

L'anti-correcteur : outil d'évaluation positive de l'orthographe et de la grammaire

Lydia-Mai Ho-Dac¹ Sophie Muller² Valentine Delbar²

(1) CLLE-ERSS, 5 Allées A. Machado, 31058 Toulouse, France

(2) Synapse Développement, 5 Rue du Moulin Bayard, 31000 Toulouse, France

hodac@univ-tlse2.fr.fr, {sophie.muller, valentine.delbar}@synapse-fr.com

RÉSUMÉ

L'objectif de cette étude est d'expérimenter l'intégration d'une nouvelle forme d'évaluation dans un correcteur orthographique et grammatical. L'« anti-correcteur » a pour objet de mesurer le taux de réussites orthographiques et grammaticales d'un texte sur certains points jugés difficiles selon la littérature et une observation d'erreurs en corpus. L'évaluation du niveau d'écriture ne se base plus uniquement sur les erreurs commises, mais également sur les réussites réalisées. Une version bêta de ce nouveau mode d'évaluation positive a été intégré dans le correcteur Cordial. Cet article a pour but de discuter de l'intérêt de ce nouveau rapport à l'orthographe et de présenter quelques premiers éléments d'analyse résultant de l'application de l'anti-correcteur sur un corpus de productions variées en matière de niveau d'écriture et genre discursif.

Ici, un résumé en français (max. 150 mots). Times, 10pt.

ABSTRACT

The "anti-correcteur": a positive evaluation module for spell and grammar checking

This study aims at testing out a new form of evaluation for spell and grammar checking. A new tool, called "anti-correcteur", was integrated in Cordial, a French spell and grammar checker, for measuring success rates in common spelling difficulties defined according to literature in French language teaching and corpus-based analysis. This module proposes to assess spelling skills not only against errors, but also by taking successes into account. This paper presents a first experiment of such a positive evaluation by exploring results given by the "anti-correcteur" applied on a diversified corpus in terms of level of literacy and genre.

MOTS-CLÉS : correction orthographique, correction grammaticale, évaluation positive, didactique du français écrit, zone de résistances, analyse de corpus diversifié

KEYWORDS: spell checking, grammar checking, positive evaluation, written French language teaching, resistance zone, diversified corpus analysis

Notre étude présente l'anti-correcteur, un outil d'évaluation positive fondé sur la détection des réussites orthographiques. Après avoir indiqué les appuis théoriques et les enjeux de l'évaluation positive, nous proposons une typologie des réussites orthographiques à détecter, puis présentons le fonctionnement et une première exploitation des résultats de la version bêta de l'anti-correcteur.

Vers une évaluation positive des pratiques orthographiques

De nombreuses études indiquent une baisse du niveau de maîtrise de l'orthographe par les français (cf. Martin Lacroux, 2015, Andreu et al., 2014 ; Manesse, Cogis, 2007) et l'apparition d'une fracture sociale entre d'un côté une « élite » qui écrit correctement et de l'autre des francophones qui ne maîtrisent pas correctement l'écrit (Chervel, 2008). Face à ce constat, de nouveaux dispositifs didactiques sont envisagés permettant à la fois une amélioration du niveau de maîtrise de l'orthographe et une dédramatisation de celle-ci. C'est dans ce contexte qu'est apparue la notion d'évaluation positive (Brissaud, Mortamet (eds), 2015).

Ces nouveaux dispositifs reposent sur l'idée que la description des réussites orthographiques permet de mesurer l'acquisition de compétences d'écriture tout en constituant une aide à l'apprentissage (Cogis et al., 2015) en portant un regard différent sur l'erreur (Gonac'h, 2015). Fisher et Nadeau (2014) montrent ainsi les effets bénéfiques des « dictées innovantes » qui proposent une correction collective des productions d'élèves, poussant ceux-ci à justifier leurs choix de graphie. D'autres projets font varier les situations de production et décrivent l'erreur non seulement comme un symptôme d'une mauvaise application d'une règle mais surtout comme une variation liée à plusieurs facteurs – le contexte stressant de la dictée par exemple (Anxionnaz, 2015, Gonac'h, 2015, Lenfant, 2015). Enfin, des dispositifs didactiques proposent des évaluations « plus justes », comme le « barème graduel » (Anxionnaz, 2015) qui permet d'évaluer le taux de réussite sur trois difficultés reconnues du français : l'accord verbal, l'accord au sein du groupe nominal et le lexique. L'enseignant identifie les « zones d'erreur potentielles » et comptabilise les zones correctement écrites. La progression de l'élève peut ainsi être évaluée et servir de moteur pour une amélioration.

Notre étude propose d'adapter l'idée d'évaluation positive à un contexte de production différent de celui du milieu scolaire : l'utilisation d'un correcteur orthographique et grammatical, dans un environnement numérique, par une population plutôt adulte. Cette proposition semble séduire un ensemble varié d'utilisateurs, comme l'indique une enquête de satisfaction menée sur le site www.cordial-enligne.fr/ et sur la page Facebook du correcteur du 24 au 30 juin 2015. Cette enquête a été réalisée auprès de locuteurs du français langue maternelle dont 79 % de femmes, 63 % de personnes entre 18 et 34 ans, 26 % entre 35 et 50 ans et 10 % ayant plus de 51 ans. Sur les 239 réponses obtenues, 87 % des participants se disent tentés par l'utilisation d'un outil évaluation positive, et ce, quelle que soit leur opinion sur leurs problèmes en orthographe qui, pour 80 % d'entre eux, sont davantage liés à de l'étourderie plutôt qu'à des difficultés récurrentes de français. Bien que cette fonctionnalité semble séduire les utilisateurs, aucun correcteur orthographique et grammatical n'a aujourd'hui proposé ce type de système.

« Zones de résistances » et réussites orthographiques

Un outil d'évaluation positive repose en premier lieu sur la détection des « zones d'erreur potentielles ». En effet, le calcul du nombre de non-erreur (taux de réussite) ou d'erreur (taux d'échec) ne peut se faire sans identification des candidats à l'erreur. Afin de définir ces candidats, nous utilisons la notion de « zones de résistances » définies par Roubaud (2014) comme des « secteurs linguistiques [qui] résistent à l'enfant » (op. cit : 120). Ces zones de résistances seraient définies à la fois par la complexité d'une règle et la fréquence d'une erreur par rapport à cette règle.

Les études des erreurs manuscrites comme tapuscrites sont variées tant au niveau de l'objet que de la population étudiée : Souque (2009, 2014) s'intéresse aux écrits d'étudiants, Laurent (2012) aux erreurs observées dans la Wikipédia française, Katoozian (2014) aux écrits d'apprenants en Français Langue Maternelle et Roubaud (2014) aux productions d'élèves du primaire. Cette série d'études permet de définir les zones de résistances comme zones sur lesquelles des erreurs sont fréquemment commises que ce soit chez des enfants ou des adultes. Leur identification repose donc sur l'identification de types d'erreurs fréquentes en français.

À cette notion de zone de résistance qui relève de la difficulté de réussir s'ajoute celle de la gravité ressentie à la lecture d'une erreur. En 2012, le Groupe R.O.¹ (2012) a mené une enquête auprès d'enseignants et d'élèves afin de relever leur sentiment vis-à-vis des difficultés de l'orthographe du français. L'enquête note par exemple que pour une majorité de personnes interrogées, une erreur d'accord du participe passé est mieux acceptée car jugée plus difficile (ou moins évidente) qu'une erreur d'homophonie de type *trait vs. très*. Cela renvoie à la notion d'acceptabilité soulignée par Souque (2014) qui explique qu'un énoncé agrammatical, soit erroné, peut être confronté à un énoncé dont la grammaticalité est douteuse, soit un énoncé dont l'acceptabilité est remise en question. Tandis que certaines erreurs sont donc acceptables, car jugées difficiles à éviter au vu des nombreuses règles qui les régissent, d'autres sont moins acceptables.

1 « Zones de résistances » connues et attestées

En croisant les zones qui résistent identifiées par Roubaud (2014), les « réussites » à mesurer dans le barème graduel d'Anxionnaz (2015) et les difficultés peu acceptables relevées par l'enquête du Groupe R.O. (2012), les zones de résistances principales peuvent être classées en cinq secteurs :

- segmentation en phrases ;
- accord verbal (dont les difficultés portant sur l'accord du participe passé) ;
- accord au sein du groupe nominal ;
- orthographe lexicale : doubles consonnes, noms composés, finales muettes, accentuation ;
- homophonie/homonymie (e.g. *à vs. a, ce vs. se, on vs. ont, et vs. est*).

2 Observatoire des erreurs sur Cordial-en-ligne

Afin de mesurer la fréquence d'apparition de ces zones de résistance, nous exploitons les données fournies par Cordial-en-ligne, correcteur orthographique qui permet de corriger gratuitement ses écrits par soumission en ligne (www.cordial-enligne.fr/). L'observation porte sur les erreurs d'une population adulte qui n'est pas particulièrement confiante avec l'orthographe et la grammaire, mais qui reste soucieuse de bien écrire. Cette population, assez différente de celles des études décrites dans la section précédente, est sans doute plus représentative des utilisateurs d'outils de correction.

En 3 mois, 8 487 257 erreurs ont été récoltées, dont 21,6 % qui nous intéressent car directement liées aux zones de résistance listées précédemment et pouvant donner lieu à des stratégies d'amélioration. Le tableau 2 donne la fréquence de ces erreurs « résistantes ». Les 78,4 % restants relèvent principalement d'erreurs jugées plutôt acceptables, e.g. usage de la virgule (8 %), espaces

¹ « Rectifications Orthographiques », rassemblement de plusieurs universitaires.

autour des signes de ponctuation (12,5 %), style « à revoir » (3,6 %). Cette répartition a été réalisée automatiquement à l'aide de l'outil de catégorisation des erreurs du correcteur utilisé.

	% des erreurs		nombre de cas	
doubles consonnes	0,4		34 676	
finales muettes (<i>noms en -é(e)(s)</i>)	0,7	(0,03)	57 594	(2 612)
phrase incomplète	6,8		577 147	
accord verbal	2,0		172 842	
accord au sein du groupe nominal	4,4		370 710	
homophonie/homonymie (<i>a vs. à</i>)	3,8	(1,4)	320 611	(120 992)
noms composés avec trait d'union	0,5		46 231	
accentuation	2,9		246 764	
Total des erreurs	21,6		1 826 575	

TABLE 2 : Répartition des erreurs dans le corpus Cordial-en-ligne

Parmi les erreurs « résistantes », les plus fréquentes sont les problèmes d'homonymie (dont les problèmes d'accentuation font partie), les accords au sein du groupe nominal et les difficultés à faire des phrases complètes. L'accord verbal, souvent mentionné dans les difficultés du français, ne représente finalement que 2 % des erreurs et les problèmes de doubles consonnes que 0,4 %. Cette répartition semble montrer que parmi les « zones de résistance » identifiées auprès des enfants certaines résistent encore à l'adulte comme l'accord du groupe nominal et la gestion des accents alors que d'autres semblent plutôt bien maîtrisées (accord verbal, trait d'union, doubles consonnes). Ce corpus d'erreurs attestées a également servi de base pour la constitution de six listes utilisées par l'anti-correcteur pour détecter les différentes zones de résistance (cf. section suivante) :

- L_TDU : 9 948 mots entre lesquels un tiret est fréquemment ajouté à tort
- L_NCP : 449 noms composés pluriels qui sont fréquemment mal orthographiés
- L_ACC : 19 331 mots fréquemment mal accentués
- L_CON : 10 438 mots fréquemment mal orthographiés à cause d'une double consonne, soit par oubli de la double consonne, soit par dédoublement erroné d'une consonne
- L_EES : 339 noms se terminant en -é(e)(s) fréquemment mal orthographiés
- L_SIM : 360 957 mots homonymes et paronymes fréquemment confondus

Fonctionnement de l'anti-correcteur

L'anti-correcteur a deux objectifs : (1) détecter des zones de résistances orthographiques et (2) vérifier que l'orthographe est correcte par rapport à la difficulté détectée. En d'autres termes, l'anti-correcteur cherche si du texte propice à une erreur "A" a été saisi et détecte si l'erreur "A" a été commise. Si l'erreur "A" n'est pas commise un message de félicitation est délivré. Par exemple, si le scripteur a écrit du texte comportant des groupes nominaux et qu'il n'a pas fait d'erreur d'accord du groupe nominal, l'outil félicite le scripteur sur ce point précis.

La version bêta de l'« anti-correcteur » traite 10 zones de résistance :

- **Z_PHR** : phrase complète (*c'est pas possible!*)
- **Z_TDU** : Trait d'union (*Que fait tu là princesse ?*)
- **Z_NCP** : Noms composés pluriels (*elle baissa les yeux sur les chou-fleurs.*)
- **Z_ACC** : Accentuation (*etre*)
- **Z_GNP** : Accord du groupe nominal masculin pluriel (*que de bon films*)
- **Z_APP** : Accord avoir + participe passé (*il les a amené.*)
- **Z_AvA** : Mot similaire à vs. a (*Elle apprenait à ce battre, a manier les armes.*)
- **Z_CON** : Doubles consonnes (*apelle*)
- **Z_EES** : Noms se terminant en -é(e)s (*activités*)
- **Z_SIM** : Mots similaires (*Désoler, j'ai étais un peu souffrante la semaine dernière.*)

Ces zones sont détectées par les « déclencheurs » suivant :

- présence d'une chaîne de caractères spécifique (Z_TDU, Z_AvA, Z_CON)
- présence d'un mot inclus dans une des listes d'erreurs fréquentes décrites dans la section précédente (Z_TDU, Z_NCP, Z_ACC, Z_CON, Z_EES, Z_SIM)
- analyse syntaxique détectant un groupe nominal (Z_GNP),
- analyse syntaxique détectant un verbe conjugué à un temps composé (Z_APP)
- déclenchement par défaut (Z_PHR). Une phrase est réputée complète tant que qu'aucun élément indispensable à sa formation n'est identifié comme manquant.

L'analyse syntaxique est réalisée par l'analyseur interne du correcteur basée sur des règles d'analyse floues prenant en compte la position des mots dans la phrase et les régularités statistiques (Laurent et al., 2015).

Pour chaque zone, l'affichage du message de félicitation est annulé si une erreur relative à la difficulté en jeu est détectée par le correcteur. Le détail de ces « annulateurs » est donné ci-dessous.

- **Z_PHR** : absence de verbe, négation incomplète, absence des COD et/ou COI d'un verbe (bi)transitif ou absence d'un déterminant devant les noms sujets
- **Z_TDU** : présence erronée ou absence d'un trait d'union dans un mot inconnu du dictionnaire ou une construction syntaxique en nécessitant un (e.g. interrogative)
- **Z_SIM** : confusion entre des homophones/homographes
- **Z_AvA** : confusion entre "à" et "a" (également détecté comme Z_SIM)
- **Z_GNP** : erreur d'accord en nombre et/ou genre dans un groupe nominal pluriel
- **Z_ACC** : erreur d'accentuation dans un mot inconnu du dictionnaire, confusion entre formes verbales ou problème d'accentuation d'une majuscule en début de mot
- **Z_CON** : doublement erroné de consonne ou oubli d'une consonne double
- **Z_EES** : mot inconnu dont la suggestion de correction est contenue dans la liste L_EES
- **Z_NCP** : mot inconnu dont la suggestion de correction est contenue dans la liste L_NCP
- **Z_APP** : erreur d'accord d'un participe passé précédé de l'auxiliaire avoir conjugué.

Première exploitation des résultats de l'anti-correcteur

La version bêta de l'anti-correcteur n'a pas encore fait l'objet d'une enquête de satisfaction mais nous proposons dans cette section une première exploitation des résultats qui permet de donner un aperçu de la couverture de l'outil en mesurant la répartition des zones de résistance et de leur taux de réussite dans un corpus diversifié. Un corpus de test de 1 867 063 mots a ainsi été constitué afin de représenter différents types de population dans différentes situations de communication :

- **écrits d'étudiants** de niveau différents (Rinck, Jacques, 2015) répondant à des consignes différentes : dissertation niveau L2 (27 152 mots), notes de lecture niveau M2 (39 843 mots), lettres de motivation niveau M1 (6 552 mots), rapports de stage niveau L3 (66 857 mots) ;
- **posts de discussions** issus de Wikipedia (64 425 mots, Ho-Dac, Laippala, 2015), de « La Cabane »² du projet Vikidia (32 861 mots) et du forum de santé enceinte.com (133 264 mots) ;
- **articles de presse** issus de Rue89 (566 282 mots) et du média citoyen AgoraVox (929 827 mots) (Ho-Dac, Küppers, 2011).

Le corpus a été segmenté en phrases et pour chaque phrase l'anti-correcteur indique si une zone a été détectée et si cette zone a donné lieu à une félicitation. Une évaluation manuelle de la détection des zones Z_APP (Accord *avoir* + participe passé) et Z_AvA (Mot similaire à *vs.a*) a été réalisée sur 349 phrases sélectionnées aléatoirement dans tous le corpus. La détection des zones Z_APP montre une précision de 100 % (81 cas) et un rappel de 96,4 % (3 zones ratées). Celle des zones Z_AvA, 100 % de précision et de rappel. La détection d'erreur sur ces zones n'a pas été évaluée. Le tableau Error: Reference source not found indique pour chaque type d'erreur :

- la proportion de phrases du corpus total présentant au moins une zone de résistance
- le corpus affichant significativement plus/moins de zones ($z > 2$)
- le taux de réussite sur la zone
- Corpus affichant le plus mauvais taux de réussite (aucun $z > 2$)

	Phrases contenant au moins une zone			Taux de réussite		
	Nb	Global (%)	Min (%)	Max (%)	Global (%)	Min (%)
Z_PHR	86761	100			87	
Z_TDU	70046	80	Vikidia (55)		96	
Z_NCP	2935	3		Rapport (10)	95	Forum (91)
Z_ACC	82839	95			92	Forum (68)
Z_GNP	47380	54			98	Forum (77)
Z_APP	15590	18		Rapport (35)	100	Forum (93)
Z_AvA	34954	40			98	Forum (77)
Z_CON	84190	97			98	
Z_EES	15404	18			99	
Z_SIM	84412	97			96	Forum (79)

TABLE 2 : Zones de résistances détectées et taux de réussite mesurés sur le corpus de test

² « le coin des discussions des jeunes vikidiens » réservé aux 7-14 ans

Les premiers éléments d'analyse montrent que les zones les plus fréquentes sont relatives aux difficultés d'ordre lexical (Z_ACC, Z_SIM, Z_CON). Certaines zones de résistance (Z_NCP, Z_APP, Z_EES) sont rarement détectées (entre 3 et 18 % des phrases) et présentent un taux de réussite proche de 100 %. De façon globale, le taux de réussite est assez bon (entre 87 % et 100 %). L'accentuation est le secteur sinistré de l'orthographe avec le plus faible taux de réussite global (92 %). Les sous-corpus présentent finalement peu de variations. Les zones Z_APP sont plus fréquentes dans les rapports de stage (35 % de phrases), les lettres de motivation (22 %) et la presse (20 % chez Rue89). Elles sont plutôt sous-employées dans les dissertations (7 %), les discussions Vikidia (8 %) et Wikipedia (9 %). Deuxième constatation, les zones Z_TDU disparaissent significativement des discussions Vikidia (55 %) et dans une moindre mesure des discussions Wikipedia (63 %).

Les forums de santé obtiennent les plus mauvais taux de réussite dans presque toutes les catégories (e.g. 68 % pour Z_ACC, 77 % pour Z_AvA et Z_GNP, 79 % pour Z_SIM). Les discussions Vikidia et Wikipedia montrent des lacunes au niveau des zones Z_CON (96 % et 95 %) et Z_EES (92% et 94 %), sans pour autant que ces écarts soient significatifs. Les corpus présentant les meilleurs taux de réussite sont les lettres de motivation, les écrits journalistiques et les rapports de stage. Si l'on prend en compte conjointement les taux de zones de résistances et les taux de réussite, il semblerait que ceux qui prennent et maîtrisent le plus de risques en matière de zone de résistance soient les rapports de stage et les écrits journalistiques, à l'inverse des lettres de motivation qui affichent des indices de prise de risque moindre.

Conclusion

Nous avons décrit dans cet article la première version d'un outil d'évaluation positive de l'orthographe. À notre connaissance, l'anti-correcteur est le premier à proposer un service d'évaluation positive de l'orthographe automatique qui ne s'adresse pas spécifiquement à des élèves apprenants. Le fait de baser la sélection des zones de résistances sur une analyse de corpus et une revue de littérature lui permet d'être un outil personnalisé pour des scripteurs francophones qui prend en compte les erreurs effectivement commises.

La liste des réussites détectées est amenée à être augmentée en fonction de l'intérêt des utilisateurs. Le système de détection bêta des zones de résistance n'évalue que peu de types d'erreurs. Par conséquent, les résultats obtenus ne sont pas représentatifs du niveau d'écriture des auteurs du fait de l'absence de couverture de l'ensemble des fautes relevées par le correcteur. Malgré cet état bêta du module, nous avons tout de même pu implémenter des zones de résistances touchant à différents niveaux de maîtrise du français et proposer une première analyse en corpus des réussites orthographiques.

Notons enfin que la taille du document n'influence en rien l'apparition des messages de félicitations. En effet, le principe est de féliciter le scripteur dès les premières phrases élaborées. Certes, une petite faute dans un document long ne reflète pas une diminution des compétences orthographiques, mais le parti pris est d'attirer néanmoins son attention sur la relecture de ce point en particulier.

Références

- ANDREU S., LE CAM M., ROCHER T. (2014). Evolution des acquis en début de CE2 entre 1999 et 2013 : les progrès observés à l'entrée au CP entre 1997 et 2011 ne sont pas confirmés, *Note d'information* 19, DEPP.
- BRISSAUD C. (2015). Entretien : La place de la dictée dans les évaluations conduites par la DEPP. *Glottopol* 26, 127-134.
- BRISSAUD C., MORTAMET C. (eds) (2015). La dictée, une pratique sociale emblématique, *Glottopol* 26.
- BOULLAUD C., CHANQUOY L., GOMBERT J. É. (2007). Cyberlangage et orthographe: quels effets sur le niveau orthographique des élèves de CM2, 5ème et 3ème?, *Bulletin de psychologie* 6, 553-565.
- COGIS D. (2005). *Pour enseigner et apprendre l'orthographe*. Paris : Delagrave.
- COGIS D., FISHER, C., NADEAU, M. (2015). Quand la dictée devient un dispositif d'apprentissage. *Glottopol* 26, 69-91.
- DEWAELE B., COLOMB A. (2011). *Maîtrisez l'orthographe avec la Certification Voltaire*. Paris : Editions Eyrolles.
- FISHER C., NADEAU M. (2014). Usage du métalangage et des manipulations syntaxiques au cours de dictées innovantes dans des classes du primaire, *Repères* 49, 169-191.
- GONAC'H J. (2015) Des variations orthographiques dans les dictées de CM1. *Glottopol* 26, 167-177.
- GROUPE R.O. (2012). Orthographe: ce qui est jugé difficile. L'avis d'enseignants et de futurs enseignants. *Glottopol* 19, 17-36.
- HO-DAC L.-M., LAIPPALA V. (2015). Les discussions Wikipédia : un corpus pour caractériser le genre « discussion ». *International Research Days: Social Media and CMC Corpora for the e-Humanities*, Oct 2015, Rennes, France.
- HO-DAC L.-M., KÜPPERS A. (2011). La subjectivité à travers les médias : étude comparée de les médias participatifs et de la presse traditionnelle. *Corpus, varia* 10, 179-200.
- KATOOZIAN K. (2014). Une revue d'articles sur les erreurs orthographiques à dominante phonétique : réflexions pour le FLE. *Linguarum Arena* 5, 31-45.
- LAURENT D. (2012). *Les vraies difficultés du français au XXIe siècle*. Toulouse : Éditions Synapse Développement.
- LAURENT D., NÈGRE S., SÉGUÉLA P. (2009). L'analyseur syntaxique Cordial dans Passage. *Actes de TALN*, 9.
- MANESSE D., COGIS D. (2007). *Orthographe : à qui la faute ?*, Paris : ESF.

MARTIN LACROUX C. (2015). *L'appréciation des compétences orthographiques en phase de présélection des dossiers de candidature : pratiques, perceptions et implications pour la GRH*. Thèse de doctorat en Gestion et management. Université de Toulon. <[tel-01236486](tel:01236486)>

RINCK F., JACQUES M.-P. (2015). Corpus de littéracie avancée : structuration et méta-données. Journée d'études Analyse linguistique de grands corpus d'écrits scolaires. Actes de *Problèmes de transcription, d'annotation et de traitement*, Sorbonne Nouvelle Paris 3, 18 mars 2015. <http://www.univ-paris3.fr/activites-ecriscol-300518.kjsp>

ROUBAUD M. N. (2014). *De la description de la langue à son enseignement*. Thèse d'habilitation à diriger des recherches en Linguistique, Université de Stendhal–Grenoble 3.

SOUQUE A. (2009). Constitution d'un corpus d'erreurs du dactylographe. Communication présentée aux 6e *Journées Internationales de Linguistique de Corpus*, Lorient.

SOUQUE A. (2014). *Modèle de vérification grammaticale automatique gauche-droite*. Thèse de doctorat en linguistique. Université Grenoble Alpes. <[tel-01247368](tel:01247368)>