



HAL
open science

De la recherche à la clinique, l'application AnaliDys : pour une meilleure caractérisation des erreurs orthographiques en écriture

Audrey Mazur, Matthieu Quignard, Agnes Witko

► To cite this version:

Audrey Mazur, Matthieu Quignard, Agnes Witko. De la recherche à la clinique, l'application AnaliDys : pour une meilleure caractérisation des erreurs orthographiques en écriture. *L'Orthophoniste*, 2025, 446, pp.28-31. <hal-05086982>

HAL Id: hal-05086982

<https://hal.science/hal-05086982v1>

Submitted on 5 Jun 2025

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

De la recherche à la clinique, l'application AnalIDys

Pour une meilleure caractérisation des erreurs orthographiques en écriture

Audrey Mazur, Matthieu Quignard,
et Agnès Witko

De la recherche à l'impact sociétal

Dans l'article précédent, les travaux cités montrent que les étudiants présentant une dyslexie ont encore de grandes difficultés en orthographe. Deux études supplémentaires sur l'orthographe ont permis d'apporter des nuances sur ces difficultés orthographiques (Mazur *et al.*, 2021 et Mazur *et al.*, 2022). Premièrement, les analyses qualitatives révèlent que les mots à risque des étudiants dyslexiques ont le même profil que ceux des contrôles : si les étudiants présentant une dyslexie déclarent être en insécurité linguistique (Mazur et Chenu, 2023) les analyses révèlent que, objectivement parlant, les étudiants dyslexiques ne s'autocensurent pas concernant leurs choix lexicaux (Mazur *et al.*, 2021). Dans l'échantillon de mots analysés, les mots à risque sont en grande partie courts et homophones. Cela étant dit, les étudiants dyslexiques font davantage d'erreurs sur ces types de mots, leur liste de mots à risque est plus importante que celle des étudiants contrôles, et ces erreurs sont assez représentatives de la totalité du groupe des étudiants dyslexiques (ce qui n'est pas le cas pour les mots à risque des

étudiants contrôles, qui concernent deux à quatre étudiants sur les 22). Deuxièmement, les analyses quantitatives confirment cette tendance : les mots à risque sont à la fois courts et homophones. En effet, une interdépendance entre la caractéristique « longueur courte » et « homophonie » est révélée, et ce d'autant plus chez les étudiants dyslexiques. Les mots à risque des étudiants dyslexiques sont donc, dans une certaine mesure, prédictibles. Ces informations sur les mots à risque peuvent être particulièrement pertinentes pour la remédiation en orthographe.

Les deux analyses mentionnées précédemment ont généré la mise en place d'une chaîne de traitement linguistique, nommée ALEX (responsable, M. Quignard ; cf. Mazur-Palandre *et al.*, 2021 pour plus d'explications). Pour chaque mot utilisé, et en nous basant sur les bases Lexiques (Lexique.org¹) et Manulex (Lété *et al.*, 2004), la chaîne de traitement indique :

- des propriétés graphémiques et phonétiques : le nombre de lettres, de graphèmes, de phonèmes, de syllabes, l'indicateur de consistance phonème / graphème, l'indicateur de consistance

graphème / phonème et la structure phonologique (succession consonnes / voyelles) ;

- des propriétés syntaxiques : la classe grammaticale, à savoir abréviation, adjectif, adverbe, conjonction, déterminant, nom commun, préposition, pronom ou verbe ;
- des propriétés de fréquence : la fréquence, selon la base lexicale Lexique.org (fréquence de lemmes, basée sur les sous-titres de films français) ;
- des propriétés de longueur : mots courts de 1 à 4 caractères ; mots moyens de 5 à 7 caractères ; mots longs à partir de 8 caractères.

Il est apparu que cette chaîne de traitement, développée à des fins de recherche en laboratoire, offre l'opportunité de développer une recherche translationnelle. Le transfert des résultats et l'adaptation de la chaîne de traitement lexical en un outil pouvant spécifier les erreurs orthographiques pourrait être apprécié. C'est dans ce contexte que l'application AnalIDys a été pensée et développée.

Présentation de l'application

AnaliDys est une application ayant pour objectif de caractériser certaines propriétés lexicales des mots utilisés dans les productions textuelles écrites. Basée sur la chaîne de traitement ALEX (responsable, M. Quignard), elle permet de donner pour chaque mot utilisé, erroné ou pas, par votre patient :

- le nombre de lettres, de graphèmes, de phonèmes, de syllabes, l'indicateur de consistance phonème / graphème, l'indicateur de consistance graphème / phonème et la structure phonologique (succession consonnes / voyelles) ;
- la classe grammaticale, à savoir abréviation, adjectif, adverbe, conjonction, déterminant, nom commun, préposition, pronom ou verbe ;
- la fréquence ;
- la longueur.

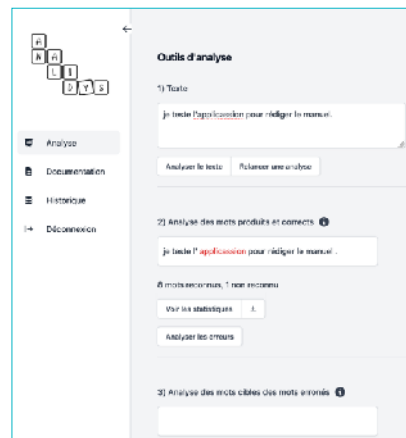
Il s'agit d'une application visant à aider l'orthophoniste à caractériser les propriétés lexicales des textes écrits de patients (dont les erreurs) présentant des troubles du langage écrit tels que la dyslexie et la dysorthographe. Ainsi, pour un texte de patient donné, pour chaque mot utilisé, l'application donne ces indications lexicales.

L'application peut également donner une cartographie des caractéristiques des mots utilisés par le patient et des erreurs commises. Elle effectue une analyse lexicale approfondie du texte en proposant des descriptions linguistiques de chaque mot sur la base d'un calcul statistique, et en pointant les erreurs par le repérage et l'identification du mot cible.

Une fois les erreurs repérées, le système établit un rapport statistique sur :

- les mots corrects utilisés ;
- les erreurs en fonction des propriétés lexicales des mots-cibles pour chacun des indicateurs lexicaux proposés par l'application (nombres d'unités, classe grammaticale, etc.).

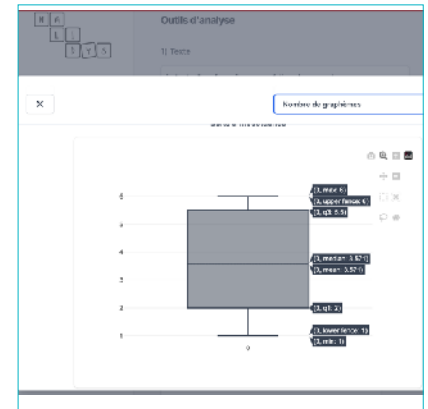
Opérationnalité



L'application permet de rentrer un texte sans limite de longueur et sans exiger d'informations à caractère personnel, permettant de sauvegarder l'anonymat du patient. Une fois le texte rentré par l'orthophoniste, deux types d'analyses sont possibles : celle sur tous les mots utilisés par le patient et celle seulement sur les mots erronés. On peut alors obtenir les statistiques des mots corrects qui permettent d'avoir différentes informations lexicales sur les mots utilisés : nombre de lettres, de syllabes à l'écrit, de phonèmes, de graphèmes, fréquence des formes dans les manuels scolaires, fréquence des formes dans la littérature, fréquence des formes dans les

sous-titres des films, correspondance orthographique phonème/graphème (CPG) et correspondance orthographique graphème/phonème (CGP). Les mêmes analyses sont possibles seulement sur les mots erronés.

Les figures ci-après donnent un output pour l'indicateur « nombre de graphèmes ».



statistique	Score
Nb de Mots (n)	7
Moyenne (μ)	3,571
Ecart-type (σ)	2,07
Minimum	1
Premier quartile	2
Médiane	3
Troisième quartile	5,5
Maximum	6

Figure 1. Statistiques de l'indicateur « nombre de graphèmes »

Les fichiers .csv de chacune des deux analyses statistiques sont téléchargeables et manipulables dans Excel. Les graphiques sont également téléchargeables en format image.

Pour plus de renseignements sur l'opérationnalité de l'application, un manuel d'utilisation complet est disponible.

Les auteurs remercient le LABEX ASLAN (ANR-10-LABX-0081) de l'université de Lyon pour son soutien financier dans le cadre du programme français « Investissements d'Avenir » géré par l'Agence nationale de la recherche (ANR).

Une plus-value pour l'orthophonie ?

De ce projet, nous entrevoyons trois principales caractéristiques pouvant apporter de la plus-value pour le domaine de l'orthophonie.

Premièrement, cette application peut enrichir la boîte à outils des orthophonistes :

- en utilisant des outils de référence comme Lexique.org et Manulex, qui sont des outils scientifiques rigoureux pour décrire les productions d'une patientèle ;
- en documentant des critères linguistiques rentables : par exemple la fréquence de la forme des mots plutôt au singulier ou au pluriel pour les noms, plutôt à la 3^e personne du singulier au présent pour les verbes, ou les mots invariables, et donc des mots indispensables à connaître ;
- en repérant des critères linguistiques logiques, comme la morphologie des mots ;
- en proposant des collections de mots structurées au patient et à ses aidants.

Deuxièmement, cette application peut permettre de mettre en place un projet thérapeutique sur la base d'outils de traçabilité :

- en obtenant une cartographie des mots utilisés par un patient ;
- en capitalisant des types de mots pour tracer le lexique utilisé ;
- en orientant finement la remédiation à partir de la cartographie des mots erronés ;
- en suivant les productions d'un patient sur une période pour voir l'évolution du répertoire lexical qu'il utilise à l'écrit (évolution du répertoire linguistique).



Troisièmement, tester et utiliser l'application AnaliDys permet une collaboration active **avec des instances de recherche, de clinique et d'éducation**:

- en tissant des relations avec les chercheurs engagés dans la pluridisciplinarité ;
- en faisant des liens avec d'autres bases de données répertoriées dans la clinique orthophonique (par exemple, <http://pontt.net/2006/08/lesbasesdedonneeslexicales/>);
- en établissant des liens avec des bases de données répertoriées dans l'éducation pour les enfants et les adolescents, entre autres <https://www.appligogiques.com/eqol> ;



<https://www.reseau-canope.fr/agence-des-usages/les-pratiques-dutilisation-des-lexiques-numeriques-chez-les-apprenantes-pistes-denseignement.html> ;

<https://eduscol.education.fr/186/liste-de-frequence-lexicale>.



Appel à participation



L'application AnaliDys étant à sa première version, nous sommes toujours à la recherche d'orthophonistes pour la tester et nous faire un retour afin que cette application soit totalement opérationnelle en cabinet.

Il est encore possible de participer à cette étude, conduite par Audrey Mazur, Matthieu Quignard (docteur en sciences du langage et sciences cognitives et ingénieur de recherche au sein du LabEx Aslan et du laboratoire Icar, CNRS UMR 5191, université Lyon 2 et ENS de Lyon) et Agnès Witko (docteure en sciences du langage, orthophoniste et MCF université Lyon 1, laboratoire DDL, CNRS UMR 5596 et université Lyon 2).

Cet outil a pour objectif de critérier les mots utilisés par vos patients présentant des troubles du langage écrit et les erreurs qu'ils peuvent faire, afin d'en avoir une visualisation plus précise. Votre but est alors de tester cet outil sur des données anonymisées d'un ou plusieurs patients et de nous faire un retour sur votre usage, en remplissant un court questionnaire.

Ainsi, si vous êtes intéressé-es, nous vous laissons nous contacter par mail : audrey.mazur@ens-lyon.fr