HAL
open science

Homophonie et maîtrise du français écrit : apport de l'apprentissage sur corpus

Maï Leray, Henry Tyne

To cite this version:

Maï Leray, Henry Tyne. Homophonie et maîtrise du français écrit : apport de l'apprentissage sur corpus. Linguistik Online, 2016, Corpus, grammaire et français langue étrangère : une concordance nécessaire, 78 (4), 10.13092/lo.78.2954 . hal-01401956

HAL Id: hal-01401956

https://hal.science/hal-01401956v1

Submitted on 24 Nov 2016

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.

Homophonie et maîtrise du français écrit : apport de l'apprentissage sur corpus

Maï Leray/Henry Tyne (Perpignan)

Abstract

Mastery of spelling in L1 French is notoriously complex, in particular due to morphographic difficulties and grammatical homophony. This study investigates the use of data-driven learning in French primary schools for the homophonous set /sE/ (*c'est*, *s'est*, *ces*, *ses*, *sais/t*). The results show that whereas initially there is no difference in comparison to traditional teaching (i.e. following pre-test, fewer errors are made in general at the time of learning by both experimental and control groups), there is less attrition for the data-driven learning group, with some spellings progressing even at the post-test stage. We suggest that data-driven learning may be used for developing spelling in L1 French as learners are exposed to enriched input. The question of how a predominantly L2-based method can be applied to L1 pedagogy is raised.

1 Introduction

Dans cet article nous présentons une étude explorant la démarche d'apprentissage sur corpus (connue en anglais sous le nom de « *data-driven learning* ») dans le cadre de la maîtrise de la langue écrite en langue première ou maternelle (L1). Cette démarche, qui a été largement développée pour la didactique des langues secondes ou étrangères (L2) (Boulton/Tyne 2014), est encore peu présente dans la didactique des L1 mais semble offrir des perspectives intéressantes notamment en ce qui concerne le problème de l'homophonie comme nous le suggérons ici. Après avoir décrit la problématique au niveau de l'acquisition de formes homophoniques en général, nous nous pencherons sur la question de l'exposition aux formes dans une perspective d'apprentissage de l'orthographe. Nous présenterons et commenterons ensuite les résultats de l'étude portant sur l'application de cette démarche à l'enseignement/acquisition de la série en /sE/ à l'école en France.

2 Homophonie et maîtrise de la langue écrite en L1

« Je ne veut pas ecrire parce que je deteste les dictee. j'ai jamai us audessue de 5 et si c'est pour faire autan derreur et que des scientifique le voie je prefére encore pas ecrire ». Cette explication, trouvée au dos d'une des copies (blanches) de notre enquête, témoigne des complexes que peut susciter le fait d'écrire lorsque celui qui écrit sait (ou pense savoir) qu'il sera jugé sur son orthographe. Cette justification découle visiblement d'une « peur de se tromper » que Brissaud (2014) impute à « l'opacité » du code écrit de la langue française dont

CC by 3.0

Catach (1973, 1988, 1995), et Fayol et Jaffré (2008) ont souligné le particularisme par rapport à d'autres langues, complétant ainsi les descriptions de Chervel et Blanche-Benveniste (1969).

La liste des « difficultés » liées à la maîtrise écrite de la langue française, notamment en raison de son évolution particulière (Catach 2001 ; Legros/Moreau 2012), est longue et a donné lieu à de nombreuses propositions de réformes (voir entre autres Chervel 1991 ; Dister et al. 2009) faisant ainsi de l'orthographe un élément plus qu'emblématique de l'enseignement du français comme L1. Si les difficultés sont de différents ordres, la question de l'homophonie[1] est, elle, souvent désignée comme étant un problème majeur. Dans leur enquête auprès d'enseignants et futurs enseignants du français, Mout et Vernel (2012) donnent une liste des difficultés orthographiques les plus problématiques (voir aussi Groupe RO 2012) : toutes les difficultés citées concernent l'homophonie. Plus précisément, à l'école en France, des paires (ou séries) d'homophones comme *a*/*à*, *son*/*sont*, *est*/*et*/*ai* sont censées être maîtrisées « sans erreur » à l'issue de la première année du cours moyen (CM1, http://eduscol.education.fr/, [20.05.2015]). Mais comme en témoignent de nombreux spécialistes, professionnels, parents, etc., l'homophonie pose encore problème à l'issue du parcours élémentaire et continue de se manifester tout au long du secondaire, voire même chez des scripteurs devenus adultes, comme le laissent constater les nombreuses réalisations erronées que l'on trouve sur des sites ou des forums en ligne.

Si, pour l'apprenant étranger, ces variantes orthographiques de suites sonores identiques constituent autant de formes utiles qui lui permettent de voir la grammaire et ses sonorités (par ex. pour le FLE, certains exercices de repérage dans Lodge et al. 1997 ; voir aussi Léon et al. 2009), pour l'élève ayant comme L1 le français, l'appréhension de ces variantes présente une double difficulté. En effet, non seulement le natif découvre la diversité écrite d'une langue déjà maîtrisée à l'oral, mais il se retrouve aussi confronté à une certaine mise à plat de différents phénomènes autour d'une même problématique : par exemple, des formes verbales (*mets*, *met* du verbe *mettre* ; ou *m'est*, *m'ait*…) sont traitées avec des conjonctions (*mais*), et autres formes possessives (*mes*). Sont alors établies des correspondances entre formes qui, sur le plan grammatical, sont très différentes. La série en /sE/ (*ses*, *ces*, *c'est*, *s'est*…), qui illustre bien ce phénomène, est donnée comme exemple par les enseignants lorsqu'on les interroge sur la nature des erreurs récurrentes de leurs élèves[2] (voir aussi Bentolila 2010). C'est cette série qui sert de support à la présente étude.

3 Corpus et apprentissage

Ce que l'on nomme en anglais « *data-driven learning* » (Johns/King 1991), et que nous traduisons ici par « apprentissage sur corpus » (ASC – voir Tyne 2013 ; Boulton/Tyne 2014) décrit avant tout une méthode d'apprentissage par la découverte de données extraites de corpus[3]. Cette approche, qui revient sur l'opposition entre lexique et grammaire, faisant une

[1] Actuellement la majorité des travaux portant sur l'homophonie traitent de l'homophonie verbale, et plus particulièrement du cas des finales en /E/ (par ex. voir David et al. 2006 ; Fayol/Pacton 2006).

[2] Cette question a été posée aux enseignants dans le cadre notre pré-enquête.

[3] Le mot « corpus » est potentiellement problématique dans la mesure où il peut renvoyer à des définitions différentes (cf. Boulton et al. 2015). En didactique le terme « corpus » peut désigner une collection de

place importante aux associations plus ou moins préférentielles ou figées dans la langue, trouve un appui théorique dans le connexionnisme (Ellis 1998) où l'on aborde l'acquisition par la détection de relations saillantes dans l'input. La pertinence d'une telle approche en L2 a été établie par Lewis (1993, 1997a, 1997b – voir Kerr 2013 et Di Vito 2015 pour une application au FLE) dans ce qu'il appelle l'« approche lexicale », où il est question de promouvoir le repérage et l'étude de différents schémas et associations récurrents.

Si le recours à l'ASC paraît encore quasi inexistant[4] dans le domaine du français L1 (comme pour d'autres langues – voir cependant pour l'anglais le travail de Sealey et al. 2004 ; Sealey 2011), il semblerait néanmoins que la progression actuelle des recherches en didactique de la L1 et de l'orthographe attache une importance de plus en plus considérable aux usages attestés et à certaines données quantitatives issues de leur observation. C'est du moins ce que nous en concluons d'après l'état de la question dressé par Brissaud (2011) qui, tout en mentionnant un recours grandissant au corpus Manulex (voir aussi Lété 2006 ; Lété et al. 2004), souligne l'importance des travaux de la psycholinguistique qui « se poursuivent et s'attachent à montrer l'impact de la fréquence des unités discursives et des régularités du système sur la performance orthographique » (2011: 211).

Si l'on admet que l'acquisition procède, entre autres, de l'identification de phénomènes réguliers et récurrents par l'exposition à la langue, au niveau de l'écrit cela signifierait que les enfants de langues et de cultures différentes auraient tendance à reproduire des phénomènes propres au système auquel ils ont été exposés par la lecture notamment (cf. Cury Pollo et al. 2009). Ainsi, par exemple, Pacton et al. (2001) montrent que les enfants acceptent des pseudo-mots plus volontiers lorsque ceux-ci respectent les règles orthographiques (connues implicitement) de leur langue. Mais on peut se demander si l'exposition aux seules formes écrites est aussi efficace pour la restitution orthographique que l'enseignement explicite de la règle (cf. Kemper et al. 2012). Les conditions qui entourent la découverte d'une forme apparaissent dès lors comme un élément crucial à prédéterminer pour établir une méthodologie propre (cf. Lewis 1997a, 1997b pour la L2).

En ce qui concerne l'effet de la fréquence, toutes les unités (phonémiques, syllabiques, n-gramiques, lexicales) ne seraient pas pertinentes au même titre pour la restitution orthographique. Dans le cas du français, l'étude de Maïonchi-Pino et al. (2010) semble indiquer que la fréquence de certains types de syllabes récurrents est un facteur particulièrement saillant lors de la reconnaissance et de la restitution écrite de différents mots. Selon Fayol et Pacton (2006), dans des cas de choix entre formes homophoniques (infinitif vs participe passé) c'est la relative fréquence d'occurrence de ces dernières dans la langue qui influe sur les réponses au test (ici le choix de l'infinitif pour les verbes dont la forme infinitive est plus courante). Bien que différentes procédures de sélection de formes soient présentes (et sans doute aussi complémentaires), la récupération en mémoire de formes sans calcul des

---

documents divers, une série de phrases ou un simple texte. Dans cet article nous nous référons à la définition communément acceptée en linguistique de corpus (cf. Gilquin/Gries 2009) comme un recueil électronique d'un grand nombre de textes (écrits comme oraux) pouvant être interrogé avec un logiciel dédié (ou concordancier).

[4] En témoigne, par exemple, le volume 15/2 de la *Revue française de linguistique appliquée* (« Apprendre à écrire : l'apport des nouvelles technologies »), où les articles traitant du recours aux corpus numériques par le scripteur ne l'envisagent que dans le cadre de l'apprentissage de la L2.

règles d'accord représente un automatisme peu coûteux en termes de traitement (cf. Ellis 1997 ; Tomasello 2003) et plus à même d'éviter la sur-généralisation des règles. Pour autant, le taux d'exposition nécessaire à une forme fréquente pour que celle-ci soit retenue n'est pas bien établi (Lété 2003: 198), et comme l'indiquent Fayol et Pacton (2006), il se peut aussi qu'il y ait des différences importantes d'un individu à l'autre. D'ailleurs, toutes les formes, tous les contextes ne sont certainement pas comparables.

La question de l'exposition aux formes écrites comme moyen d'avancer dans la maîtrise de l'orthographe semble ainsi suffisamment pertinente pour que l'on s'y intéresse par le biais de l'ASC. L'interrogation qui en découle est la suivante : en quoi l'ASC renforcerait-il les effets précédemment décrits pour la restitution exacte et/ou l'utilisation appropriée d'une graphie ? Le travail sur corpus (et a fortiori sur concordances) conduirait en L1 à privilégier avant tout la relation visuelle (le code oral étant maîtrisé) avec l'input et le développement d'automatismes au niveau de la lecture. Nombreuses sont les études ayant mis en avant le fait que l'exposition à la langue écrite (via la lecture) joue un rôle déterminant dans le développement de l'orthographe : ce sont en général les enfants qui orthographient bien qui sont aussi de bons lecteurs (Fayol et al. 2009). Concernant ensuite la relation visuelle avec l'input, l'étude de Chaves et al. (2012) montre qu'en dehors du contexte de décodage initial (découverte de nouvelles formes), la mémorisation de l'orthographe est meilleure lorsque le regard porte sur une forme entière et non découpée, mais qu'elle en pâtit si ce même regard n'est pas de qualité ou ne se fait que partiellement (lecture trop rapide, par exemple). Justement, le recours à des concordances comme input peut servir à renforcer autant l'aspect visuel (cf. l'input « enrichi » de Cobb 2014) que les mécanismes de lecture, étant donnée la présentation particulière, centrée sur la page et répétée, de la forme choisie. Enfin, si on se réfère aux contributions de l'ASC à la didactique de la L2, le fait de pouvoir mettre en avant des éléments récurrents permet de relever des informations que l'on n'aurait pas nécessairement détectées (ou en tout cas pas assez rapidement) dans l'input naturel que constitue tous les contacts ordinaires avec la langue (apprentissage de la L2) ou avec les formes écrites (lecture et orthographe en L1), faisant ainsi mieux ressortir les conditions d'utilisation de l'élément cible.

4 Etude

Ainsi que nous l'avons précisé plus haut, c'est essentiellement pour son caractère « symptomatique » que notre choix s'est porté sur la série des formes en /sE/. Nous ajoutons que ce choix se justifie d'autant plus lorsqu'on tient compte de la prononciation systématiquement fermée par les locuteurs du sud[5] alors que les manuels opposent en général *ces/ses* et *c'est/s'est* (par ex. François-Denève 2012). La présente étude porte ainsi sur six morphographies différentes (*ces*, *ses*, *c'est*, *s'est*, *sais*, *sait*), même si, comme nous le verrons, nous n'avons pas considéré *sais* et *sait* comme étant des formes en opposition (voir la note 13 ci-dessous), étant surtout intéressés par la restitution d'une des formes du verbe *savoir* (*sais* ou *sait* selon les exercices) par rapport aux autres formes dans la série. Sont exclues de la

[5] Cette étude concerne un public d'enfants du sud de la France pour qui les oppositions dites « standard » pour la voyelle /E/ en position finale ne se font pas à l'oral (*été*/*ét**ais*** = [ete]).

série des formes rares (par ex. comme dans *il faut qu'il s'ait pour lui-même*) et des formes lexicales, par ailleurs rares également (par ex. *la saie*, ou la ville de *Sées*).

4.1 Objectifs et hypothèses

S'agissant d'élèves natifs pour la totalité, l'expérimentation de l'ASC n'allait, pour ainsi dire, rien apprendre de nouveau aux élèves, la distinction sémantique des formes en /sE/ étant parfaitement opérée lors de la lecture. D'ailleurs, chacune de ces formes a déjà été étudiée en classe (« les pronoms possessifs », « les articles », « les verbes pronominaux », etc.). Notre objectif n'était donc pas de faire en sorte que les enfants associent mieux une forme à un sens donné ou à du métalangage, mais bien de les aider à faire le bon « choix » graphique au moment de retranscrire la suite /sE/. Même si ce choix devait demeurer imparfait, nous faisions l'hypothèse que des séances de redécouverte de ces formes par le biais de l'ASC entraîneraient un meilleur usage des formes en général, une diminution des formes dominantes ayant tendance à parasiter tous les contextes d'écriture, et la diminution des graphies autres que celles représentant les homophones en question (écritures phonographiques, récupération de phonèmes existants, etc.).

4.2 Méthodologie

L'ensemble de l'expérimentation s'est déroulé sur six mois (janvier-juin 2015) et a concerné un total de 169 enfants du niveau CE2 (7-8 ans) à la sixième (11-12 ans), répartis dans cinq établissements dans la zone géographique du Conflent dans le département des Pyrénées-Orientales en France. Les données pour les élèves de CE2, de CM1 (8-9 ans) et de sixième n'ont été récoltées que lors du pré-test afin de mieux situer le niveau global des élèves de CM2 (10-11 ans) qui constituaient le public cible de notre expérimentation, correspondant à la dernière année de l'école primaire.

4.2.1 Population concernée et dispositif expérimental

La population de CM2 ciblée dans cette étude représente un effectif de 50 élèves en tout : le groupe expérimental (ayant testé l'ASC) et le groupe témoin (enseignement sans corpus) comportant respectivement 25 élèves chacun. Chaque groupe se compose de deux sous-groupes de taille similaire issus d'écoles et de villes différentes ; aucun d'entre eux n'avait entamé de leçon ou de révision sur /sE/ pendant les mois ayant précédé le pré-test[6].

Dans les classes du dispositif expérimental, les enseignants[7] (qui découvraient l'ASC à l'occasion de cette étude) ont initié un apprentissage implicite des divers homophones à partir d'une série d'exercices à réaliser en autonomie et basés sur l'analyse et la description de concordances présentées sur papier (approche dite « *hands-off* », c'est-à-dire sans accès direct aux données à l'aide d'un ordinateur – Boulton 2012). A l'opposé de David et al. (2006) qui ont choisi de charger un autre enseignant des séances expérimentales afin d'éviter l'effet

[6] Par ailleurs, nous faisons abstraction des variables socioculturelles ainsi que de la répartition fille/garçon ; aucun cas de déficience intellectuelle n'a été signalé dans notre échantillon.

[7] Nous remercions tous les enseignants qui ont accepté de participer à l'étude.

« maître », nous avons préféré laisser nos supports aux enseignants habituels pour que l'effet « nouveauté » ne réside que, et uniquement, dans la méthode.

En ce qui concerne les élèves du dispositif contrôle, les enseignants ont suivi leur programme habituel[8] et n'ont fait qu'accorder leurs séances autour de la même période de réalisation des séances par l'ASC dans les groupes expérimentaux. Nos seules consignes ont été destinées aux modalités des divers tests.

4.2.2 Choix des corpus, extraction des formes et didactisation

Au risque d'avoir eu recours à des données langagières qui dépassent le quota d'exposition à la langue des enfants de dix ans (type de lexique, thématiques), nous avons opté pour un corpus « varié et équilibré » (Bilger 2008: 277) comportant différents sous-corpus écrits et oraux[9] :

- 2 corpus écrits d'un million de mots chacun (Presse et Littérature[10]) ;
- le Corpus de Référence de Français Parlé (*CRFP*) de 440 000 mots incluant 3 situations de parole (privée, professionnelle, publique).

Les analyses effectuées sur ce corpus ont été réalisées à l'aide du concordancier Antconc de Laurence Anthony (version 3.4.3, www.laurenceanthony.net/software/antconc/ [20.05.2015]). La distribution de la série d'homophones hétérographes en /sE/ dans ce corpus est présentée dans le Tableau 1.

[8] Nous comptons ici sur l'opposition simple entre approche avec corpus (telle que nous la définissons dans le cadre de l'ASC) et approche sans corpus, quelle qu'elle soit, que nous appelons « habituelle ». Des entretiens avec les enseignants ont permis de constater l'utilisation de consignes et des exercices assez « classiques » de type déductif (par ex. rappel des règles suivi textes à trous). La didactique de l'orthographe a connu des approches différentes au fil des années et les recherches continuent d'alimenter cette réflexion. Pour une présentation récente voir dans Brissaud et al. (2008) et Brissaud et Cogis (2011), mais aussi Duret et Zecca (2013) pour une étude portant précisément sur l'enseignement des homophones grammaticaux.

[9] Si de nombreuses études ayant conduit des expérimentations auprès d'enfants du primaire ont eu recours au corpus Manulex (Lété et al. 2004), un tel corpus ne pouvait correspondre à nos besoins dans la mesure où il ne permet a priori pas de produire des concordances et ne contient a priori pas d'usages « ordinaires » puisés dans le langage de tous les jours et a fortiori oral.

[10] Corpus « Presse » fait d'articles du *Monde*, de *Libération* et du *Nouvel Observateur* ; corpus « Littérature » fait d'extraits des œuvres suivantes : *Le bossu*, *Madame Bovary*, *Germinal*, *Le château des Carpathes*, *Eugénie Grandet*.

	c'est	*s'est*	*ces*	*ses*	*sait*	*sais*
Nombre total d'occurrences	6057	408	1108	1654	160	454
ECRIT	1065 (18%)	242 (60%)	829 (75%)	1565 (95%)	85 (53%)	101 (22%)
Littéraire	*394 (7%)*	*31 (8%)*	*424 (38%)*	*925 (56%)*	*28 (18%)*	*80 (18%)*
Presse	*671 (11%)*	*211 (52%)*	*405 (37%)*	*640 (39%)*	*57 (35%)*	*21 (4%)*
ORAL	4992 (82%)	166 (40%)	279 (25%)	89(5%)	75 (47%)	353 (78%)
Privé	*4063 (67%)*	*149 (36%)*	*232 (21%)*	*75 (4%)*	*58 (36%)*	*305 (67%)*
Professionnel	*929 (16%)*	*17 (4%)*	*47 (4%)*	*14 (1%)*	*17 (11%)*	*48 (11%)*

Tableau 1 : Distribution des /sE/ dans les corpus écrits et oraux

Des extraits des corpus mettant en exergue les différents usages des diverses graphies de /sE/ ont ensuite été sélectionnés afin de permettre à l'enfant de mieux cerner les informations pertinentes (éléments contextuels récurrents, tendances) dans l'input sous forme de concordance. Cette démarche en entonnoir, certes coûteuse en termes de préparation et d'analyse préalables des formes, résout en même temps de nombreux conflits concernant le choix en amont des occurrences à traiter (cf. Kerr 2013). Parfois de légères modifications ont été effectuées dans les données orales (en retirant par exemple certains symboles et autres bribes dans une retranscription) afin de rendre les extraits plus « accessibles » aux lecteurs (cf. Blanche-Benveniste 2010). Nous rappelons en effet qu'une appréhension trop rigide de l'authentique (cf. Duda et al. 2009) peut parfois comporter des risques en nous empêchant de conserver un certain bon sens, en l'occurrence le fait que des élèves de CM2 ne sont guère familiers des modalités de la transcription ou que du lexique (vraiment) trop rare détournerait leur attention sur un autre élément que la notion visée et compliquerait l'accès à la compréhension.

La mise en valeur de différents aspects (souvent variables d'un corpus à l'autre pour une même forme)[11] a ainsi nécessité pour chaque forme de /sE/ la création d'une à trois séries de concordances comportant entre 10 et 20 lignes chacune ; chaque forme faisant ensuite l'objet d'une séquence didactique indépendante.

Les diverses tâches (voir un exemple en Annexe) comportaient essentiellement des exercices de repérage visuel directement sur les concordances ou nécessitant une réponse écrite, des exercices d'extractions de constructions ou de collocations à réemployer, des transformations de séquences présentes dans les concordances, des exercices de reconstitution partielle de la règle, des exercices portant sur l'interprétation grammaticale et sémantique. Toutes les requêtes (notamment de repérage ou de tri des formes) ont été guidées par des mises en relief en gras ou en couleurs des éléments cibles dans les concordances. Ce guidage a ainsi permis

[11] Par exemple, les occurrences de *ces* dans le corpus Oral font essentiellement apparaître des constructions comme « *ces [NOM PL]-là* ») alors que le corpus Presse (qui contient beaucoup séquences telles que *ces derniers jours/temps/mois* ou encore *ces dernières années/semaines*) amène à faire mention de l'aspect temporel des constructions en *ces* dans ce type d'écrit.

de proposer aux élèves un travail similaire à celui du chercheur, les exercices et tâches d'observation remplaçant certaines fonctionnalités des logiciels ou des moteurs de recherche.

4.2.3 Progression des séquences et modalités de l'évaluation

La mise en place de l'expérimentation a suivi la chronologie suivante :

- fin février : participation de tous les niveaux au pré-test ;
- début mars : introduction de l'ASC auprès du groupe expérimental de CM2 à raison d'une séance d'environ 45 minutes par semaine ;
- fin avril : exercice de contrôle pour les groupes expérimental et témoin de CM2 ;
- fin mai : post-test pour les groupes expérimental et témoin de CM2.

Les pré- et post-tests ont été proposés sous la forme de courtes dictées illustrées. L'exercice de contrôle, quant à lui, proposait un modèle écrit des différentes graphies pour compléter les bulles d'une bande dessinée.

4.3 Résultats et analyses

Dans cette partie, nous présenterons et commenterons d'abord les résultats au pré-test des 161 élèves[12] âgés de 9 à 11 ans (38 élèves en CM1, 50 en CM2 et 73 en sixième), qui nous ont permis non seulement de mieux situer le niveau de notre groupe d'étude (CM2) mais aussi d'établir certaines tendances quant à la maîtrise écrite et au taux d'utilisation de diverses morphographies en fonction des différents niveaux. Nous analyserons ensuite les résultats témoignant de la progression des deux groupes impliqués dans l'expérimentation (le groupe ayant testé l'ASC et le groupe contrôle) avant de nous intéresser, pour finir, à l'évolution de la maîtrise de chaque homophone.

4.3.1 Résultats globaux : résultats au pré-test du CM1 à la sixième

La dictée du pré-test contenait deux occurrences respectives des formes *c'est*, *s'est*, *ses*, *ces* et *sais*/*sait*[13]. Le Graphique 1 fait apparaître la moyenne sur 10 obtenue par l'ensemble des enfants, mais aussi le nombre moyen de formes différentes présentes dans les copies (dans ce cas *sais* et *sait* ont valeur de 1 – voir la note 13).

---

[12] Ce nombre correspond à 169 (nombre initial de participants) moins 8 (nombre d'élèves de CE2, traités à part dans cette étude étant donné le faible effectif).

[13] Afin de ne pas alourdir la tâche pour les enfants, il a été décidé de traiter le singulier du présent du verbe *savoir* comme une seule forme, c'est-à-dire que nous n'avons pas mis en opposition *sais* et *sait* avec toutes les réalisations possibles (*je*, *tu*, *il*, *elle*, *on*, sujet lexical ou autre) dans une même dictée lors des premiers tests (ce qui permettait également de limiter le nombre total d'occurrences à 10 tout en présentant deux utilisations identiques de chaque forme). Dans tous les cas, comme pour la prise en compte des autres formes, il s'agissait de deux occurrences de /sE/ (du verbe *savoir*) pour lesquelles il y avait deux bonnes réponses possibles (*sais* ou *sait* selon l'exercice).

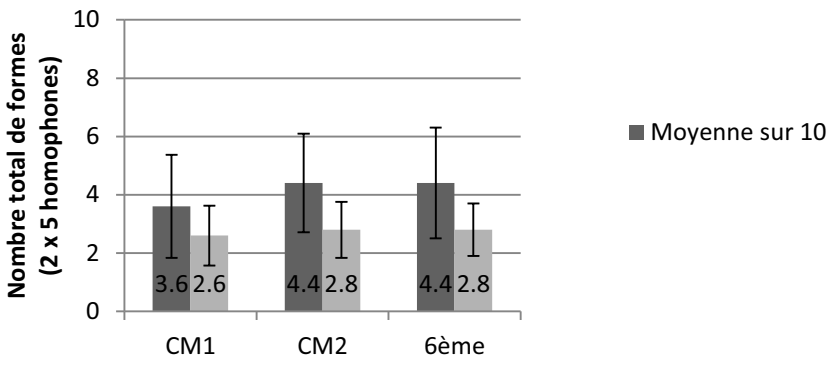


Figure 1 : Nombre moyen de formes correctement orthographiées au pré-test et nombre moyen de formes utilisées

Si ces résultats montrent une légère progression entre le CM1 et le CM2, les moyennes identiques obtenues par les enfants de CM2 et de sixième semblent indiquer que le CM2 est loin de constituer une étape charnière (telle que supposée par les programmes) dans la maîtrise des différents homophones avant l'entrée en sixième. Tout au plus l'écart-type légèrement plus important en sixième (2,1 contre 1,9) reflète-t-il les disparités au sein d'une population plus élevée (73 contre 50). Il est également intéressant de noter que le nombre de graphies différentes de /sE/ auxquelles ont recours les enfants ne varie pratiquement pas tout au long de ces trois années. Le Graphique 2, qui détaille le pourcentage de réussite des enfants en fonction des différentes formes, permet de mieux voir à quoi correspondent ces moyennes.

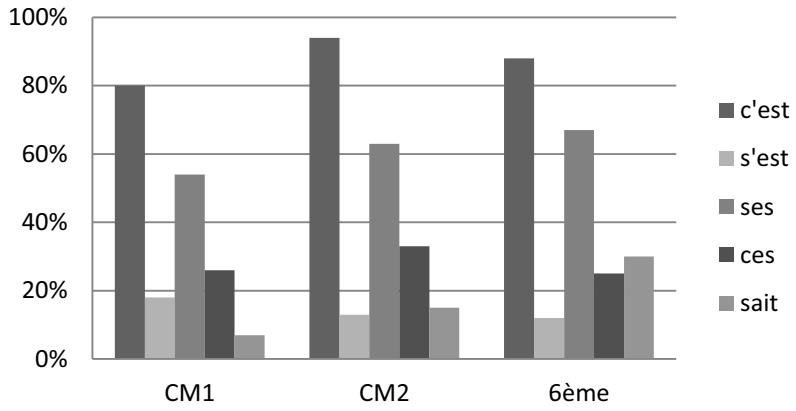


Figure 2 : Taux de réussite de placement des différents homophones

Ces données montrent que les formes *c'est* et *ses* sont celles qui, de prime abord, sont le mieux connues quel que soit le niveau scolaire. On constate ensuite que le bond dans l'emploi de la graphie *sait* qui arrive en sixième se fait au détriment des autres formes, mais surtout que les formes *sait*, *ces* et *s'est* sont en compétition plus qu'elles ne s'équilibrent au fil du temps, faisant ainsi de la moyenne 2,8 un reflet potentiel de la masse critique du nombre de formes que les enfants de cet âge-là sont capables de gérer. Par ailleurs, on peut s'étonner du fait que les CM1 aient mieux réussi à placer *s'est* que les deux autres niveaux.

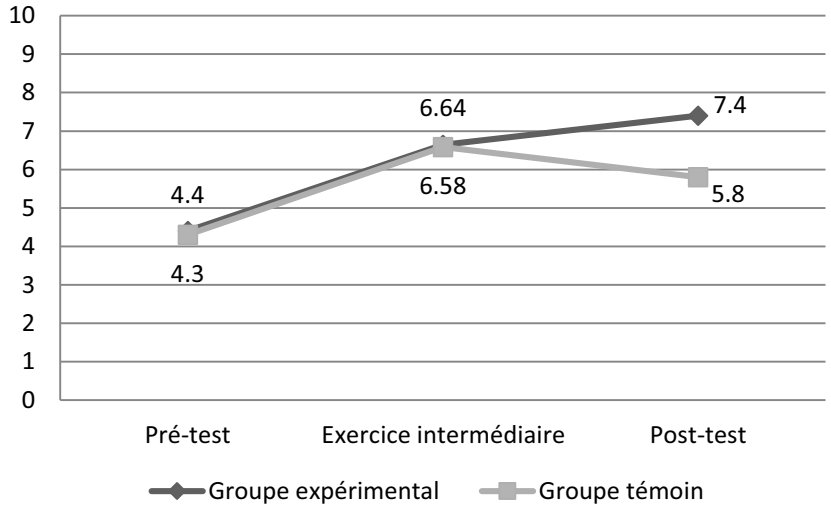
Il est à noter toutefois que les formes *c'est* et *ses* ont représenté à elles seules près de 70% des écritures de /sE/ par les enfants. Autrement dit, il est possible que de nombreuses réponses « justes » aux emplacements de *c'est* et de *ses* (voir le Graphique 2) l'aient été par défaut puisque les élèves utilisent plus volontiers ces formes en général. Par ailleurs, les graphies qui

diffèrent de celles des homophones retenus (et qui ne figurent pas dans le Graphique 2) constituent une part non négligeable des réponses puisqu'elle dépasse respectivement la part des formes *s'est* et *sait* à chaque niveau. En dehors des erreurs portant sur la sur-généralisation de certaines graphies normées (ou « stratégies de point d'appui » – cf. Delsol 2003) nous retrouvons la plupart des autres types d'erreurs répertoriées par Delsol (2003) dans son étude concernant l'acquisition de l'orthographe du /sE/ en cycle 3, et par Brissaud (2011) en ce qui concerne l'orthographe en général, à savoir : l'erreur phonologique (les enfants transcrivent ce qu'ils entendent dans des graphies inventées), la reconstitution du son par récupération de monosyllabes existantes (et leur juxtaposition d'après des phénomènes connus de la chaîne parlée) ainsi que la production de graphies existantes correspondant (ou pas) au son recherché. Ces divergences (autour de 8% pour les trois niveaux) illustrent aussi bien l'étendue de la variété des écritures possibles du son /sE/ que le trouble suscité par la conscience des combinaisons potentielles du français écrit (en témoignent plus particulièrement les propositions de type *s'es* ou *c'ait*).

Enfin, ces données nous ont permis de constater que le niveau du public ciblé (CM2), étant donné l'effectif non négligeable des groupes de CM1 et de sixième au sein duquel il se positionne, se situait dans une certaine norme, du moins relativement à l'espace géographique concerné.

4.3.2 Résultats du groupe expérimental et du groupe témoin

Le Graphique 3 montre les résultats moyens des deux groupes à chaque évaluation.



Ecarts-type :	**Pré-test**	**Exercice**	**Post-test**
Groupe expérimental	1,5	2,4	2
Groupe témoin	2,3	2,3	2,6

Figure 3 : Moyennes sur 10 des groupes expérimental et témoin

Avec une première évaluation qui a donné lieu à des moyennes pratiquement identiques et a révélé une hétérogénéité plus prononcée dans le groupe témoin, il apparaît que les résultats à l'exercice de contrôle reflètent a priori la même efficacité des deux enseignements lorsque les

enfants disposent des formes écrites sous les yeux pour compléter un texte. Lors de la troisième évaluation, l'écart entre les résultats du pré-test et ceux du post-test du groupe expérimental est très exactement deux fois supérieur à celui du groupe témoin (3 points de plus contre 1,5). Si nous demeurons prudents quant à l'interprétation de la hausse des résultats du groupe test (il nous est en effet impossible de déterminer dans quelle mesure l'orthographe des formes a pu se trouver renforcée pendant la période qui a suivi l'exercice de contrôle), nous précisons toutefois que l'un des enseignants d'un des sous-groupes du groupe témoin s'est servi de la dernière dictée comme support d'évaluation « réel » et a intégré les différentes formes en rappel dans les consignes, ce qui laisse supposer que la moyenne générale de ce groupe aurait été encore diminuée si l'absence de référence orthographique en rappel avait été respectée.

Dans ces circonstances, et conscients des vérifications à plus grande échelle requises dans l'interprétation de ces résultats, il semblerait néanmoins que l'ASC ne conduise pas à de meilleurs résultats immédiats mais favorise plutôt le développement d'un terrain plus propice à assimiler de nouvelles compétences (pour certains plus que pour d'autres étant donné le creusement de l'écart-type final). Autrement dit, l'ASC constituerait ici une base plus solide pour que l'enfant puisse ensuite progresser par lui-même (que les résultats finaux reflètent un seuil atteint ou une dynamique encore en cours). Enfin, bien que les écarts-types soient élevés dans les deux groupes, on constate que l'ASC a généré des disparités moins importantes que dans le groupe témoin même si elles se trouvent être plus importantes qu'au début.

En ce qui concerne le nombre de graphies inappropriées précédemment décrites, le rapport du pré-test au post-test dans le groupe témoin est passé de quatre cas relevés chez un élève à cinq cas répartis sur quatre élèves. Dans le groupe expérimental ce même rapport est passé de huit cas répartis sur quatre élèves à trois sur deux élèves. Alors que l'ASC semble avoir conduit à une diminution de ces graphies aussi bien du point de vue quantitatif qu'au niveau du nombre d'élèves concernés, les enseignements habituels semblent avoir eu l'effet inverse puisque plus d'élèves commettent davantage d'erreurs à l'arrivée dans le groupe témoin.

4.3.3 Résultats par variable morphographique

Les taux de réussite de l'orthographe des formes aux emplacements requis dans les groupes témoin et expérimental (voir les Graphiques 4 et 5) font apparaître des progressions différentes pour chacune des formes, notamment entre le pré-test (T1) et l'exercice intermédiaire (T2).

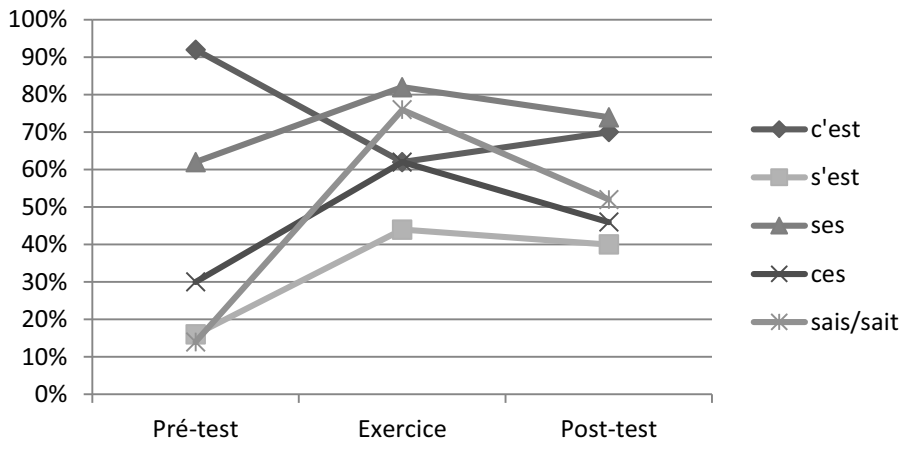


Figure 4 : Taux de réussite de placements des homophones dans le groupe témoin

Dans le groupe témoin, toutes les formes sauf *c'est* ont subi une nette progression entre T1 et T2 (un test de khi deux donne un résultat significatif, $p < .05$) mais les niveaux (sauf pour *c'est*) baissent légèrement deux mois après (et ce de façon significative pour *ces* et *sais/t*). La mise à disposition des élèves de toutes les morphographies au moment de l'exercice montre que la présence simultanée des autres formes a conduit les élèves à ne plus autant recourir à *c'est* en tant que forme dominante, ce qui confirme le caractère semi-aléatoire des résultats positifs concernant cette forme pendant le pré-test. De plus, le fait que la remontée de *c'est* au post-test soit loin de rejoindre les 92% de réussite du début laisse supposer que sa confrontation avec les autres homophones n'a fait qu'accroître les confusions liées à son utilisation. Il est cependant difficile de déterminer si cette baisse générale, dont l'évolution ne semble progressive que par sa représentation en courbe, reflète une réelle tendance (des résultats qui continueraient à baisser au fil du temps) ou l'état d'une compétence orthographique qui s'est améliorée au point de représenter un niveau désormais acquis (avec toutefois un pic lors de l'exercice intermédiaire).

Quant au groupe expérimental (Graphique 5), la même remarque concernant l'évolution significative (sauf pour *ses*) de la maîtrise des formes entre T1 et T2 s'impose.

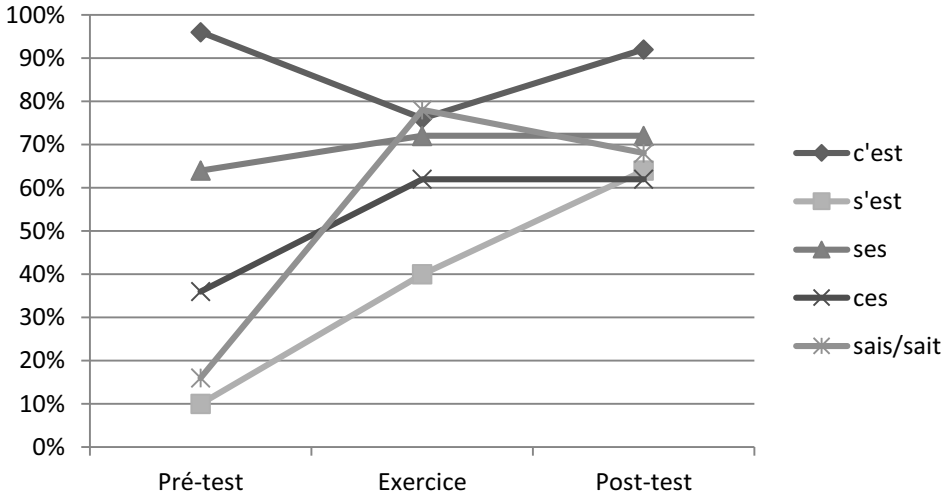


Figure 5 : Taux de réussite de placements des homophones dans le groupe expérimental

Toutefois, si la baisse des résultats pour *c'est* lors de l'exercice peut s'interpréter de la même manière que précédemment (une confusion attribuable à la présence des autres formes), nous remarquons que les élèves en ont récupéré quasiment le même taux de réussite un mois après. On remarque ensuite une avancée significative pour *s'est* avec le temps T1-T3 (χ^2 50,69, ddl=2, $p < .01$) (même si l'évolution globale pour toutes les formes prises ensemble n'atteint pas un niveau significatif). Si les formes *ces* et *ses* semblent s'être stabilisées, la progression de *s'est* est d'autant plus unique qu'il s'agissait de la forme la moins maîtrisée avant l'expérimentation. Par ailleurs, concernant les formes *sait/sais*, la proportion de réussite plus détaillée de ces formes en fonction de la variante requise aux évaluations (voir le Tableau 2) indique que la progression de la variante sollicitée à l'exercice (*sais*) a, contrairement à ce que laisse voir le graphique, finalement suivi le même parcours que les formes *c'est* et *s'est*, à savoir une meilleure maîtrise encore lors du post-test (+6 points) pour le groupe expérimental et un léger déclin pour le groupe témoin (-8 points).

	Pré-test	Exercice	Post-test	
	sait	***sais***	***sait***	***sais***
Groupe témoin	14%	76%	36%	68%
Groupe expérimental	16%	78%	52%	84%

Tableau 2 : Taux de réussite de placements des formes *sait* et *sais* dans les deux groupes

En conclusion, on note que les enfants du groupe témoin ont progressé par rapport à leur niveau initial mais qu'aucune des formes n'est mieux (ou aussi bien) utilisée dans la dictée finale que lorsqu'elles sont visuellement présentes au moment de l'écriture. A l'inverse, chez le groupe ayant expérimenté l'ASC, les progrès sont notables pour la totalité des formes homophoniques (notamment en ce qui concerne les formes *s'est* et *sais/t*) et ne régressent dans aucun des cas.

5 Discussion

Tout d'abord, l'approche que nous avons développée (choix des occurrences, typologie des exercices) est à ce jour très semblable à celle que nous aurions suivie dans le cadre de la L2. Toutefois, il est possible qu'une meilleure prise en compte des connaissances des natifs sur leur propre langue produise de meilleurs résultats, ou en tout cas donne lieu à des tâches ou des considérations différentes de ce que l'on décrit pour la L2. Le travail sur corpus en L1 reposerait ainsi sur une considération autre du bagage linguistique en vigueur qui ne ferait que renforcer l'étendue de son champ d'action (essentiellement mis en avant en L2) permettant d'en inclure a priori bien davantage qu'en L2. Cependant, le potentiel de transversalité des diverses formes ou notions que l'on cherche à faire acquérir par le biais des corpus en L1 peut très bien aussi exister en L2 d'une manière qui reste à explorer. La question sur l'adaptation des angles d'approche en fonction du caractère natif ou non natif de celui qui « apprend sur corpus » demeure ouverte donc.

Une deuxième remarque concerne ensuite l'âge des enfants (ou des apprenants s'il s'agit d'une L2) étant donné que, jusqu'à présent, l'ASC a typiquement été testé avec des

apprenants adultes (à l'université pour la plupart des études – Boulton/Tyne 2014). En effet, si la question de l'utilisation de concordances se pose tout naturellement pour des apprenants de L2, tant ils ne sont pas habitués à les utiliser et ont parfois besoin d'encadrement (Kerr 2013) ou d'un certain savoir-faire sur corpus afin de profiter de ce nouvel input (Kübler 2014), elle se pose aussi pour l'apprentissage de la L1 notamment en raison de l'âge et du degré de maîtrise du code écrit du public concerné, à savoir des enfants. Or, s'il s'avère que pour Sealey et al. (2004 ; Sealey 2011), comme pour nous, le recours aux corpus comme input ne soulève pas de difficultés particulières auprès d'enfants âgés de 9-10 ans, qu'en serait-il pour un public encore plus jeune ? Ainsi que nous l'avons précisé, le faible effectif (huit élèves) des CE2 (7-8 ans) et l'absence d'un groupe témoin similaire ne nous a pas permis d'inclure pleinement ces enfants dans notre étude[14]. Cependant, compte-tenu du peu de recherches explorant l'ASC auprès d'enfants en bas âge, il demeure intéressant de présenter, à titre parenthétique, les résultats obtenus avec ce public (voir le Tableau 3), tout en restant naturellement prudent quant aux conclusions éventuelles.

Pré-test	***c'est***	***s'est***	***ses***	***ces***	***sait***		Moyenne/10	Ecart-type
	81%	0%	44%	25%	0%		3	1,8
Exercice	***c'est***	***s'est***	***ses***	***ces***	***sais***		Moyenne/10	Ecart-type
	56%	19%	44%	50%	31%		3,9	1,2
Post-test	***c'est***	***s'est***	***ses***	***ces***	***sait/sais***		Moyenne/10	Ecart-type
	64%	7%	64%	36%	62%	62%	4,9	1,9

Tableau 3 : Taux de réussite de placement des homophones chez les CE2

Le premier constat (rassurant) est que l'ASC, en l'absence d'enseignement explicite, n'aura pas porté préjudice aux enfants puisque leurs résultats au post-test demeurent meilleurs qu'à la première évaluation. Ces résultats semblent suivre la même dynamique que pour le groupe des CM2, à savoir une amélioration progressive de leur compétence orthographique au-delà de l'exercice, une maîtrise des variantes *sais* et *sait* et une diminution de l'emploi correct (par défaut) de *c'est* au profit d'un meilleur emploi des autres formes. En ce qui concerne le nombre (relativement élevé) de graphies erronées présentes lors du pré-test (22 soit 28,8% du total des formes), il s'est vu réduit à un seul (*sai*) au moment du post-test, ce qui suggère que c'est sur ce plan que l'ASC semble avoir été le plus profitable, c'est-à-dire dans la restriction des formes « sauvages » au profit de formes choisies dans une palette de formes existantes. L'ensemble de ces résultats laissent ainsi penser que le facteur âge ne constitue pas nécessairement un frein à l'ASC et que des exercices plus adaptés auraient peut-être conduit à de meilleurs résultats.

6 Conclusion

Si l'apprentissage sur corpus, dont bien des avantages ont été identifiés pour l'enseignement-apprentissage des L2, n'est pas très présent dans l'apprentissage de la L1 (tout comme il n'est

[14] Leur participation s'explique par leur appartenance à la même classe (double niveau) qu'une partie des élèves de CM2 impliqués dans cette étude.

pas utilisé avec de jeunes apprenants en général), les résultats de cette étude semblent montrer qu'il y aurait à gagner à en faire une méthode (parmi d'autres) pour l'enseignement de la L1, en s'inspirant des approches décrites pour l'enseignement de la L2. De plus, la question du choix des données est posée avec le recours aux corpus de données authentiques, avec de surcroît la double question du développement des exercices selon qu'il s'agit de la L1 ou de la L2 (incluant la possibilité que l'approche de la L1 serve à enrichir le travail sur la L2, par exemple). Enfin, bien que l'objectif de l'ASC ait été de rendre saillant aux yeux des enfants un ensemble de phénomènes qui les aide à résoudre divers conflits, de répondre à différentes questions, il n'est pas pour autant évident de déterminer quels aspects dans cette approche (l'alignement de la forme, son association à des collocatifs récurrents, la lecture répétée, le rappel implicite du sémantisme, le nombre/la diversité des exemples, le style des exercices, etc.) auront eu le plus d'impact, d'autant que cet impact a pu être variable d'un individu à l'autre pour divers facteurs : difficultés résiduelles au niveau du décodage, perception et investissement variables face à l'autonomie, effet nouveauté, esthétique des supports, etc. Mieux cerner les apports réels de ce type d'approche dans le développement de l'orthographe nécessiterait de creuser encore davantage cette problématique tant au niveau de l'expérimentation qu'au niveau pratique et méthodologique.

Bibliographie

Bentolila, Alain (2010): *Parle à ceux que tu n'aimes pas. Le défi babel*. Paris: Jacob.

Bilger, Mireille (2008): « De l'intérêt des corpus diversifiés pour les descriptions en (morpho)syntaxe. Réflexions et illustration avec le pronom relatif lequel ». *Verbum* 30/4: 275–286.

Blanche-Benveniste, Claire (2010): *Approches de la langue parlée en français.* 2e éd. Paris: Ophrys.

Boulton, Alex (2012): "Hands-on / hands-off: varying approaches in data-driven learning". In: James, Thomas/Boulton, Alex (eds.): *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno, Masaryk University Press: 152–168.

Boulton, Alex et al. (2015 [2013]): « Corpus et appropriation de L1 et L2 ». *LINX* 68/69: 9–32.

Boulton, Alex/Tyne, Henry (2014): *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*. Paris: Didier.

Brissaud, Catherine/Jaffré, Jean-Pierre/Pellat, Jean-Christophe (eds.) (2008): *L'orthographe aujourd'hui: regards croisés*. Limoges: Lambert.

Brissaud, Catherine (2011): « Didactique de l'orthographe: avancées ou piétinements? ». *Pratiques* 149/150: 207–226. http://pratiques.revues.org/1740 [30.06.2015].

Brissaud, Catherine (2014): PU – CSP Contribution. http://cache.media.education.gouv.fr/file/CSP/38/7/BRISSAUD_Catherine_-_PU_-_CSP_Contribution_374387.pdf [30.06.2015].

Brissaud, Catherine/Cogis, Danièle (2011): *Comment enseigner l'orthographe aujourd'hui?* Paris: Hatier.

Catach, Nina (1973): « La structure de l'orthographe française ». *La Recherche* 39/4: 949–956.

Catach, Nina (1988): « L'écriture en tant que plurisystème, ou théorie de L prime ». In: Catach, Nina (ed.): *Pour une théorie de la langue écrite*, actes de la table ronde internationale CNRS-Heso. Paris, CNRS: 243–259.

Catach, Nina (1995): *Dictionnaire historique de l'orthographe française*. Paris: Larousse.

Catach, Nina (2001): *Histoire de l'orthographe française*. Paris: Champion.

Chaves, Natalie et al. (2012): « La mémorisation de l'orthographe des mots lus en CM2: effet du traitement visuel simultané ». *L'Année psychologique* 112/2: 175–196.

Chervel, André (1991): « L'école républicaine et la réforme de l'orthographe (1879-1891) ». *Mots* 28: 35–55. www.persee.fr/web/revues/home/prescript/article/mots_0243-6450_1991_num_28_1_2033 [30.06.2015].

Chervel, André/Blanche-Benveniste, Claire (1969): *L'orthographe.* Paris: Maspéro.

Cobb, Tom (2014): "A resource wish-list for data-driven learning in French". In: Tyne, Henry et al. (eds.): *French through corpora: Ecological and data-driven perspectives in French language studies.* Newcastle, Cambridge Scholars: 255–290.

Cury Pollo, Tatiana/Kessler, Brett/Treiman, Rebecca (2009): "Statistical patterns in children's early writing". *Journal of Experimental Child Psychology* 104: 410–426.

David, Jacques/Guyon, Odile/Brissaud, Catherine (2006): « Apprendre à orthographier les verbes: le cas de l'homophonie des finales en /E/ ». *Langue française* 15: 109–126.

Delsol, Alain. (2003): « L'acquisition de l'orthographe des homophones non homographes de /sE/? ». *Les Dossiers des Sciences de l'Education* 9: 77–88.

Di Vito, Sonia (2015 [2013]): « L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE ». *LINX* 68/69: 159–176.

Dister, Anne et al. (2009): *Penser l'orthographe de demain*. Paris: CILF.

Duda, Richard et al. (2009): « Table ronde: faut-il aménager les documents authentiques en vue de l'apprentissage? ». *Mélanges CRAPEL* 31: 273–286.

Duret, Laura/Zecca, Adeline (2013): « Peut-on enseigner efficacement les homophones grammaticaux en les étudiant séparément? Vers une nouvelle approche de l'enseignement des homophones grammaticaux ». *Education* 2013. http://dumas.ccsd.cnrs.fr/dumas-00958890 [30.05.2016].

Ellis, Nick (1997): "Vocabulary acquisition: Word structure, collocation, grammar, and meaning". In: McCarthy, Michael/Schmidt, Norbert (eds.): *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge, Cambridge University Press: 122–139.

Ellis, Nick (1998): "Emergentism, connectionism and language learning". *Language Learning* 48/4: 631–664.

Fayol, Michel/Jaffré, Jean-Pierre (2008): *Orthographier*. Paris: PUF.

Fayol, Michel/Pacton, Sébastien (2006): « L'accord du participe passé: entre compétition de procédures et récupération en mémoire ». *Langue française* 151/3: 59–73.

Fayol, Michel/Zorman, Michel/Lété, Bernard (2009): "Associations and dissociations in reading and spelling French: Unexpectedly poor and good spellers". *British Journal of Educational Psychology* 6: 63–75.

François-Denève, Corinne (2012): *50 règles essentielles: orthographe française*. Levallois-Perret: Studyrama-Vocatis.

Groupe RO (2012): « Orthographe: ce qui est jugé difficile. L'avis d'enseignants et de futurs enseignants ». *Glottopol* 19: 17–36.

Gilquin, Gaëtanelle/Gries, Stefan (2009): "Corpora and experimental methods: A state-of-the-art review". *Corpus Linguistics and Linguistic Theory* 5/1: 1–26.

Johns, Tim/King, Philip (eds.) (1991): "Classroom Concordancing". *English Language Research Journal* 4: 27–45.

Kemper, Marion/Verhoeven, Ludo/Bosman, Anna (2012): "Implicit and explicit instruction of spelling rules". *Learning and Individual Differences* 22: 639–649.

Kerr, Betsy (2013): "Grammatical description and classroom application: Theory and practice in data-driven learning". *Bulletin suisse de linguistique appliquée* 97: 17–39.

Kübler, Natalie (2014): « Mettre en œuvre la linguistique de corpus à l'université: vers une compétence utile pour l'enseignement/apprentissage des langues? ». *Recherches en didactique des langues et des cultures: Les Cahiers de l'Acedle* 11/1: 37–77.

Léon, Pierre/Léon, Monique/Léon, Françoise/Thomas, Alain (2009): *Phonétique du FLE. Prononciation: de la lettre au son*. Paris: Colin.

Lété, Bernard (2003): "Building the mental lexicon by exposure to print: A corpus-based analysis of French reading books". In: Bonin, Patrick (ed.): *Mental lexicon. 'Some words to talk about words'*. Hauppauge/New York, Nova Science Publisher: 187–214.

Lété, Bernard (2006): « L'apprentissage implicite des régularités statistiques de la langue et l'acquisition des unités morphosyntaxiques. L'exemple des homophones-hétérographes ». *Langue française* 151/3: 41–58.

Lété, Bernard/Sprenger-Charolles, Liliane/Colé, Pascale (2004): "MANULEX: A grade-level lexical database from French elementary-school Readers". *Behavior Research Methods, Instruments and Computers* 36: 156–166.

Lewis, Michel (1993): *The Lexical Approach: The State of ELT and the Way Forward*. Hove: Language Teaching Publications.

Lewis, Michel (1997a): *Implementing the Lexical Approach: Putting Theory into Practice*. Hove: Language Teaching Publications.

Lewis, Michel (1997b): "Pedagogical implications of the lexical approach". In: Coady, James/Huckin, Thomas (eds.): *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge, CUP: 255–270.

Lodge, R. Anthony/Shelton, Jane/Ellis, Yvette/Armstrong, Nigel (1997): *Exploring the French Language*. Londres: Arnold.

Maïonchi-Pino, Norbert/Magnan, Annie/Ecallé, Jean (2010): "Syllable frequency effects in visual word recognition: Developmental approach in French children". *Journal of Applied Developmental Psychology* 31: 70–82.

Mout, Tiphaine/Vernet, Samuel (2012): « Réforme de l'orthographe du français: qu'en pense le monde enseignant? Enquête dans six pays francophones ». *CMLF 2012 SHS Web of Conferences*: 2203–2215.

Pacton, Sébastien/Perruchet, Pierre/Fayol, Michel/Cleeremans, Axel (2001): "Implicit learning out of the lab: The case of orthographical regularities". *Journal of Experimental Psychology: General* 130/3: 401–426.

Sealey, Alison (2011): "The use of corpus-based approaches in building children's knowledge about language". In: Ellis, Sue/Elspeth, McCartney (eds.): *Applied Linguistics and Primary School Teaching*. Cambridge, Cambridge University Press: 93–106.

Sealey, Alison/Thompson, Paul/Scott, Mike (2004): "An investigation into corpus-based learning about language in the primary school". *ESRC Research Award* R000223900. http://www.leeds.ac.uk/educol/documents/189413.pdf [30.05.2016].

Tomasello, Michael (2003): *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge MA: Harvard University Press.

Tyne, Henry (2013): « Corpus et apprentissage-enseignant des langues ». *Bulletin suisse de linguistique appliquée* 97: 7–15.

Sealey, Alison (2011): "The use of corpus-based approaches in building children's knowledge about language." In Ellis, Sue & Elspeth McCartney (éds.): *Applied Linguistics and Primary School Teaching*. Cambridge, Cambridge University Press: 93-106.

Sealey, Alison, Paul Thompson & Mike Scott (2004): "An investigation into corpus-based learning about language in the primary school." ESRC Research Award R000223900. http://www.leeds.ac.uk/educol/documents/189413.pdf

Tomasello, Michael (2003): *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge MA: Harvard University Press.

Tyne, Henry (2013): "Corpus et apprentissage-enseignant des langues." *Bulletin suisse de linguistique appliquée* 97: 7-15

ANNEXE : Extraits de séquences pour *s'est* et *ces*

Extraits de la presse :

1	C'est exactement l'inverse qui	s'est	produit.	Le Nouvel Observateur
2	Martine Aubry	s'est	aussitôt mise en travers de son chemin.	Libération
3	mais, à cause de ce qui	s'est	passé, j'ai décidé cette fois de dire non. »	Libération
4	L'issue de la réunion ministérielle qui	s'est	déroulée, lundi 25 août à Matignon,	Le Monde
5	Fin de l'histoire ? Ken	s'est	engagé à dix-huit ans dans les brigades	Le Monde
6	La jeune fille	s'est	révélée combative et résistante.	Le Monde
7	Tout du long de sa carrière politique, il	s'est	battu pour ses convictions,	Libération
8	Martini ne	s'est	jamais demandé si ce poste lui était destiné.	Libération
9	Lari, actuel ministre de l'intérieur, qui	s'est	battu avec une belle pugnacité contre le Conseil	Le Nouvel Observateur
10	après sa sortie de prison en 1995, Tyson	s'est	marié avec Monica Turner, une étudiante	Libération

1) Quel est le point commun des mots surlignés en violet ?

.....

2) Parmi les 4 participes passés soulignés, lequel n'est pas un synonyme de « avoir lieu » ?

.....

3) Remplace le pronom « il » par « elle » dans l'extrait n°7 et accorde le participe passé :

➢ Tout au long de sa carrière politique, ... s'est ..... pour ses convictions.

Extraits d'enregistrements :

1	c'est un peu l'habitude chez	ces	gens-là euh il vivait à nos crochets quoi	PAROLE privée
2	il faut être plus ouvert par rapport à	ces	personnes-là	PAROLE privée
3	le fait de voir tous	ces	gens réunis + c'était super	PAROLE privée
4	références de l'ouvrage en question + donc dans	ces	cas-là soit on l'a en magasin + soit on	PAROLE privée
5	tu n'as jamais été dans	ces	pays-là	PAROLE privée
6	+ et bien toutes	ces	personnes trouvent très facilement du travail	PAROLE privée
7	+ on s'est aperçu de	ces	choses-là +	PAROLE privée
8	il faut se mettre un peu à la place de	ces	gens-là	PAROLE privée
9	je me suis aperçue au cours de	ces	soirées que	PAROLE privée
10	c'est possible oui + seulement si	ces	conditions-là sont réunies sinon	PAROLE privée

1) Quel est le **nom** le plus fréquent <u>après</u> « ces » ? Réponse : .....

2) Mets ces deux extraits au singulier :
   - (n°4) dans ces cas-là ➔ dans.....
   - (n°9) au cours de ces soirées ➔ au cours de.....

3) Regarde les extraits et devine le mot qui manque :
   - En plus, j'étais à l'internat à cette époque-.....
   - À ce moment-....., on ne savait pas encore ce qui allait se passer.
   - Il faut se méfier de ces gens-.....