + NV	.88	1 + NV + NMI + COE1
vitesse	.89	1 + NV	.91	1 + NV + COE1

La Fig. 3 compare les deux textes pour les juges humains et les scores prédits par les modèles, pour chacune des dimensions. Les juges humains tendent à mieux noter le texte 2 que le texte 1. Les scores prédits sont plus proches avec un léger avantage pour le texte 1. Avec un R2 significativement plus élevé, on observe également que les scores prédits sont plus consistants que les scores des juges humains.

Ainsi la méthode de prédiction des scores présentés ici peut l'être avec différents types de texte. Les textes peuvent induire quelques différences, notamment d'expressivité, ces différences étant cependant moins importantes qu'avec des juges humains.

Le processus décrit ci-dessus (alignement, extraction des paramètres verbaux et prosodiques, calculs des scores à chaque dimension à partir des modèles) a été ensuite automatisé pour évaluer directement les lectures à partir des données audio brutes.

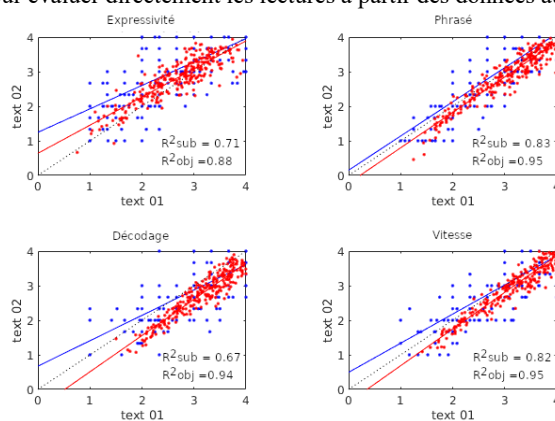


Fig. 3. Scores au texte 2 en fonction des scores au texte 1 pour les dimensions de phrasé et d'expressivité. Scores subjectifs (bleu) et scores calculés (rouge).

### 3 Application à une étude longitudinale

#### 3.1 Analyses statistiques

Nous analysons ici les données du corpus longitudinal. Les analyses ont été réalisées avec le logiciel R. L'effet du niveau a été testé en utilisant des modèles linéaires multiniveaux de croissance, en utilisant le sujet comme paramètre aléatoire de niveau 2 et le parcours comme paramètre aléatoire de niveau 3, avec la fonction lmer du package lme4 (Bates et al., 2015). Les participants sont répartis sur plusieurs classes chaque année (quatre classes en CE1, cinq en CE2 et six en CM1). Cette répartition évolue chaque année. Les sujets suivent donc des parcours différents pendant ces trois années. Au total, les participants se répartissent sur 12 parcours du CE1 au CE2 (année 2) et sur 25 parcours du CE1 au CM1 (année 3). Pour chaque modèle, l'effet aléatoire de niveau 3 est significatif. Il est donc conservé pour la suite des analyses. Quand l'effet du niveau est significatif, une analyse post-hoc a été menée pour déterminer les différences entre les niveaux deux à deux. Nous avons utilisé le test HSD de Tukey. Pour cette analyse nous rapportons la p-value la plus élevée parmi les tests deux à deux. Nous avons également testé l'effet du texte et son interaction avec le niveau.

Le tableau 3 présente les scores obtenus pour chaque niveau et chaque texte les scores aux dimensions de l'EMDF. La 4 présente les modèles de croissance pour chacune des dimensions de la fluence pour chacun des textes évalués.

Tableau 3. Scores moyens (écart-type) prédits à l'EMDF pour le texte 1 par l'outil automatique pour chaque année du corpus longitudinal.

année	Texte 1			Texte2		
	1	2	3	1	2	3
niveau	CE1	CE2	CM1	CE1	CE2	CM1
Expressivité	2.16(0.53)	2.35(0.53)	2.57(0.52)	2.22(0.37)	2.55(0.42)	2.82(0.46)
Phrasé	2.35(0.49)	2.71(0.50)	2.99(0.50)	2.21(0.44)	2.71(0.43)	3.03(0.42)
Décodage	2.62(0.43)	2.92(0.39)	3.10(0.40)	2.19(0.33)	2.60(0.35)	2.91(0.37)
Vitesse	2.53(0.49)	2.98(0.49)	3.24(0.48)	2.32(0.41)	2.83(0.44)	3.14(0.40)

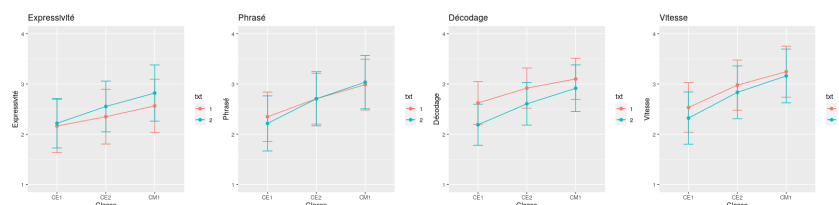


Fig. 4. Modèle de croissance du CE1 au CM1 pour chacune des quatre dimensions de la fluence pour un texte narratif (rouge) et un dialogue (bleu)

### 3.2 Expressivité

Le modèle de croissance montre un effet fixe positif du niveau sur le score à la dimension expressivité. L'ajout du niveau comme paramètre aléatoire de niveau 2 n'améliore pas notre modèle. Il n'y a donc pas de variance des pentes entre les élèves. On observe par contre un effet d'interaction entre niveau et texte. Le texte 2 a un taux de croissance légèrement plus élevé pour le texte 2, 0.08 points en plus par an par rapport au texte 1. L'analyse post-hoc de Tukey confirme la croissance progressive de l'expressivité avec des scores significativement différents pour chaque année ( $p < .001$ ). Les scores à la dimension expressivité augmentent en moyenne de 0.25 points par an.

### 3.3 Phrasé

Le modèle de croissance montre un effet fixe positif du niveau sur le score à la dimension phrasé. L'ajout du niveau comme paramètre aléatoire de niveau 2 améliore notre modèle. On observe donc un effet du niveau et une variance des pentes entre élèves. Le rythme de progression est donc différent entre les élèves. L'interaction entre niveau et texte améliore également notre modèle. On constate à la fois un effet du texte sur les scores, significativement plus faibles pour le texte 2 (-0.23 points), et sur le taux de croissance légèrement plus élevé pour le texte 2 (0.07 points en plus par an). Comme on peut le constater sur la figure 4, la croissance des scores à la dimension phrasé semble assez linéaire du CE1 au CM1. L'analyse post-hoc de Tukey confirme la croissance progressive de l'expressivité avec des scores significativement différents pour chaque année ( $p < .001$ ). Les scores à la dimension expressivité augmentent en moyenne de 0.37 points par an.

### 3.4 Décodage

Le modèle de croissance montre un effet fixe positif du niveau sur le score à la dimension décodage. L'ajout du niveau comme paramètre aléatoire de niveau 2 n'améliore pas notre modèle. On observe donc un effet du niveau mais pas de différence entre élèves dans les rythmes d'acquisition. L'ajout de l'interaction du niveau avec le texte améliore grandement notre modèle. Les scores du texte 2 sont significativement inférieurs à ceux du texte 1 (-0.57 points) et son taux de croissance significativement supérieur (0.10 points de plus par an). Comme on peut le constater sur la figure 4, la croissance des scores à la dimension décodage semble changer du CE2 au CM1. Les scores, compris entre 1.83 et 3.85, pour le texte 1, ne plafonnent pas en CM1. L'analyse post-hoc de Tukey confirme la croissance progressive du décodage avec des scores significativement différents pour chaque année ( $p < .001$ ). Les scores à la dimension décodage augmentent en moyenne de 0.31 points par an.

### 3.5 Vitesse

Le modèle de croissance montre un effet fixe positif du niveau sur le score à la dimension vitesse. L'ajout du niveau comme paramètre aléatoire de niveau 2 améliore notre modèle. On observe donc un effet fixe du niveau et une variance des pentes au niveau inter-élèves. Tous les élèves ne progressent donc pas au même rythme. L'ajout de l'interaction entre niveau et texte améliore le modèle. Le score à la dimension vitesse est significativement plus faible pour le texte 2 (-0.30 points) et le taux de croissance légèrement plus élevé. La différence reste cependant faible (+ 0.04 points par an). Comme on peut le constater sur la figure 3, la croissance des scores à la dimension vitesse semble ralentir du CE2 au CM1. Les scores, compris entre 1.7 et 4 plafonnent en CM1. L'analyse post-hoc HSD de Tukey confirme la croissance progressive de la vitesse avec des scores significativement différents pour chaque année ( $p < .001$ ). Les scores à la dimension vitesse augmentent en moyenne de 0.39 points par an.

## 4 Discussion

### 4.1 Modèle d'évaluation utilisé

Le modèle d'évaluation automatique proposé utilise des paramètres objectifs verbaux et prosodiques issus d'une comparaison avec de multiples références. Ce modèle permet de prédire les scores subjectifs à l'EMDF pour chacune des dimensions à partir de données acoustiques objectives. Les scores obtenus sont très corrélés à ceux donnés par des juges humains, tout en étant plus simple et rapide à mettre en œuvre. Ce modèle permet ainsi de s'abstraire de la variabilité inter-évaluateurs pouvant poser problème dans une analyse avec de multiples points de suivis. Cet outil est également moins dépendant du texte que l'évaluation humaine. Les scores entre deux textes différents sont plus consistants. De plus les juges humains sont soumis à plusieurs biais : les attendus selon l'âge supposés des lecteurs, l'interdépendance perceptive des dimensions de la fluence, la tendance à utiliser l'intégralité de l'échelle (y compris pour des élèves proches). Finalement, les modèles présentés s'appuyant sur des données objectives, donnent des scores continus. Ces scores sont donc beaucoup plus précis que ceux obtenus avec des juges humains, scores discrets. Cet outil permet donc de capter des variations plus fines de fluence chez les jeunes lecteurs.

Ainsi cet outil est un outil objectif, valide et précis tout à fait utilisable en recherche. Il est par ailleurs totalement adapté aux études longitudinales. Son objectivité permet de réaliser des évaluations multiples sans se soucier de la variabilité inter-évaluateurs. Les scores aux évaluations répétées sont directement comparables et les variations fines facilement détectables. Ainsi il pourrait être utilisé pour le suivi des progrès dans le cadre de la mise en place d'atelier spécifique ou, dans le cadre d'un entraînement, pour donner un feedback aux élèves et de visualiser leur progrès.

## 4.2 Développement de la prosodie

Les modèles de croissance des quatre dimensions de la fluence montrent des scores en augmentation du CE1 au CM1 dans toutes les dimensions. Cependant les profils sont différents d'une dimension à l'autre. En effet, les scores en CE1 sont plus élevés dans les dimensions d'automatisme (vitesse et décodage) que dans les dimensions prosodiques (expressivité et phrasé). Le rythme d'acquisition moyen dépend aussi de la dimension étudiée. Les scores de phrasé et de vitesse augmentent plus rapidement que ceux de décodage et d'expressivité. On peut supposer que les enfants de CE1 ayant déjà acquis un bon niveau de décodage, les scores plafonnent très rapidement. L'expressivité est une dimension qui s'acquiert plus tardivement et plus lentement.

Ces modèles nous montrent également des différences entre les enfants dans leurs rythmes d'acquisition des compétences de phrasé et de vitesse. On pourrait expliquer cette différence de rythme pour la vitesse par une forte hétérogénéité en CE1 dans cette compétence, avec des élèves plus faibles en CE1 qui « rattrapent » leurs camarades en CM1. Les compétences en expressivité sont moins hétérogènes en CE1 et progressent de façon similaire chez tous les enfants entre le CE1 et le CM1. Il serait intéressant de tester d'autres compétences pouvant expliquer ces différences (vocabulaire ou compréhension par exemple).

Finalement, on constate une différence entre les rythmes de progression sur les deux textes. Les scores progressent plus vite pour le dialogue, dans toutes les dimensions. Ces différences sont particulièrement fortes pour l'expressivité et le décodage. Il est probable que les enfants de CM1 soient plus à l'aise pour mettre en voix le dialogue et obtiennent donc des scores rapidement plus élevés que pour le texte narratif. Il est également possible que cet exercice soit peu familier aux enfants de CE1 et beaucoup plus en CM1. Il serait intéressant de prolonger cette étude longitudinale pour mieux cerner le profil d'acquisition spécifique à chaque dimension et particulièrement pour l'expressivité dont le développement se prolonge au-delà du CM1.

Ce type d'étude longitudinale donne des outils pour un meilleur diagnostic des difficultés rencontrées par les enfants. D'autre part, elle permet de mieux cibler les dimensions à entraîner à quel moment de l'apprentissage

## 5 Conclusion

Nous avons montré ici qu'en français, les dimensions prosodiques de la fluence se développent plus lentement et plus tardivement que les dimensions d'automatisme. Leur suivi est indispensable pour mieux comprendre l'acquisition des diverses compétences en lecture et repérer les élèves en difficulté. L'outil objectif proposé et utilisé ici permet des mesures répétées fiables, il est particulièrement adapté à ce type d'études longitudinales. L'intérêt pour l'enseignement de la lecture de cet outil réside dans sa simplicité et rapidité d'utilisation. Son utilisation simple en classe permet un suivi régulier et un feedback aux élèves. Son implémentation dans des applications d'entraînement de la fluence pourrait également permettre de faire le suivi de leur efficacité en classe.

## Références

1. Meisinger, E., Bradley, B. Schwanenflugel, P., Kuhn, M. Morris, R. : Myth and reality of the word caller : The relation between nominations and prevalence among elementary school children. *School Psychology Quarterly* 24(3), 147(2009).
2. Rasinski, T. *Assessing Reading Fluency*. Pacific Resources for Education and Learning (PREL), Honolulu (2004)
3. Kuhn, M., Schwanenflugel, P.: Aligning theory and assessment of reading fluency : Automaticity, prosody and definitions of fluency. *Reading Research Quarterly* 45(2), 230-251 (2010).
4. Godde, E., Bosse, M.L., Bailly, G., : A Review of Reading Prosody Acquisition and Development. *Reading and Writing*, 33, 399-426 (2020)
5. Wolters A., Kim YS, Szura J.: Is reading Prosody Related to Reading Comprehension ? A Meta-analysis. *Scientific Studies of Reading* (2020).
6. Álvarez-Cañizo, M., Suárez-Coalla, P. et Cuetos, F. : Reading prosody development in spanish children. *Reading and Writing* 31(1), 35–52 (2018).
7. Miller, J. et Schwanenflugel, P. J. : A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading research quarterly* 43(4), 336–354(2008).
8. Godde, E., Bosse, M.L., Bailly, G., Echelle Multi-Dimensionnelle de Fluence : nouvel outil d'évaluation de la fluence en lecture prenant en compte la prosodie, étalonné du CE1 à la 5ème. *L'Année Psychologique*. (Sous presse)
9. Moser, G. P., Sudweeks, R. R., Morrison, T. G. et Wilcox, B. : Reliability of ratings of children's expressive reading. *Reading Psychology* 35(1), 58–79(2014).
10. Bolanos, D., Cole, R. A., Ward, W. H., Tindal, G. A., Schwanenflugel, P. J. et Kuhn, M. R. : Automatic assessment of expressive oral reading. *Speech Communication* 55(2), 221–236(2013).
11. Hirst, D., Rilliard, A. et Aubergé, V. Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis. In *The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*.(1998)
12. Bailly G., Raidt S. et Elisei, F.: Gaze, conversational agents and face-to-face communication", *Speech Communication - special issue on Speech and Face-to-Face Communication*, 52(3): 598–612(2010).