

# De l'inégalité des consonnes doubles en français face à l'erreur orthographique

*Jean-Luc Manguin*

Laboratoire GREYC, CNRS – UMR 6072, Université de Caen, 1 4052 Caen  
Cedex, France  
[jean-luc.manguin@unicaen.fr](mailto:jean-luc.manguin@unicaen.fr)

**Résumé.** Cet article étudie les erreurs d'omission dans les motifs à consonne double des mots français. Il se base sur les fréquences des formes correctes et erronées en corpus, et vise à dégager les facteurs qui favorisent ou freinent l'apparition de ces erreurs, en faisant référence aux travaux de psycholinguistique sur l'apprentissage implicite de l'orthographe. On y observe ainsi les influences de la fréquence du motif, de celle de son concurrent, de la difficulté du contexte et de la présence du motif dans un morphème grammatical.

**Abstract. On the disparity of the double consonants in French in the problem of spelling errors.** In this article we study the omission errors in the French words with double consonant patterns. We rely on the frequencies of the correct and erroneous forms in corpora, and we aim at identifying the factors which favor or slow down the appearance of these errors, with reference to works of psycholinguists on implicit learning of spelling. We also detail here the influences of the frequency of the pattern, that of its concurrent, the difficulty of the context and the presence of the pattern in a grammatical morpheme.

## 1 Introduction

Depuis de nombreuses années, les erreurs orthographiques constituent un riche champ d'études, que l'on s'intéresse à leur correction par le traitement automatique des langues, depuis les premiers travaux de Damereau [Damereau, 1964], jusqu'à la thèse récente de Marion Baranes [Baranes, 2015], ou que l'on essaye de les expliquer par des mécanismes cognitifs [Borchardt, 2012]. A ce point de vue, il y a eu depuis quelques années une remise en cause de la prédominance de la phonologie, et l'idée d'une relative indépendance de l'orthographe a été étayée par de nombreux travaux dans les langues dont le système orthographique est peu consistant, comme le français ou l'anglais [Fayol et Jaffré, 2016]. En particulier, sur les consonnes doubles du français, il a été montré que l'apprentissage des régularités graphotactiques se fait de manière implicite par l'intermédiaire de la lecture, et que par conséquent les fréquences des motifs jouent un rôle à ce niveau [Pacton et al.

2001]. Cet apprentissage implicite des statistiques de la langue peut toutefois, en français en particulier, être gênant dans certains cas et nécessiter une instruction explicite pour certaines formes en contradiction avec ces régularités [Pacton et al. 2005]. Les erreurs commises étant révélatrices de ces mécanismes complexes, une étude en corpus apparaît intéressante pour compléter les données obtenues en production chez les jeunes enfants. Dans le cas du français, nous savons que les erreurs les plus fréquentes sont, après les erreurs d'accentuation, celles qui concernent les consonnes doubles [Lucci et Millet, 1994] ; nous avons donc choisi de nous intéresser à cette difficulté orthographique, d'une manière plus élargie que par le passé, où nous n'avions étudié que le seul motif « onn » [Manguin, 2009], mais dans le seul but de valider les données de notre base d'erreurs orthographiques. Cette fois, nous embrassons la totalité des motifs à consonnes doubles du français, sous le seul aspect de l'omission, autrement dit de la transformation de la consonne double en consonne simple. Il va de soi qu'une étude exhaustive des autres phénomènes liés aux consonnes doubles (ajout, triplement, décalage, compensation) dépasserait probablement l'espace imparti à cette communication.

D'autre part, notre intérêt pour l'omission se justifie par le besoin d'élargir – si possible – les résultats obtenus sur le motif « onn » à propos du contexte intra-mot. Nous avons en effet remarqué que la consonne double est traitée de manière inégale par le scripteur suivant certains contextes (que nous préciserons par la suite) ; ici, nous allons montrer que d'autres facteurs, comme la fréquence absolue, l'impact phonologique, ou l'appartenance à un morphème flexionnel jouent des rôles importants dans l'apparition d'une erreur d'omission dans le motif à consonne double.

## **2 Données et corpus**

Cette étude s'appuie sur les données de la base Ortholexies, présentée en juillet 2016 à Tours lors du Congrès Mondial de Linguistique Française [Manguin, 2016], ainsi que sur la partie française du corpus Wacky [Baroni et al., 2009]. Ortholexies est une base de données qui contient 556 000 formes distinctes, dont plus de 200 000 ont été traitées ; ces dernières se répartissent en 128 000 formes correctes et 75 000 formes erronées, appariées à leur(s) correction(s). A chaque forme sont associées deux fréquences, la première étant le nombre de fois où elle a été demandée en tant que requête sur le dictionnaire des synonymes du Crisco [Crisco, 1998], et la seconde provient des données fréquentielles de la partie française du corpus Wacky, constitué de textes tirés du Web ; cette partie française contient environ 1,6 milliard de mots.

## **3 Méthode**

Dans le but déjà mentionné de compléter les études précédentes, nous avons adopté la même méthode, à savoir que pour chaque mot et ses formes erronées, nous faisons le relevé de leurs fréquences absolues dans la base et dans le corpus. Nous travaillerons ensuite sur les pourcentages d'erreurs associés à chaque forme correcte, c'est-à-dire par exemple, que si pour une forme correcte qui possède une consonne double, on relève plusieurs formes erronées (« plusieurs » en raison des erreurs d'accents éventuelles) dans lesquelles cette consonne double est devenue simple, nous faisons le total des fréquences de ces formes erronées, puis nous divisons ce total par la fréquence de la forme correcte pour obtenir ses pourcentages d'erreurs dans la base et dans le corpus.

Bien évidemment, certains mots sont exclus de ces comptages ; tout d'abord, les couples de formes correctes qui ne diffèrent que par le doublement d'une consonne, comme par

exemple « casser » et « caser ». Ensuite, certaines formes qui présentent deux formes correctes possibles en raison de la récente réforme de l'orthographe, comme « interpellier » et « interpeler ».

Enfin, précisons que nous avons étudié les mots à une ou deux consonnes doubles, soit en détail :

- 5149 mots à une consonne double et leurs 6211 formes erronées.
- 1514 mots à deux consonnes doubles et leurs 2216 formes erronées.

Les mots à deux consonnes doubles sont examinés dans la deuxième partie de cette étude, consacrée à l'influence du contexte intra-mot sur l'apparition des erreurs.

Enfin, répétons qu'il y a plus de formes erronées que de formes correctes en raison des éventuelles erreurs d'accents ; par exemple, « aberration » possède deux formes erronées : \*aberation et \*abération. A cela s'ajoutent les possibilités combinatoires pour les mots à deux consonnes doubles ; on trouve ainsi le cas de « accessible » qui nous offre 4 formes erronées : \*accesible, \*accessible, \*accésible et \*accèsible.

## 4 Rôle de la fréquence

Commençons par quelques remarques d'ordre quantitatif ; les ensembles de mots mentionnés précédemment représentent chacun un nombre d'occurrences qui est donné dans le tableau ci-après :

**Tableau 1** : nombre de mots étudiés et leurs occurrences

	Nb de mots	Nb occ. dans la base	Nb occ. dans le corpus
Corrects, 1 c. double	5149	23 588 676	70 526 104
Erronés, 1 c. double	6211	1 040 839	599 515
Corrects, 2 c. doubles	1514	1 336 426	3 288 340
Erronés, 2 c. doubles	2216	184 482	35 524

On peut déjà remarquer que les mots à deux consonnes doubles sont bien plus rares (environ 20 fois) que les mots à une consonne double. C'est pourquoi, dans une première approche, nous allons nous contenter d'étudier les mots à une consonne double ; les mots qui en possèdent deux seront examinés dans la deuxième partie de cette étude.

La répartition en nombre d'occurrences (et pas en nombre de mots distincts) montre clairement qu'un groupe de 4 consonnes (b,d,g et z) se sépare des autres par une différence d'un ordre de grandeur (voire même deux), et ceci pour les deux sources de données.

**Tableau 2** : nombre d'occurrences des consonnes doubles

lettre	N base	N corpus
s	5282487	14820571
l	3578720	11717113
n	2731945	10317510
f	2422295	5113778
t	2393011	6572845
p	2178850	5280834
r	1915238	4710048
m	1679887	8934429
c	1215706	2649415
g	89336	127973
d	41463	32195
b	14597	30717
z	12603	18396

Parallèlement à cela, la répartition des pourcentages d'erreurs dans les corpus en fonction de la consonne redoublée fait ressortir ce même groupe de 4 consonnes, dont les pourcentages d'erreurs sont plus élevés que ceux des autres consonnes ; cette différence est plus flagrante dans le corpus textuel, où l'on a un seuil très net entre les deux groupes, avec un passage de 0,54 % (lettre r) à 0,76 % (lettre b).

**Tableau 3** : taux d'erreur des consonnes doubles

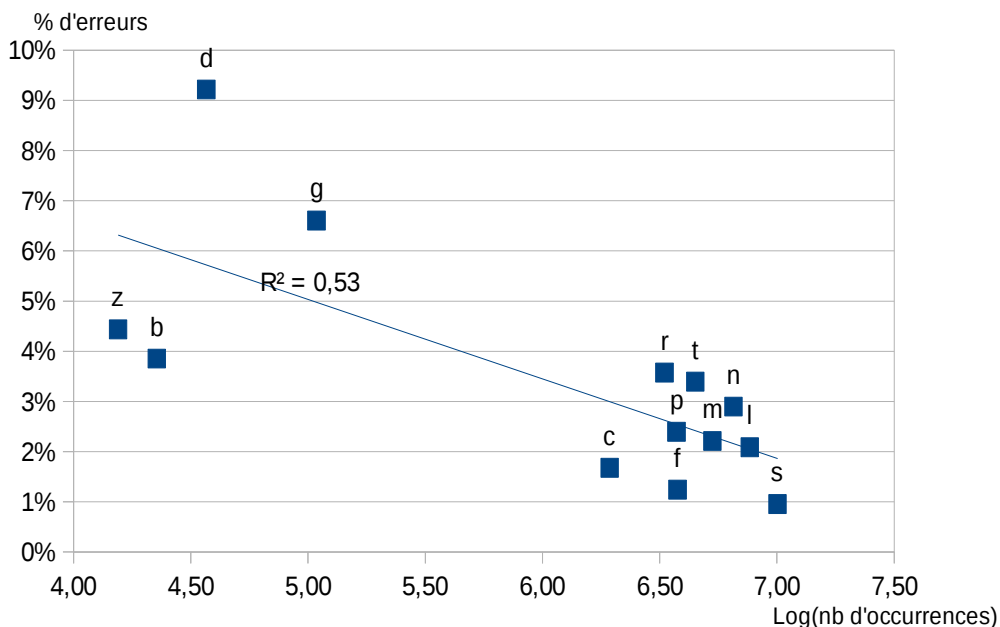
lettre	% base	% corpus
s	1,77%	0,15%
f	2,33%	0,15%
c	3,19%	0,17%
l	3,95%	0,22%
t	4,12%	0,31%
p	4,52%	0,27%
m	5,55%	0,25%
n	6,52%	0,27%
r	6,61%	0,54%
g	6,68%	1,03%
z	8,06%	0,82%
d	11,19%	2,03%
b	17,68%	0,76%

On peut donc dire dans un premier temps, que pour les consonnes doubles, la fréquence d'occurrence joue un rôle crucial sur l'apparition des erreurs, et que plus le motif est rare, plus il est exposé au risque d'erreur. Ceci corrobore les conclusions des études psycholinguistiques comme par exemple la thèse de Gaëlle Borchartd [Borchartd, op. cit. ]

ou celle de Marie-Claire Hazard [Hazard, 2009], et plus généralement les idées développées par Patrick Bonin [Bonin, 2007].

Il convient maintenant de préciser cet effet de manière quantitative ; pour cela, nous allons étudier le pourcentage d'erreurs en fonction du nombre d'occurrences du motif. Comme nous disposons de deux sources de données, nous employons les moyennes, et en raison de la différence d'ordre de grandeur des occurrences, nous prenons le logarithme du nombre d'occurrences. Le résultat est montré dans la figure ci-après.

**Figure 1** : taux d'erreur en fonction du nombre d'occurrences



Le graphique montre clairement la séparation du groupe des « consonnes doubles rares », et permet, grâce à la droite de régression, d'affirmer que l'effet inverse de la fréquence d'occurrences sur le taux d'erreurs est approximativement logarithmique. Cet effet logarithmique assez classique confirme un effet que nous avons déjà observé lors de travaux applicatifs de la base Ortholexies, présentés lors du congrès de l'Association for Psychological Science en 2013 [Le Goff et al., 2013].

## 5 Rôle de l'impact phonologique

L'examen du graphique précédent et des deux premiers tableaux révèle cependant des disparités au sein de ce que nous appellerons le « groupe principal », et celles-ci nous conduisent à envisager d'autres hypothèses pour expliquer les différences. En effet, si le faible pourcentage d'erreur du « ss » peut s'expliquer encore par un effet inverse de la fréquence, il n'en va pas de même pour le « cc » qui a un taux d'erreur comparable, mais une fréquence bien inférieure.

Notre hypothèse est de prendre en considération l'impact phonologique du doublement de la consonne pour comprendre le phénomène ; l'impact du « ss » est évident dans la prononciation, et celui du « cc » l'est aussi quand cette consonne double précède un « e » ou

un « i ». C'est pourquoi nous avons distingué différents cas, et mesuré de manière séparée les taux d'erreur.

Nous avons fait les distinctions suivantes :

- cc : 2 cas, prononcé [k] comme dans « occuper » ou [ks] comme dans « accident »
- ff : 2 cas, précédé ou non d'un « e ».
- ll : 3 cas, précédé d'un « e », précédé d'un « i » et prononcé [j], et autres cas.
- tt : 2 cas, précédé ou non d'un « e ».

**Tableau 4** : taux d'erreur des consonnes doubles, avec distinctions

motif	% base	% corpus
eff	1,38%	0,046%
cc [ks]	0,72%	0,052%
ill [j]	0,87%	0,104%
ell	3,91%	0,112%
ss	1,77%	0,146%
ett	2,52%	0,162%
ff	2,67%	0,197%
cc [k]	3,91%	0,234%
mm	5,55%	0,247%
pp	4,52%	0,269%
nn	6,52%	0,270%
ll	6,20%	0,383%
tt	5,06%	0,444%
rr	6,61%	0,536%

Nous avons volontairement séparé les deux groupes qui émergent de ces résultats, triés suivant le taux d'erreur dans le corpus Wacky. Précisons, pour être clair, que dans ce tableau, le motif « ff » correspond au cas où cette consonne double n'est pas précédée du « e », et qu'il en est de même pour les autres consonnes. Il apparaît de manière évidente (excepté une petite anomalie pour les données de la base du motif « ell ») que si le doublement de la consonne a un impact phonologique (que ce soit sur la consonne elle-même ou bien sur la voyelle qui la précède), la consonne ainsi redoublée résistera mieux à l'erreur orthographique. Ceci s'explique très probablement par la dominance de la phonologie dans l'apprentissage de l'écriture [Fayol et Jaffré, 2016].

Terminons ce paragraphe par deux remarques ; tout d'abord, il serait intéressant de prolonger cette partie de l'étude par l'examen de la « compensation », autrement dit de regarder dans quels cas le scripteur qui, par exemple, supprime un des « t » du motif « ett » va en même temps ajouter un accent grave au « e » afin d'être en accord avec sa prononciation.

D'autre part, il existe des cas où le doublement de la consonne a un impact sur la prononciation du « e » qui la précède, mais que nous n'avons pas détaillés dans le tableau ; ce sont les motifs « emm », « enn » et « err ». Ces cas sont des contre-exemples, en ce sens

que cette fois, la consonne double précédée d'un « e » ne résiste pas mieux à l'erreur ; c'est pourquoi nous allons les examiner en détail dans le paragraphe qui suit, afin d'expliquer ce phénomène.

## 6 Cas où l'impact phonologique n'a pas d'influence sur le taux d'erreur

L'explication est relativement simple et fait appel à ce que nous avons pris en compte au début de cette étude, à savoir la fréquence du motif. Mais ici, le pas en avant consiste à considérer le motif, par exemple « emm » par rapport à son concurrent graphique « em + voyelle ». Ce qui nous amène forcément à reconsidérer les résultats concernant les motifs « eff », « ell » et « ett » dans un but de cohérence globale.

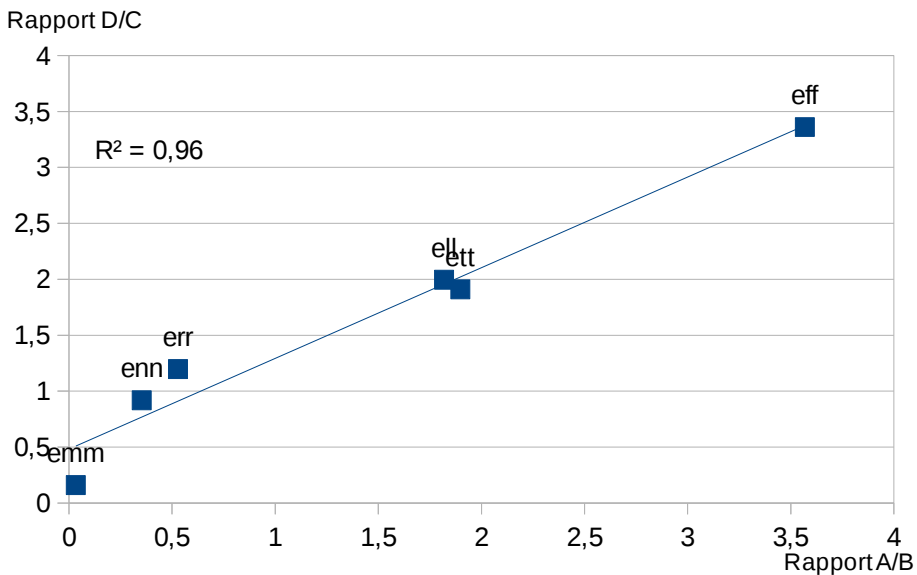
Nous avons donc d'une part relevé dans le corpus Wacky le nombre d'occurrences de chaque motif et de son concurrent, et d'autre part mesuré les taux d'erreur du motif et de la consonne double qui en fait partie. Les observations et calculs sont donnés dans le tableau ci-après :

**Tableau 5** : taux d'erreur et fréquences des motifs et de leurs concurrents

motif	A=N motif	B=N motif concur.	A/B	C = % motif	D = % consonne	D/C
eff	1683320	471679	3,569	0,046%	0,153%	3,361
ell	8906238	4896118	1,819	0,112%	0,223%	1,996
ett	5223949	2753073	1,897	0,162%	0,309%	1,911
err	2283058	4317356	0,529	0,448%	0,536%	1,197
enn	1880353	5327798	0,353	0,293%	0,270%	0,919
emm	636439	19025738	0,033	1,538%	0,247%	0,160

Ce que l'on remarque d'abord dans ce tableau, c'est que plus le motif est fréquent (A) par rapport à son concurrent (B), plus son taux d'erreur (C) est faible. Mais si l'on veut aller plus loin et quantifier les résultats, on peut mettre en relation ce taux d'erreur (C) avec celui de la consonne double (D) quelle que soit la voyelle qui la précède. Il apparaît alors une très forte corrélation entre les rapports A/B et D/C, comme le montre la figure 2 ci-après.

Certes, ce résultat assez surprenant dépend peut-être du corpus considéré (ici le corpus Wacky), mais dans la mesure où ce corpus est un reflet assez fidèle de ce qui tombe sous nos yeux, il est nécessaire de le mentionner. Enfin, pour être complet, remarquons que cet effet du « motif concurrent » ne se manifeste vraiment que dans le cas du « e » précédant le « mm », et que cette concurrence est spécialement due aux finales en « emment », opposées à celles en « ement » qui sont plus nombreuses.

**Figure 2** : corrélation entre les rapports de fréquences et les rapports de taux d'erreur

En conclusion de cette première partie, nous pouvons affirmer que la fréquence d'un motif à consonne double joue un rôle prépondérant dans la sensibilité de ce motif à l'erreur orthographique, et en outre que son effet est logarithmique. Pour des motifs dont la fréquence est du même ordre de grandeur, c'est l'impact phonologique du motif qui le rend plus solide, quand toutefois l'effet de la fréquence du motif concurrent à une seule consonne ne se manifeste pas.

## 7 Rôle de la difficulté du contexte

Dans une étude présentée en 2009 et consacrée au motif « onn », nous avons remarqué que le taux d'erreur (en omission) sur ce motif augmentait de manière significative lorsqu'il faisait partie d'un mot contenant une autre consonne double [Manguin, op. cit.]. Nous en avons déduit que la surcharge cognitive qui résulte de la difficulté du mot entraînait cette augmentation des erreurs sur le motif « onn », auxquelles venaient s'ajouter les erreurs sur la seconde consonne double, bien entendu. Parmi les perspectives de cette étude, nous avons prévu d'élargir si possible ce résultat pour le généraliser à toutes les consonnes doubles, toujours pour les erreurs d'omission.

Nous avons donc mesuré les taux d'erreur dans les mots à deux consonnes doubles suivant le même principe que dans la première partie ; cependant il n'a pas été possible de comparer ces taux à tous ceux des mots à une consonne double, pour la simple et bonne raison que certains motifs sont introuvables dans les mots à deux consonnes doubles, comme par exemple « bb », « gg » et « zz ». Nous n'avons pas non plus mentionné ci-après le motif « dd » pour lequel nous n'avons que trop peu de données (4 mots).

Le tableau ci-après donne les résultats des mesures ; chaque motif est détaillé dans deux lignes, la première pour les mots à une consonne double, et la seconde pour les mots à deux



consonnes doubles. Pour tous les motifs sauf un, les taux de la deuxième ligne sont nettement supérieurs à ceux de la première ligne (en comparant colonne à colonne), ce qui confirme notre résultat de 2009.

Le seul motif pour lequel cette remarque générale ne se vérifie pas est le « cc » prononcé [k], mais on voit clairement que l'écart entre les deux lignes est très faible. Il suffit en effet d'un mot (sur les 220 étudiés) pour faire basculer le résultat dans l'autre sens : ainsi par exemple, la suppression des données concernant le mot « impeccable » (qui représente à lui seul 6 % des erreurs) conduit à des taux moyens de 3,75 % (pour la base) et 0,216 % (pour le corpus) ; il est donc clair que cette légère inversion résulte simplement d'une particularité, qui est la suivante : le « cc » est toujours après la première voyelle du mot dans les mots à deux consonnes doubles, alors que ce n'est pas le cas pour les mots à une consonne double « cc » où l'on rencontre des exemples comme « désaccord », « impeccable » ou « inaccoutumé ».

Cette explication se base sur les constatations faites par Wing et Baddeley [Wing et Baddeley, 2009] sur les mots anglais, et la prise en compte de la position du motif dans le mot nous conduit naturellement au paragraphe suivant.

**Tableau 6** : taux d'erreur dans les mots à une ou deux consonnes doubles

<b>motif</b>	<b>% base</b>	<b>% corpus</b>
cc [k]	3,91%	0,234%
cc [k] 2c	3,87%	0,221%
cc [ks]	0,72%	0,052%
cc [ks] 2c	2,43%	0,165%
ell	3,91%	0,112%
ell 2c	9,42%	0,438%
ett	2,52%	0,162%
ett 2c	13,59%	1,33%
ff	2,67%	0,197%
ff 2c	14,24%	0,658%
ill [j]	0,87%	0,104%
ill [j] 2c	2,18%	0,186%
ll	6,20%	0,383%
ll 2c	14,95%	4,22%
mm	5,55%	0,247%
mm 2c	40,56%	2,84%
nn	6,52%	0,270%
nn 2c	14,12%	1,10%
pp	4,52%	0,269%
pp 2c	7,99%	1,28%
rr	6,61%	0,536%
rr 2c	47,42%	6,34%
ss	1,77%	0,146%
ss 2c	3,28%	0,255%
tt	5,06%	0,444%
tt 2c	23,22%	11,46%

## 8 Rôle de la position

Dans l'étude de 2009 déjà mentionnée, nous avons prolongé la réflexion en examinant les taux d'erreur en fonction de la position du motif étudié « onn » par rapport à l'autre consonne double présente dans le mot. Et nous avons observé que lorsque ce motif était en seconde position, son taux d'erreur était significativement plus élevé. Il s'agit maintenant de tenter de généraliser ce résultat à tous les motifs à consonne double dont nous disposons dans le corpus.

Précisons en premier lieu qu'il n'a pas été possible de faire la comparaison pour tous les motifs présents dans les mots à deux consonnes doubles, car certains ne se trouvent jamais en seconde position : ce sont « pp » et « cc » prononcé [ks].

Le résultat de nos mesures ne nous permet pas de généraliser complètement le fait que la seconde consonne double présente un taux d'erreur supérieure à la première, dans les mots à deux consonnes doubles. Nous allons voir en effet, que pour certaines consonnes, d'autres paramètres entrent en jeu. Néanmoins, la plus forte sensibilité à l'erreur d'omission se retrouve pour les groupes présents dans le tableau suivant, savoir « cc » prononcé [k], « ff » (non précédé d'un « e »), « mm », « nn » et « rr ». On remarquera la différence élevée entre la première et seconde position, ce qui montre que sur cette seconde consonne réside une difficulté orthographique dont l'importance n'a peut-être pas été prise en compte jusqu'ici.

Signalons enfin le cas particulier du motif « ll » dans les cas qui ne relèvent ni du motif « ell » ni du « ill » prononcé [j] : dans les mots à deux consonnes, il ne se rencontre qu'en première position, et se trouve par conséquent absent du tableau suivant.

**Tableau 7** : taux d'erreur suivant la position dans les mots à deux consonnes doubles (le chiffre indique la position de la consonne double)

Motif	% base	% corpus
cc [k] 1	3,82%	0,174%
cc [k] 2	8,73%	12,930%
ff 1	4,33%	0,164%
ff 2	71,74%	15,674%
mm 1	15,36%	0,246%
mm 2	50,58%	8,809%
nn 1	7,84%	0,626%
nn 2	19,39%	1,824%
rr 1	36,69%	4,232%
rr 2	70,37%	16,504%
tt 1	13,71%	7,515%
tt 2	67,97%	35,94%

## 9 Cas où la position joue un rôle secondaire

Avant de passer en revue les motifs pour lesquels la seconde position est moins sensible à l'erreur que la première, signalons d'abord le cas du groupe « ss » pour lequel il n'y a (quasiment) aucune différence dans les taux d'erreur observés :

**Tableau 8** : taux d'erreur du motif « ss » dans les mots à deux consonnes doubles

Motif	% base	% corpus
ss 1	3,28%	0,229%
ss 2	3,29%	0,269%

Ce phénomène s'explique par plusieurs raisons qui agissent dans le même sens ; tout d'abord, le motif « ss » est comme nous l'avons vu, le plus courant parmi les consonnes doubles. Ensuite, son impact phonologique est renforcé car il est « intrinsèque » : le doublement de la consonne modifie la prononciation du groupe consonantique et non celle de la voyelle qui précède. Le motif « ss » est par conséquent un point de faible difficulté, peu sensible aux erreurs.

Pouvons-nous extrapoler ce résultat aux autres motifs à consonne double qui sont peu sensibles aux erreurs, comme le « ill », « ell » ou « ett » ? Ces motifs n'ont pas d'impact intrinsèque, néanmoins nous avons vu que leur impact phonologique et leur fréquence permettaient de les rendre plus résistants aux erreurs. Voyons d'abord les motifs en « ell » et « ill » :

**Tableau 9** : taux d'erreur des motifs « ell » et « ill » dans les mots à deux consonnes doubles

Motif	% base	% corpus
ell 1	12,28%	0,330%
ell 2	8,22%	0,412%
ill [j] 1	3,06%	0,364%
ill [j] 2	1,46%	0,106%

En apparence, ces observations sont en désaccord avec celles faites pour autres motifs, pour lesquels le taux d'erreur est plus élevé quand le motif est en seconde position. En réalité, ces résultats sont en trompe-l'œil, et plutôt en accord avec ceux du motif « ss ».

Prenons tout d'abord le cas du « ill » ; une plongée dans le détail des erreurs observées nous montre un biais « lexical » : en effet, sur les 34 mots analysés dans cette série, « gentillesse » représente 71 % des erreurs dans le corpus de la base et 56 % dans le corpus Wacky, et les taux calculés sans tenir compte de ce mot ne sont plus les mêmes : 1,38 % (base) et 0,177 % (Wacky) pour le motif « ill » [j] en première position. Ainsi, on se retrouve avec des taux d'erreurs comparables en position 1 et 2, ce qui nous replace dans le cas du motif « ss ».

Reste à expliquer pourquoi le mot « gentillesse », qui ne représente que 10 % des occurrences de la série des 34 mots, pèse à lui tout seul 56 % des erreurs ? Son taux d'erreur propre est 1,94 %, autrement dit 10 fois plus élevé que la moyenne des 33 autres mots. L'explication est ici une convergence de deux phénomènes : d'abord la position médiane du « ill » opposée à la position finale (au sens de dernière consonne) du « ss », car l'on sait depuis Wing et Baddeley [Wing et Baddeley, 2009] que les milieux de mots sont plus sensibles aux erreurs que les fins de mots, et ensuite la présence du « ss » dans un morphème final typiquement féminin « esse » ce qui le rend encore plus résistant aux erreurs, comme l'ont montré Deacon et Bryant [Deacon et Bryant, 2006].

Il nous reste enfin à examiner en détail le cas du « ett », dont les résultats sont particuliers puisque nous n'avons pas rencontré de mots à deux consonnes dans lesquels le motif « ett » serait en première position ; néanmoins, il est intéressant de séparer le cas du « ette » final des autres cas (comme par exemple « commettre ») :

**Tableau 10** : taux d'erreur du motif « ett »

Motif	% base	% corpus
t 2 ett	27,45%	3,53%
t 2 fin ette	2,70%	0,53%

On s'aperçoit ici que la finale en « ette », caractéristique des noms (ou adjectifs) féminins est beaucoup plus résistante aux erreurs, comme c'était le cas dans l'exemple de « gentillesse » ; nous avons donc une nouvelle confirmation des observations faites par Deacon et Bryant [Deacon et Bryant, 2006].

Le cas du motif « ell » est assez particulier, c'est pourquoi nous lui avons consacré le paragraphe qui suit.

## 10 Remarques sur la présence du motif dans un morphème grammatical

La présence (ou pas) du motif final dans un morphème grammatical mérite d'être détaillée pour le motif « ell ». Nous distinguons d'abord deux cas : soit le motif « ell » n'est pas dans la dernière syllabe (ex. « exceptionnellement », « appellation »), soit il en fait partie, et dans ce cas il peut être inclus dans un morphème grammatical, en l'occurrence dans un adjectif féminin (ex. « exceptionnelle »). Les taux d'erreur sont donnés ci-après :

**Tableau 11** : taux d'erreur du motif « ell »

Motif	% base	% corpus
ell 2	11,19%	0,850%
elle 2 final	4,99%	0,295%

On voit ainsi que si le motif n'est pas final, son taux d'erreur est au mieux comparable à celui qui lui est associé quand il est en première position parmi les consonnes doubles du mot. Par contre, la position finale rend le motif plus résistant à l'erreur, et nous pouvons aller plus loin en remarquant que ceci est surtout vrai lorsque le doublement du « l » résulte de l'ajout du morphème qui indique le féminin de l'adjectif (toujours dans les mots à deux consonnes doubles) :

**Tableau 12** : taux d'erreur du motif final « elle »

Motif	% base	% corpus
« elle » fin ADJ	1,26%	0,024%
« elle » fin NON ADJ	16,12%	0,82%

Les exemples non adjectivaux viennent de noms (« aisselle », « cannelle » et vaisselle ») et des verbes « appeler » et « rappeler » ; pour ces derniers les taux d'erreur sont de 27,6 % (base) et 0,89 % (corpus) , et l'on peut donc s'interroger sur la décision prise dans le cadre de la dernière réforme de l'orthographe consistant à modifier les formes conjuguées des verbes en « eler », mais en conservant l'exception du verbe « appeler » et de ses composés, sous prétexte que leur orthographe était bien ancrée dans l'usage. Nous ne sommes pas ici en mesure de préciser si cet ancrage n'était en fait pas si bien ancré que cela, ou si la réforme a semé la confusion en introduisant une exception qui complique les règles ; en revanche, nous pouvons dire avec certitude que l'orthographe de l'adjectif féminin en « elle » est très bien ancrée dans l'usage, ceci parce que l'ajout du morphème flexionnel féminin est intrinsèquement lié au doublement de la consonne.

Avant de conclure, revenons un peu en arrière et signalons que ces observations sur le motif « ell » dans les mots à deux consonnes doubles sont confirmées dans les mots à une consonne double dont nous avons parlé en première partie. En effet, le motif en question montre des taux d'erreur bien différents suivant qu'il est ou non en position finale et suivi d'un « e » :

**Tableau 13** : taux d'erreur du motif « ell » dans les mots à une consonne double

Motif	% base	% corpus
« ell » non final	5,53%	0,39%
« elle » final	1,64%	0,041%

Et comme pour les mots à deux consonnes doubles, nous observons également une nette baisse du taux d'erreur lorsque le motif « elle » final se trouve dans un adjectif féminin :

**Tableau 14** : taux d'erreur du motif final « elle » dans les mots à une consonne double

Motif	% base	% corpus
« elle » final ADJ	0,94%	0,026%
« elle » final NON ADJ	2,71%	0,054%

## 11 Conclusion

Au cours de cette étude, nous avons mis en évidence le fait que la résistance à l'erreur orthographique d'un motif à consonne double dépend de plusieurs facteurs qui peuvent agir en s'opposant les uns aux autres. Par exemple, nous avons mentionné le cas des fins en « emment » qui, malgré l'effet positif de l'impact phonologique du « mm », subissent l'effet négatif de la fréquence du motif concurrent des fins de mots en « ement ».

Pour résumer, voici les paramètres ayant un effet positif :

- fréquence du motif
- impact phonologique
- situation dans un morphème grammatical

A l'inverse, les paramètres ayant un effet négatif :

- difficulté du mot

- fréquence du motif concurrent

Nous avons montré aussi que la position du motif n'est pas un paramètre déterminant, car son influence est très vite dépassée par d'autres causes ayant un effet plus prononcé, comme par exemple l'inclusion dans un morphème grammatical. Pour quantifier ces effets, il serait intéressant de détailler des exemples assez nombreux où les différentes causes pourraient être clairement séparées. Dans l'état actuel de notre travail, cela reste encore dans le domaine des perspectives. Cependant, ce qui est exposé ici permet de comprendre pourquoi une telle inégalité face aux erreurs orthographiques existe pour les motifs à consonnes doubles.

## Bibliographie

- Baranes, M. (2015). *Normalisation orthographique de corpus bruités*. Thèse de l'Université Paris Diderot.
- Baroni M., Bernardini S., Ferraresi A., Zanchetta E. (2009). The WaCky Wide Web : A collection of very large linguistically processed Web-crawled corpora. *Language Resources and Evaluation* 43(3), pp. 209-226.
- Bonin, P. (2007). *Psychologie du langage*. Louvain, De Boeck.
- Borchardt, G. (2012). *L'influence des connaissances graphotactiques sur l'acquisition de l'orthographe lexicale : étude chez l'enfant d'école élémentaire et chez l'adulte*. Thèse de l'Université Paris Descartes.
- Crisco (1998). *Dictionnaire électronique des synonymes*. <http://www.crisco.unicaen.fr/> .
- Deacon S. H., et Bryant P. (2006). This turnip's not for turning : children's morphological awareness and their use of root morphemes in spelling. *British Journal of Developmental Psychology*, Vol. 24, pp. 567-575.
- Fayol, M. et Jaffré J.-P. (2016). L'orthographe : des systèmes aux usages. *Pratiques*, 169-170.
- Hazard, M-C (2009). *Consistance orthographique et construction du lexique chez l'enfant d'âge scolaire*. Thèse de l'Université de Nice Sophia Antipolis.
- Le Goff K., Pacton S., Manguin, J.-L. & Rey, A. (2013). The interference of misspellings on spelling performance. *poster for the 25th Association for a Psychological Science (A.P.S.) Convention*. Washington D.C.
- Lucci V., et Millet, A. (1994). *L'orthographe de tous les jours : enquête sur les pratiques orthographiques des français*. Paris, Honoré Champion.
- Manguin, J.-L. (2009). Les requêtes sur un site Web : un corpus pour étudier la variation orthographique. *Journées Internationales de Linguistique de Corpus*. Lorient.
- Manguin, J.-L. (2016). « Ortholexies », une base de données publique pour l'orthographe lexicale. *5<sup>ème</sup> Congrès Mondial de Linguistique Française*. Tours.
- Pacton, S., Fayol, M. & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities in French. *Child Development* 76, p. 324-339.
- Pacton, S., Perruchet, P., Fayol, M. & Cleeremans, A. (2001). Implicit learning out of the lab.: The case of orthographic regularities. *Journal of Experimental Psychology: General* 130, p. 401-426
- Wing, A., et Baddeley, A. (2009). Righting errors in writing errors: The Wing and Baddeley (1980) spelling error corpus revisited, *Cognitive neuropsychology*, vol. 26, 2, pp. 223-226.