



HAL
open science

Performances des correcteurs orthographiques pour la correction des erreurs dys

Johana Bodard, Céline Jost, Isis Truck, Gérard Uzan

► **To cite this version:**

Johana Bodard, Céline Jost, Isis Truck, Gérard Uzan. Performances des correcteurs orthographiques pour la correction des erreurs dys. Handicap 2024, IFRATH, Jun 2024, Paris, France. hal-04611887

HAL Id: hal-04611887

<https://hal.science/hal-04611887v1>

Submitted on 14 Jun 2024

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.

Performances des correcteurs orthographiques pour la correction des erreurs dys

Johana Bodard
Laboratoire CHArt
Université Paris 8
Saint-Denis, France
0000-0002-6556-2640

Céline Jost
Laboratoire CHArt
Université Paris 8
Saint-Denis, France
0000-0002-5870-9640

Isis Truck
Laboratoire CHArt
Université Paris 8
Saint-Denis, France
0000-0003-2052-2546

Gérard Uzan
Laboratoire CHArt
Université Paris 8
Saint-Denis, France
0009-0009-1396-6983

Résumé—Cet article présente les principaux résultats d’une évaluation des performances de plusieurs correcteurs orthographiques sur un corpus de textes rédigés par des personnes dyslexiques-dysorthographiques. Pour identifier les points forts et les points faibles de ces outils, nous avons évalué leurs performances en détection et en correction en fonction de différentes caractéristiques des mots erronés telles que le type des erreurs. Les résultats indiquent que les performances de ces outils sont insuffisantes pour aider les personnes dyslexiques-dysorthographiques à compenser leurs difficultés en orthographe.

Mots clés—correction orthographique, erreur d’orthographe, dyslexie, dysorthographie

I. INTRODUCTION

En France, la maîtrise de l’orthographe est une compétence socialement valorisée et le non-respect des normes orthographiques ou grammaticales est généralement mal perçu, notamment dans le milieu professionnel [1], [2]. Cela peut être particulièrement handicapant pour les personnes dyslexiques-dysorthographiques (dys) en raison de leurs difficultés importantes en orthographe.

Plusieurs études ont comparé les erreurs d’orthographe des personnes dys avec celles de personnes non dys [3]–[6]. Ces études retrouvent globalement un profil d’erreurs similaire entre les deux groupes : ils produisent les mêmes types d’erreurs dans les mêmes proportions. La principale différence entre les écrits des deux groupes réside dans la quantité d’erreurs produites : les personnes dys font significativement plus d’erreurs que les personnes non dys. Certaines études relèvent également une inconsistance des erreurs [3], [6]. De plus, l’étude de Mazur-Palandre réalisée auprès d’étudiants à l’université montre que même chez cette population ayant atteint un niveau d’études élevé, des difficultés importantes persistent et notamment des erreurs inhabituelles pour leur niveau d’études.

Une question qui se pose est de savoir quelles sont les performances des logiciels de correction orthographique avec des textes écrits par des personnes dys. Par exemple, la correction de certaines erreurs nécessite de prendre en compte le contexte (les mots voisins). Cependant, plus une personne produit d’erreurs, plus le contexte autour de ces erreurs a de chances d’être erroné. Cela peut avoir un impact sur les performances des correcteurs.

Une étude d’Antoine *et al.* suggère que les correcteurs ne répondent pas complètement aux besoins en correction d’enfants dyslexiques [7]. Il peut donc être intéressant d’étudier pourquoi : quels sont les points forts (les problèmes déjà résolus) et les points faibles (les problèmes qui restent à résoudre) des correcteurs à disposition des personnes dyslexiques.

Dans cet article, nous présentons les principaux résultats d’une évaluation des performances de neuf correcteurs orthographiques sur un corpus d’écrits dys. La section II explique comment évaluer les performances des systèmes de correction orthographique. La section III décrit la méthodologie que nous avons utilisée pour mener notre évaluation. Les résultats de cette évaluation sont présentés dans la section IV. La section V discute des résultats, tandis que la section VI conclut ce papier.

II. ÉVALUATION DE CORRECTEURS ORTHOGRAPHIQUES

L’évaluation des performances des systèmes de correction orthographique nécessite la collecte de données d’évaluation et le choix de métriques appropriées.

A. Données

Pour évaluer les performances des systèmes de correction orthographique, il est nécessaire de rassembler trois ensembles de données :

- les **données de test** : textes contenant des erreurs fournis en entrée des correcteurs orthographiques pour évaluer leurs performances ;
- les **données de référence** : version sans erreur des données de test ;
- les **données des systèmes** : résultats de la correction des données de test par les correcteurs orthographiques.

B. Métriques

Les métriques les plus fréquemment utilisées pour évaluer les performances des correcteurs orthographiques sont le **rap- pel**, la **précision** et le **F-score**. Ces métriques sont calculées à partir des taux de **TP** (*True Positive*), **FP** (*False Positive*), **TN** (*True Negative*), et **FN** (*False Negative*). Dans un premier temps, il est nécessaire de définir à quoi correspondent ces quatre classes pour la détection et pour la correction.

En détection, le travail du correcteur peut être assimilé à un problème de classification : le correcteur doit prédire si un mot appartient à la classe des mots corrects ou à celle des mots erronés. L'évaluation consiste à vérifier la fiabilité de ces prédictions. Il existe quatre cas possibles :

- le mot est erroné et le système prédit qu'il est erroné \Rightarrow TP ;
- le mot est correct et le système prédit qu'il est correct \Rightarrow TN ;
- le mot est correct, mais le système prédit qu'il est erroné \Rightarrow FP ;
- le mot est erroné, mais le système prédit qu'il est correct \Rightarrow FN.

Autrement dit, T signifie que la prédiction du correcteur est correcte (*True*), F qu'elle est fautive (*False*). P signifie que le correcteur a prédit que le mot appartenait à la classe positive (ici, les mots erronés) et N signifie que le correcteur a prédit que le mot appartenait à la classe négative (les mots corrects).

En correction, on cherche à déterminer, d'une part, si la correction prédite par le système pour un mot erroné est identique à celle proposée dans la référence, et d'autre part, si le système n'a pas modifié un mot qui est correct. Chodorow *et al.* distinguent les cinq cas suivants [8] :

- le mot est erroné et le système prédit la bonne correction \Rightarrow TP ;
- le mot est correct et le système ne l'a pas modifié \Rightarrow TN ;
- le mot est correct, mais le système l'a modifié \Rightarrow FP ;
- le mot est erroné, mais le système ne l'a pas modifié \Rightarrow FN ;
- le mot est erroné, mais le système n'a pas prédit la bonne correction \Rightarrow il s'agit à la fois d'un FN (le mot n'est pas corrigé) et d'un FP (la correction prédite n'est pas valide), c'est-à-dire d'un FPN.

Une fois que l'on a déterminé les taux de TP, TN, FP et FN en détection et en correction, on peut évaluer les performances des systèmes avec les métriques précédemment citées :

1) *Le rappel (r)* : il évalue la capacité d'un système à détecter ou corriger l'ensemble des mots erronés.

$$r = \frac{TP}{TP + FN} \quad (1)$$

2) *La précision (p)* : elle évalue l'exactitude des détections ou des corrections réalisées par un système.

$$p = \frac{TP}{TP + FP} \quad (2)$$

3) *Le F-score (F_β)* : il combine la précision et le rappel dans une mesure unique.

$$F_\beta = (1 + \beta^2) \times \frac{p \times r}{\beta^2 \times p + r} \quad (3)$$

Avec $\beta = 1$, on accorde un poids équivalent à la précision et au rappel. Avec $\beta = 0,5$, on accorde deux fois plus de poids à la précision qu'au rappel.

A. Correcteurs orthographiques

Nous avons sélectionné les neuf outils suivants :

- **Cordial** (CO) et **Antidote** (AN) : deux logiciels de correction orthographique et grammaticale ;
- **le filtre de correction d'ADELE-TEAM** (ADC) : ADELE-TEAM est un outil d'aide à la lecture et à l'écriture qui inclut un filtre de correction ;
- **Microsoft Word** (MW) et **Pages** (PA) : deux logiciels de traitement de texte qui intègrent un correcteur orthographique et grammatical ;
- **Grammalecte** (GR) : un outil gratuit et libre pour la correction des erreurs orthographiques, grammaticales et typographiques ;
- **Scribens** (SC), **LanguageTool** (LT), **BonPatron** (BP) : trois outils en ligne.

B. Données

Les données de test sont composées des 78 textes du Corpus DYS [9]. Les mots erronés de ce corpus ont été extraits et annotés manuellement. Le résultat de cette annotation constitue les données de référence. Pour obtenir les données des systèmes, nous avons fait corriger chaque texte du corpus par les correcteurs orthographiques et nous avons extrait manuellement les informations suivantes : la forme signalée par le correcteur (mot ou groupe de mots), les suggestions de correction (jusqu'à cinq suggestions), les éventuelles explications fournies par le correcteur, la position de la forme signalée dans le texte.

C. Métriques

Pour évaluer les performances globales des correcteurs (c'est-à-dire sur l'ensemble des mots du corpus), nous avons utilisé les métriques présentées dans l'état de l'art : le rappel, la précision, et le F-score (avec $\beta = 1$). Les définitions des TP, TN, FP et FN correspondent également à celles présentées dans l'état de l'art. En correction, on considère qu'un mot est corrigé, si et seulement si la première suggestion est correcte.

Puis, nous nous sommes intéressés aux performances des correcteurs en fonction des caractéristiques des mots erronés. Nous présentons dans ce papier les résultats obtenus pour les caractéristiques suivantes :

- le type des erreurs contenues dans les mots erronés ;
- le fait que le mot erroné soit une erreur lexicale (non-mot, par exemple, *aprle / parle*) ou une erreur non lexicale (erreur qui produit un mot que l'on peut trouver dans un dictionnaire, par exemple, *peut / peu*) ;
- le fait que le mot erroné soit phonologiquement plausible (par exemple, *bato / bateau*) ou non phonologiquement plausible (par exemple, *réusite / réussite*).

Pour ces trois caractéristiques, nous avons regardé la proportion de mots erronés détectés et la proportion de mots erronés bien corrigés parmi ceux qui sont détectés par les correcteurs (ce qui équivaut au rappel en détection et en correction respectivement).

De plus, les correcteurs signalent parfois des formes composées de plusieurs mots (par exemple, *les gens touché*). Dans ce cas, il n'est pas possible de déterminer de manière systématique quels mots sont considérés comme corrects et quels mots sont considérés comme erronés par le correcteur. Nous avons donc séparé l'analyse des formes composées d'un seul mot de celle des groupes de mots.

IV. RÉSULTATS

Remarque : une flèche vers le bas (\downarrow) dans l'entête d'une colonne indique que cette colonne a été utilisée pour trier les données par ordre décroissant.

A. Performances générales en détection

Le tableau I présente les performances en détection de chaque correcteur sur l'ensemble des mots du corpus (hors groupes de mots). Le rappel varie fortement selon les correcteurs, de seulement 38,5 % pour Pages à 75,9 % pour Cordial (moyenne = 58 %, écart-type = 9,4 %). Ainsi, même dans le meilleur des cas, un quart des mots erronés ne sont pas détectés. En revanche, tous les correcteurs affichent une bonne précision (moyenne = 96 %, écart-type = 3 %), ce qui signifie que, dans la très grande majorité des cas, les mots signalés par les correcteurs sont bien erronés.

TABLE I
DÉTECTION – PERFORMANCES DE CHAQUE CORRECTEUR SUR L'ENSEMBLE DU CORPUS (HORS GROUPES DE MOTS)

	TP	FP	TN	FN	R	P	F ₁ ↓
CO	1515	101	4449	482	0,759	0,938	0,839
AN	1465	39	4846	862	0,630	0,974	0,765
GR	1458	86	4817	889	0,621	0,944	0,749
LT	1318	52	4688	854	0,607	0,962	0,744
BP	933	24	4058	672	0,581	0,975	0,728
MW	1307	11	4837	984	0,570	0,992	0,724
SC	1254	44	4812	1039	0,547	0,966	0,698
ADC	1229	151	4722	1120	0,523	0,891	0,659
PA	898	3	4870	1433	0,385	0,997	0,556

On retrouve également cette bonne précision avec les groupes de mots, comme le montre le tableau II. En moyenne, 93,6 % des groupes de mots signalés contiennent au moins une erreur (écart-type = 5,2 %). Les précisions obtenues pour les groupes de mots sont proches de celles obtenues pour les formes composées d'un seul mot, à l'exception de Grammalecte qui affiche une précision inférieure de 0,11 point pour les groupes de mots.

B. Performances générales en correction

Le tableau III présente les performances en correction des correcteurs pour l'ensemble des mots du corpus (hors groupes de mots). Globalement, ces performances sont très faibles. En moyenne, les correcteurs corrigent 30,9 % des mots erronés (écart-type = 9,6 %) et seul Cordial parvient à corriger un peu plus de la moitié des mots erronés (51,3 %). Bien qu'ils restent faibles, les résultats en précision sont meilleurs : en moyenne, 53,7 % des corrections proposées sont valides (écart-type = 9,2 %).

TABLE II
DÉTECTION – PRÉCISION POUR LES GROUPES DE MOTS ERRONÉS DÉTECTÉS PAR LES CORRECTEURS

	TP	FP	Total	P ↓
PA	52	0	52	1,00
MW	78	1	79	0,987
BP	639	24	663	0,964
AN	44	2	46	0,957
LT	185	9	194	0,954
SC	69	4	73	0,945
CO	268	21	289	0,927
ADC	37	6	43	0,860
GR	25	5	30	0,833

TABLE III
CORRECTION – PERFORMANCES DE CHAQUE CORRECTEUR SUR L'ENSEMBLE DU CORPUS (HORS GROUPES DE MOTS)

	TP	FP	TN	FN	FPN	R	P	F ₁ ↓
CO	1024	524	4454	973	428	0,513	0,661	0,578
AN	935	453	4849	1392	417	0,402	0,674	0,503
MW	799	517	4837	1492	506	0,349	0,607	0,443
SC	678	558	4822	1615	524	0,296	0,549	0,384
GR	696	816	4830	1651	743	0,297	0,460	0,361
LT	614	647	4704	1558	611	0,283	0,487	0,358
ADC	610	600	4771	1739	498	0,260	0,504	0,343
PA	451	414	4873	1880	414	0,193	0,521	0,282
BP	302	507	4064	1303	489	0,188	0,373	0,250

Ces faibles résultats en correction s'expliquent notamment par les faibles résultats en détection : les mots erronés non détectés ne peuvent pas être corrigés. Nous avons donc également calculé les performances en correction des correcteurs en prenant en compte uniquement les mots erronés qu'ils ont détectés. Les résultats sont présentés dans le tableau IV. En moyenne, les correcteurs corrigent 52,6 % des mots erronés qu'ils détectent (écart-type = 10 %). La précision est légèrement meilleure : en moyenne 55,7 % des corrections proposées sont valides (écart-type = 9,8 %).

TABLE IV
CORRECTION – PERFORMANCES DE CHAQUE CORRECTEUR POUR LES MOTS ERRONÉS DÉTECTÉS (HORS GROUPES DE MOTS)

	TP	FP	TN	FN	FPN	R	P	F ₁ ↓
CO	1024	428	0	491	428	0,676	0,705	0,690
AN	935	417	0	530	417	0,638	0,692	0,664
MW	799	506	0	508	506	0,611	0,612	0,612
SC	678	524	0	576	524	0,541	0,564	0,552
ADC	610	498	0	619	498	0,496	0,551	0,522
PA	451	414	0	447	414	0,502	0,521	0,512
LT	614	611	0	704	611	0,466	0,501	0,483
GR	696	743	0	762	743	0,477	0,484	0,480
BP	302	489	0	631	489	0,324	0,382	0,350

Le tableau V présente les performances en correction des groupes de mots pour chaque correcteur. Celles-ci sont parfois très différentes de celles obtenues pour les mots erronés détectés par les correcteurs. Par exemple, Pages corrige à peine plus de la moitié des mots erronés qu'il détecte, alors qu'il corrige deux tiers des groupes de mots qu'il signale. À l'inverse, Cordial corrige deux tiers des mots erronés qu'il détecte et un peu moins de la moitié des groupes de mots qu'il signale. BonPatron ne corrige quasiment aucun des groupes de

mots qu'il signale, parce qu'il ne propose quasiment jamais de correction. Il faut noter également que le nombre de groupes de mots signalés varie fortement d'un correcteur à l'autre : 663 pour BonPatron et seulement 30 pour Grammalecte.

TABLE V
CORRECTION – PERFORMANCE EN CORRECTION DES GROUPES DE MOTS

	TP	FP	TN	FN	FPN	R	P	F ₁ ↓
PA	35	16	0	17	16	0,673	0,686	0,680
MW	51	28	0	27	27	0,654	0,646	0,650
AN	28	18	0	16	16	0,636	0,609	0,622
GR	13	16	1	12	12	0,520	0,448	0,481
CO	132	152	0	136	131	0,493	0,465	0,478
SC	27	27	2	42	25	0,391	0,500	0,439
LT	51	76	1	134	68	0,276	0,402	0,327
ADC	2	31	1	35	26	0,054	0,061	0,057
BP	4	7	24	635	7	0,006	0,364	0,012

C. Types d'erreurs

1) *Performances en détection* : Le tableau VI présente les principales statistiques du rappel en détection (proportion de mots détectés parmi les mots erronés – hors groupes de mots) en fonction du type des erreurs contenues dans les mots erronés pour l'ensemble des correcteurs évalués. Pour certains types d'erreurs, les résultats varient peu en fonction des correcteurs. Les erreurs les moins bien détectées par les correcteurs sont celles qui portent sur les mots dans leur ensemble (confusions entre homophones, substitution de mots, ajout ou omission de mots). En revanche, les mots contenant des erreurs de transposition de caractères ou d'omission de caractères et ceux contenant des erreurs de phonétisation ou d'approximation sont bien détectés dans l'ensemble. La proportion de mots erronés détectés varie fortement pour les erreurs de majuscules, les confusions entre homophones, et les erreurs d'accord en genre et en nombre et de conjugaison.

Les résultats pour chaque correcteur sont présentés dans le tableau VII. Tous les correcteurs détectent les mots contenant des erreurs de déplacement de caractères. Cependant, le corpus contient uniquement deux mots contenant ce type d'erreurs. Dans ces conditions, il est difficile de conclure quoi que ce soit pour ce type d'erreurs. C'est également le cas pour les types d'erreurs « mot non reconnu » et « liaison erronée », car il y a moins de dix occurrences de mots erronés contenant ce type d'erreurs dans le corpus.

2) *Performances en correction* : Le tableau VIII présente les principales statistiques du rappel en correction des mots détectés par les correcteurs en fonction du type des erreurs contenues dans les mots erronés pour l'ensemble des correcteurs évalués. En moyenne, les erreurs les mieux corrigées sont les erreurs de majuscules (moyenne = 60,8 %). Pour tous les autres types d'erreurs, en moyenne, moins de la moitié des mots erronés détectés sont bien corrigés par les correcteurs.

Comme pour la détection et pour les raisons évoquées précédemment, il est difficile de tirer des conclusions pour les erreurs de déplacement de caractères et les erreurs de liaison. De plus, aucun correcteur n'ayant réussi à détecter plus de dix mots erronés contenant des erreurs d'omission, d'ajout ou

TABLE VI
DÉTECTION – PERFORMANCES EN FONCTION DU TYPES D'ERREUR POUR L'ENSEMBLE DES CORRECTEURS

Type	Min	Max	Moyenne ↓	Médiane	Écart-type
DPL*	1,000	1,000	1,000	1,000	0,000
TRS	0,750	0,966	0,912	0,931	0,063
OMS	0,667	0,913	0,836	0,863	0,066
APX	0,711	0,897	0,828	0,845	0,053
PHN	0,691	0,879	0,817	0,845	0,059
AJT	0,667	0,871	0,765	0,765	0,054
SEG	0,616	0,919	0,746	0,740	0,103
MNR*	0,600	0,800	0,733	0,800	0,094
SUB	0,488	0,744	0,638	0,643	0,067
MAJ	0,248	0,876	0,547	0,544	0,220
LSN*	0,400	0,800	0,541	0,400	0,168
FLX	0,285	0,734	0,538	0,537	0,139
HOM	0,073	0,621	0,327	0,340	0,166
LEX	0,000	0,188	0,068	0,056	0,061
MOT	0,000	0,212	0,049	0,028	0,064

* moins de 10 formes erronées contiennent ce type d'erreurs dans le corpus. DPL : déplacement d'un caractère de plus d'une position ; TRS : transposition de deux caractères ; OMS : omission d'un caractère ; APX : approximation ; PHN : phonétisation ; AJT : ajout d'un caractère ; SEG : erreur de segmentation ; MNR : mot non reconnu ; SUB : substitution de caractères ; MAJ : erreur de majuscule ; LSN : liaison erronée ; FLX : erreur d'accord ou de conjugaison ; HOM : confusion entre homophones ; LEX : substitution de mots ; MOT : ajout, omission ou déplacement d'un mot.

TABLE VII
DÉTECTION – RAPPEL EN FONCTION DES TYPES D'ERREURS POUR CHAQUE CORRECTEUR

Type	CO	AN	ADC	MW	GR	PA	SC	LT	BP
FLX	0,73	0,71	0,59	0,40	0,64	0,28	0,49	0,54	0,45
PHN	0,86	0,85	0,84	0,81	0,85	0,74	0,69	0,85	0,88
APX	0,90	0,84	0,84	0,81	0,87	0,77	0,71	0,86	0,85
HOM	0,62	0,28	0,11	0,45	0,45	0,07	0,34	0,39	0,21
MAJ	0,88	0,54	0,29	0,79	0,39	0,25	0,74	0,68	0,37
SEG	0,92	0,81	0,62	0,77	0,74	0,62	0,63	0,74	0,87
OMS	0,86	0,82	0,87	0,85	0,87	0,81	0,67	0,86	0,91
SUB	0,74	0,64	0,63	0,60	0,71	0,65	0,49	0,65	0,62
MOT	0,21	0,00	0,03	0,00	0,03	0,00	0,09	0,03	0,05
AJT	0,87	0,71	0,76	0,76	0,77	0,76	0,67	0,81	0,76
TRS	0,89	0,89	0,93	0,97	0,93	0,93	0,75	0,96	0,96
LEX	0,12	0,00	0,06	0,00	0,11	0,00	0,19	0,06	0,08
LSN	0,80	0,40	0,40	0,40	0,80	0,40	0,60	0,40	0,67
MNR	0,80	0,60	0,80	0,80	0,80	0,80	0,60	0,60	0,80
DPL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

de substitution de mots, il est également difficile de tirer des conclusions pour ces types d'erreurs.

Les résultats pour chaque correcteur, présentés dans le tableau IX, montrent cependant une forte variabilité pour certains types d'erreurs. Par exemple, Cordial corrige 90 % des erreurs de majuscule qu'il détecte, tandis qu'ADELE-TEAM en corrige seulement 4,1 %. De même, Cordial corrige 82,9 % des confusions entre homophones qu'il détecte, tandis qu'ADELE-TEAM n'en corrige aucune. Les autres types concernés par ces fortes variations sont les erreurs d'accord en genre et en nombre et de conjugaison, et dans une moindre mesure les erreurs de transposition et d'ajout de caractères. En revanche, les résultats varient peu pour les erreurs d'approximation, qui obtiennent le deuxième meilleur rappel en moyenne.

D. Erreurs lexicales (EL) et non lexicales (ENL)

1) *Performances en détection* : Le tableau X présente le rappel en détection des erreurs lexicales (EL, erreurs qui

TABLE VIII
CORRECTION – PERFORMANCE EN FONCTION DU TYPE D’ERREUR POUR L’ENSEMBLE DES CORRECTEURS

Type	Min	Max	Moyenne ↓	Médiane	Écart-type
MAJ	0,041	0,900	0,608	0,768	0,316
APX	0,443	0,571	0,499	0,499	0,041
HOM	0,000	0,829	0,481	0,624	0,304
DPL*	0,000	1,000	0,444	0,500	0,369
AJT	0,115	0,481	0,389	0,409	0,105
PHN	0,256	0,522	0,384	0,399	0,083
FLX	0,064	0,668	0,380	0,433	0,183
TRS	0,000	0,519	0,294	0,320	0,150
SUB	0,074	0,429	0,267	0,261	0,095
OMS	0,085	0,362	0,262	0,278	0,082
SEG	0,110	0,405	0,247	0,255	0,083
MOT ²	0,000	1,000	0,190	0,000	0,363
LEX ²	0,000	0,500	0,056	0,000	0,157
LSN*	0,000	0,500	0,056	0,000	0,157

* moins de 10 formes erronées contiennent ce type d’erreurs dans le corpus.

² tous les correcteurs ont détecté moins de 10 mots erronés contenant ce type d’erreur.

TABLE IX
CORRECTION – RAPPEL EN FONCTION DES TYPES D’ERREURS POUR CHAQUE CORRECTEUR

Type	CO	AN	GR	ADC	LT	SC	BP	MW	PA
FLX	0,56	0,67	0,43	0,51	0,24	0,45	0,06	0,30	0,19
PHN	0,48	0,44	0,26	0,52	0,33	0,31	0,31	0,40	0,41
APX	0,57	0,52	0,44	0,49	0,50	0,45	0,46	0,54	0,53
HOM	0,83	0,65	0,62	0,00	0,54	0,65	0,13	0,82	0,08
MAJ	0,90	0,79	0,72	0,04	0,77	0,79	0,05	0,88	0,54
SEG	0,25	0,41	0,18	0,24	0,16	0,11	0,27	0,31	0,29
OMS	0,34	0,22	0,21	0,08	0,22	0,28	0,30	0,36	0,33
SUB	0,28	0,26	0,21	0,07	0,29	0,38	0,26	0,23	0,43
MOT	0,71	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00
AJT	0,48	0,36	0,44	0,12	0,46	0,41	0,36	0,46	0,40
TRS	0,40	0,40	0,15	0,00	0,32	0,29	0,18	0,39	0,52
LEX	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
LSN	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,50
DPL	1,00	0,00	0,50	0,00	0,50	0,00	0,50	1,00	0,50

produisent des non-mots) et des erreurs non lexicales (ENL, erreurs qui aboutissent à des mots du dictionnaire), ainsi que la différence entre les deux rappels. Dans l’ensemble, les erreurs lexicales sont très bien détectées par tous les correcteurs (moyenne = 96 %, écart-type = 4,8 %). On note cependant qu’à l’exception de Language Tool, aucun correcteur ne détecte 100 % des erreurs lexicales. En revanche, les performances sont très variables pour les erreurs non lexicales. Si Cordial arrive à détecter un peu moins des deux tiers de ces erreurs, Pages, à l’inverse, n’en détecte quasiment aucune. En moyenne, seulement 31,5 % de ces erreurs sont détectées (écart-type = 16 %).

2) *Performances en correction* : Le tableau XI présente le rappel en correction des erreurs lexicales (EL) et des erreurs non lexicales (ENL) détectées par les correcteurs, ainsi que la différence entre les deux rappels. En moyenne, seulement 45,2 % des erreurs non lexicales détectées sont corrigées (écart-type = 7,2 %). Seuls trois correcteurs, Cordial, Antidote et Pages corrigent plus de la moitié des erreurs non lexicales qu’ils détectent. Les performances obtenues pour les erreurs lexicales sont meilleures pour tous les correcteurs sauf pour Pages et BonPatron. Si l’on exclut ces deux correcteurs, les

TABLE X
DÉTECTION – RAPPEL DES ENL ET DES EL POUR CHAQUE CORRECTEUR

	ENL		EL		EL – ENL
	Total	R ↓	Total	R	R
CO	1062	0,626	860	0,966	0,340
AN	1279	0,416	939	0,978	0,562
GR	1292	0,378	953	0,997	0,619
SC	1232	0,357	951	0,835	0,478
LT	1174	0,340	898	1,000	0,660
MW	1231	0,301	951	0,965	0,664
ADC	1287	0,221	930	0,992	0,771
BP	708	0,175	821	0,965	0,790
PA	1288	0,017	917	0,942	0,925

sept autres corrigent en moyenne 77,6 % des erreurs lexicales qu’ils détectent (écart-type = 10 %).

TABLE XI
CORRECTION – RAPPEL DES ENL ET DES EL PARMIS CELLES QUI SONT DÉTECTÉES POUR CHAQUE CORRECTEUR

	ENL		EL		ENL – EL
	Total	R ↓	Total	R	R
MW	371	0,919	918	0,496	0,423
AN	532	0,876	918	0,510	0,366
CO	665	0,817	831	0,573	0,244
SC	440	0,807	794	0,403	0,404
GR	488	0,727	950	0,359	0,369
ADC	285	0,646	923	0,462	0,184
LT	399	0,642	898	0,398	0,244
PA	22	0,364	864	0,513	-0,149
BP	124	0,129	792	0,360	-0,231

E. Erreurs phonologiquement plausibles (EPP) et non phonologiquement plausibles (ENPP)

1) *Performances en détection* : Le tableau XII présente le rappel en détection des erreurs phonologiquement plausibles (EPP, mots erronés qui se prononcent comme leur forme correcte) et des erreurs non phonologiquement plausibles (ENPP, mots erronés qui se prononcent différemment de leur forme correcte), ainsi que la différence entre les deux rappels. En moyenne, 72,3 % des ENPP sont détectées (écart-type = 6,3 %), tandis que seulement 51,7 % des EPP sont détectées (écart-type = 12,6 %).

TABLE XII
DÉTECTION – RAPPEL DES EPP ET DES ENPP POUR CHAQUE CORRECTEUR

	EPP		ENPP		ENPP – EPP
	Total	R ↓	Total	R	R
CO	1090	0,747	820	0,824	0,078
AN	1309	0,601	894	0,738	0,137
GR	1329	0,561	901	0,764	0,203
LT	1208	0,546	849	0,746	0,199
SC	1286	0,526	882	0,629	0,104
MW	1276	0,520	891	0,697	0,177
BP	839	0,456	678	0,779	0,322
ADC	1312	0,441	890	0,701	0,260
PA	1309	0,253	881	0,625	0,373

2) *Performances en correction* : Le tableau XIII présente le rappel en correction des erreurs phonologiquement plausibles

(EPP) et des erreurs non phonologiquement plausibles (ENPP) détectées par les correcteurs, ainsi que la différence entre les deux rappels. En moyenne, 60,8 % des EPP détectées sont corrigées (écart-type = 14 %), tandis que seulement 45,2 % des ENPP détectées sont corrigées (écart-type = 7 %).

TABLE XIII
CORRECTION – RAPPEL DES EPP ET DES ENPP PARMI CELLES QUI SONT DÉTECTÉES POUR CHAQUE CORRECTEUR

	EPP		ENPP		ENPP – EPP
	Total	R ↓	Total	R	R
CO	814	0,773	676	0,577	0,196
AN	787	0,762	660	0,506	0,256
MW	664	0,720	621	0,512	0,208
ADC	579	0,656	624	0,369	0,288
SC	676	0,641	555	0,436	0,204
GR	745	0,561	688	0,404	0,157
PA	331	0,538	551	0,495	0,042
LT	660	0,527	633	0,419	0,109
BP	383	0,298	528	0,354	-0,057

V. DISCUSSION

Les résultats de cette évaluation montrent que les performances des correcteurs orthographiques sélectionnés sont insuffisantes pour aider les personnes dys à compenser leurs difficultés en orthographe. Dans le meilleur des cas, un quart des mots erronés ne sont pas détectés et près de la moitié des mots erronés ne sont pas corrigés.

En détection, un nombre important de mots erronés ne sont pas détectés par la plupart des correcteurs. Or, dans une précédente étude de MacArthur *et al.*, réalisée auprès d'une population de faibles scripteurs, aucun des participants n'a identifié de mots erronés en plus de ceux détectés par le correcteur [10]. Il paraît donc important que les correcteurs proposés aux personnes dys identifient le plus de mots erronés possible. Les erreurs qui posent le plus de difficultés aux correcteurs sont des erreurs qui nécessitent une analyse approfondie du contexte : ajout, omission, substitution ou déplacement de mots, confusion entre homophones et erreurs non lexicales. À l'inverse les erreurs non lexicales, qui ne nécessitent pas de prise en compte du contexte, sont très bien détectées. C'est le cas également des erreurs de déplacement, de transposition ou d'omission de caractères, ainsi que des erreurs de phonétisation ou d'approximation. Ces erreurs génèrent la plupart du temps des erreurs non lexicales.

En correction, les performances sont globalement très faibles. On note cependant qu'à l'exception de Pages et BonPatron, lorsque les correcteurs détectent des erreurs non lexicales, ils parviennent à corriger la plupart de ces erreurs.

VI. CONCLUSION

Dans cet article, nous avons présenté une évaluation de neuf correcteurs orthographiques sur des textes rédigés par des personnes dyslexiques. Cette évaluation montre que ces correcteurs présentent d'importantes limites et qu'ils ne sont pas suffisants pour compenser complètement les difficultés en orthographe des personnes dyslexiques. Cependant, même une personne non dyslexique peut avoir des difficultés pour

corriger un texte rédigé par une personne dyslexique. Pour aider les personnes dyslexiques dans la tâche d'écriture et de révision de textes, il faut des solutions d'aide à l'écriture (et à la lecture) qui intègrent diverses technologies d'assistance : synthèse et reconnaissance vocale, prédiction de mots, repérage et navigation facilitée dans un grand document.

Il serait également intéressant d'étudier le potentiel des grands modèles de langues, tels que celui utilisé par ChatGPT. Ces outils permettent de générer facilement des textes orthographiquement et grammaticalement corrects. Cependant, leur utilisation soulève de nombreuses questions (présence de biais dans les données d'apprentissage, hallucinations des modèles, etc.).

Enfin, le français a une orthographe complexe. Sa simplification fait fréquemment l'objet de débat. Une réforme de l'orthographe (par exemple, la simplification de l'accord du participe passé), même si elle ne résoudrait pas toutes les difficultés en orthographe des personnes dyslexiques, serait bénéfique pour ces personnes.

RÉFÉRENCES

- [1] P. H. François, "Effets psychologiques des erreurs d'orthographe dans une lettre de motivation," *L'orientation scolaire et professionnelle*, no. 47/3, pp. 469–494, Sep. 2018, number : 47/3 Publisher : Institut national d'étude du travail et d'orientation professionnelle (INETOP). [Online]. Available : <https://journals.openedition.org/osp/8478>
- [2] C. Martin-Lacroux, "'Without the spelling errors I would have shortlisted her...': The impact of spelling errors on recruiters' choice during the personnel selection process," *International Journal of Selection & Assessment*, vol. 25, no. 3, pp. 276–283, Sep. 2017.
- [3] G. de Weck and M. Fayol, "L'orthographe en production de textes chez les enfants avec et sans dysorthographe," in *Langage & pratiques*, 2009, vol. 43, pp. 46–58.
- [4] A. Plisson, D. Daigle, and I. Montesinos-Gelet, "The Spelling Skills of French-Speaking Dyslexic Children," *Dyslexia*, vol. 19, no. 2, pp. 76–91, May 2013.
- [5] D. Daigle, A. Costerg, A. Plisson, N. Ruberto, and J. Varin, "Spelling Errors in French-speaking Children with Dyslexia : Phonology May Not Provide the Best Evidence," *Dyslexia (10769242)*, vol. 22, no. 2, pp. 137–157, May 2016, publisher : John Wiley & Sons, Inc.
- [6] A. Mazur-Palandre, "La dyslexie à l'âge adulte : la persistance des difficultés orthographiques," *SHS Web of Conferences*, vol. 46, p. 10003, 2018, dyslexia in adulthood : the persistence of orthographic difficulties. [Online]. Available : https://articles/shsconf/abs/2018/07/shsconf_cmlf2018_10003/shsconf_cmlf2018_10003.html
- [7] J.-Y. Antoine, M. Crochetet, C. Arbizu, E. Lopez, S. Pouplin, A. Besnier, and M. Thebaud, "Ma copie adore le vélo : analyse des besoins réels en correction orthographique sur un corpus de dictées d'enfants," in *TALN 2019*, Toulouse, France, Jun. 2019. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-02375246>
- [8] M. Chodorow, M. Dickinson, R. Israel, and J. Tetreault, "Problems in Evaluating Grammatical Error Detection Systems," in *Proceedings of COLING 2012*. Mumbai, India : The COLING 2012 Organizing Committee, Dec. 2012, pp. 611–628. [Online]. Available : <https://aclanthology.org/C12-1038>
- [9] J. Bodard, "Corpus DYS," Mar. 2022. [Online]. Available : <https://www.nakala.fr/10.34847/nkl.ced0370u>
- [10] C. A. MacArthur, S. Graham, J. B. Haynes, and S. DeLaPaz, "Spelling Checkers and Students with Learning Disabilities : Performance Comparisons and Impact on Spelling," *The Journal of Special Education*, vol. 30, no. 1, pp. 35–57, Apr. 1996. [Online]. Available : <http://journals.sagepub.com/doi/10.1177/002246699603000103>