

# Difficultés orthographiques d'apprenants débutants plurilingues en Français Langue Seconde

**Nathalie GETTLIFFE**

UR2310 Laboratoire interuniversitaire des sciences de l'éducation et de la communication (LISEC), Université de Strasbourg, de Lorraine et de Haute Alsace  
ngettliffe@unistra.fr

Although the French orthographic system has been the subject of numerous studies highlighting some of its inherent complexity, researchers in Applied Linguistics rarely rely on these results to establish the areas of difficulty for learners of French as a Foreign or Second Language. Our research aims to understand the challenges of an adult audience ( $n = 20$ ) as they enter the French orthographic system (first 20 hours of A1 level). We based our study on the multi-system approach of Catach (1995), to which we added phonemico-graphic and ideogrammic components to analyze the errors produced during five dictations proposed at the end of the circle of literacy (neurolinguistic approach). Results and analyses show that 63% of the words were error-free. Most errors were phonemico-graphic (59%), meaning that the learners had the wrong acoustic perception of the sounds composing a word or failed to choose the correct graphemes to transcribe certain sounds at the heart of the French phonological system (archiphoneme [OE] and nasal vowels). The learners also struggled with specific components of the French orthographic system such as apostrophes or the use of capitals (ideogramms). Morphogrammic, logogrammic, and phonogrammic errors are residual and represent each less than 10% of the errors detected.

## 1. Introduction

Les travaux portant sur l'apprentissage de l'orthographe pour les publics étudiant le Français Langue Étrangère et Seconde (FLES) sont relativement peu nombreux. Cette compétence est scrutée par de nombreux spécialistes (psycholinguistes, sociolinguistiques, didacticiens, enseignants, sociologues) en français langue maternelle, en revanche les apprenants de FLES semblent un peu oubliés. Or, en interrogeant ces derniers, il est évident que l'orthographe est au cœur de leurs préoccupations alors qu'ils sont confrontés à un système qu'ils perçoivent comme complexe (Li, 2021). En effet, alors que certains systèmes scripturaux privilégient très largement la représentation phonographique (transcription des unités sonores avec des lettres ou des syllabes), le français s'appuie en partie sur la représentation sémiographique (Jaffré & Pellat, 2008) avec des lettres qui ne représentent pas des sons mais, soit un sens (exemple : la lettre *s* pour marquer le pluriel), soit une étymologie (exemple : la lettre *h* dans le mot *huile*).

Quelques études sur l'apprentissage de l'orthographe par des publics FLES (Luzzati, 2010 ; Makassikis & Pellat, 2011 ; Li, 2021) montrent que les zones de difficultés (phonétique, morphologie, homophonie) semblent varier selon le niveau en langue (intermédiaire, avancé) ; cependant, les méthodes de collecte de données sont variées (dictée ; productions libres ; grilles d'observation personnelles) et ne permettent pas de conclure avec certitude quelles sont les zones de difficultés pour chaque niveau. De plus, aucune étude ne fait état des difficultés pour les débutants.

Notre étude propose de s'appuyer sur un corpus de dictées d'apprenants en FLES, débutants et faux-débutants, afin d'identifier les zones de fragilité orthographique des premiers écrits. En mobilisant plusieurs descriptions synchroniques du système orthographique du français (Catach, 1995 ; Lété, 2008 ; Fayol & Jaffré, 2014), nous proposons dans un premier temps de cartographier sa complexité. Nous compléterons notre propos en synthétisant les difficultés déjà répertoriées par les quelques recherches portant sur l'apprentissage de l'orthographe par des apprenants en FLES. Nous continuerons en décrivant le contexte de notre étude qui s'est appuyée sur un dispositif d'enseignement particulier, à savoir l'approche neurolinguistique ; cette dernière, par le biais de la boucle de la littéracie, met l'accent sur le développement de la précision linguistique (notamment orthographique). Finalement, nous présenterons et discuterons les résultats de notre étude.

## **2. Revue de la littérature**

### **2.1 *L'orthographe du français : approches synchroniques***

Plusieurs spécialistes ont tenté de décrire de manière synchronique l'état actuel du système orthographique français qui, de par son figement, inclut de nombreuses lettres dites silencieuses et laisse apparaître des transcriptions plus sémiographiques que phonographiques (ex : huent = [y]).

Catach (1995) propose, en s'appuyant sur les recherches du laboratoire CNRS - HESO (Histoire et Structure des Orthographe et Systèmes d'écriture) une approche plurisystémique de l'orthographe du français avec une

répartition des graphèmes en quatre zones – phonogrammique, morphogrammique, logogrammique, étymologique et historique.

Concernant la première zone, l'auteure rappelle, tout d'abord, que le système orthographique français est à 80% phonogrammique, à savoir que l'écriture d'un mot reflète dans sa très grande majorité l'ensemble des phonèmes qui sont prononcés. Cependant, une des premières difficultés du français réside dans l'existence d'un grand nombre de transcriptions pour un même phonème (par exemple pour /e/ ou /s/). C'est pourquoi elle propose de regrouper dans la catégorie des archigraphèmes un ensemble de transcriptions d'un même phonème ou archiphonème, ce qui lui permet de mettre en évidence les points de variation du système de transcription des phonèmes en français. Certains archigraphèmes sont stables, ex : [U] s'écrit presque toujours *ou*, de même que [WA] se marque systématiquement par le bigramme *oi*. D'autres archigraphèmes ont des réalisations concurrentielles, ex : [O] = *o*, *au*, *eau* (Catach, 1995, p. 10-15).

La seconde zone du système orthographique du français retranscrit, selon Catach (1995), des informations sur la morphologie, soit flexionnelle (le genre, le nombre, les conjugaisons), soit dérivationnelle (ex : retard avec un *d* muet pour former retarder, retardataire, retardement, etc.). Ces parties de mots sont appelés les morphogrammes et forment environ 5% des écrits courants.

La troisième partie du système orthographique contient des mots qui se transcrivent de manière différente afin de refléter des sens distincts : ce sont les logogrammes (ex : *ou/où* ; *a/à* ; *bon/bond*). Ils représentent aussi environ 5% des mots des écrits courants.

Finalement, les mots français peuvent contenir des lettres étymologiques muettes (qui reflètent des stades antérieurs : sculpteur) ou des lettres historiques (qui ont été ajoutées au gré des réformes : huile). Malheureusement, cette orthographe sans fonction particulière (Catach, 2011, p. 57) représente une partie importante (10%) des mots que nous mobilisons régulièrement, même si elle a déjà fait l'objet de réformes récentes afin de réduire le nombre d'idiosyncrasies (mots composés avec ou sans trait d'union,

accent circonflexe ...). Ainsi, la description synchronique de Catach (1995) permet d'avoir une vision simplifiée de l'orthographe française avec notamment un nombre réduit d'archigraphèmes (n=33) qui permettent de transcrire phonétiquement la plupart des mots (80%) ; ce système pluri-systémique est aussi constitué de sept morphogrammes grammaticaux (s, x, e, t, d, z, r) qui sont mobilisés dans les écrits courants et de 12 logogrammes à connaître pour éviter la plupart des erreurs (Massarenti, 1971).

Fayol et Jaffré (2014) confirment les spécificités du système orthographique français déjà énoncées par Catach (1995), à savoir sa polyvalence graphique (130 phonogrammes pour une trentaine de phonèmes) et la présence d'un nombre important de lettres non prononcées du fait de l'amuïssement des fins de mots au cours de l'histoire. Ils évoquent aussi l'existence des accents et des lettres étymologiques qui complexifient l'écriture de certains mots ; finalement, pour eux, la morphologie grammaticale apparaît comme assez désordonnée avec plusieurs marques pour une même notion (ex : pluriel= -s, -x, changement phonologique de -al/ vers -aux ou absence de marques).

Lété (2008) propose une analyse statistique du système orthographique du français en calculant la consistance de la relation phonème/graphème ou graphème/phonème pour chaque mot, à savoir « la proportion de mots pour lesquels le phonème est associé à un graphème particulier relativement au nombre total de mots dans lesquels le phonème apparaît, quelle que soit sa façon d'être orthographiée » (p.85). Par exemple, le phonème /ɛ̃/ s'écritra dans 18% des cas avec le graphème *ain* en fin de mot. Afin de mener des analyses statistiques, il a constitué avec des collègues (Peerman & al., 2007) une base de données nommée *Manulex-infra* composée de textes de lecture proposés aux enfants de CP, CE1 et Cycle 3 (au total 45 000 mots).

Plusieurs résultats portant sur ce corpus permettent de dessiner certaines particularités du système orthographique du français. Tout d'abord, Peerman et al. (2007) ont pu montrer que la consistance des mots variait selon que le phonème ou le graphème se situait en début, milieu ou fin de mot. Respectivement dans le sens graphème-phonème (lecture à voix haute), la

consistance moyenne était de 94% (début), 78% (milieu) et 85% (fin). Dans le sens inverse phonème-graphème (écriture), la consistance moyenne selon les positions était de 83% (début), 74% (milieu) et 44% (fin). Ces résultats montrent qu'il est plus facile de lire que d'écrire en français mais surtout que le risque d'erreurs est plus important en fin de mots puisque le lien phonème-graphème est ténu, notamment à cause des morphologies flexionnelle et dérivationnelle, qui sont pour la plupart silencieuses. Une deuxième analyse statistique a porté sur la consistance des mots en fonction de leur nature (mono- ou polysyllabiques). Ainsi, plus un mot est court et plus sa consistance est faible, aussi bien dans le sens lecture à voix haute qu'écriture. Par exemple, en fin de mot, la consistance pour les monosyllabiques est de 39% alors qu'elle est de 52% pour les mots de plus de quatre syllabes.

Finalement, à partir de la même base de données, Lété (2008) a testé l'assertion : « plus un mot est fréquent et plus sa consistance est faible » : or, la corrélation entre ces deux facteurs est très faible ( $r < .06$ ). Les mots fréquents ne sont donc pas plus difficiles à lire à voix haute (en déchiffrant des graphèmes) qu'à écrire (en transcrivant des phonèmes) par rapport aux autres mots.

## 2.2 *Les difficultés orthographiques des apprenants en FLS*

En ce qui concerne l'acquisition de l'orthographe par des élèves ou des adultes ayant le français comme langue étrangère ou seconde, les études sont peu nombreuses mais quelques tendances émergent.

Plusieurs contributions issues d'un numéro thématique de la revue *TRANEL* (2011) dédié à cette question, notamment celles de Amokrane (2011), Gonac'h et Mortamet (2011), ont pu montrer que les erreurs étaient plutôt à dominance phonogrammique. Les erreurs à dominance morphogrammique constituent le deuxième point faible des écrits en FLES (autour de 30% pour des étudiants de niveau avancé selon Makassikis & Pellat, 2011). Elles portent souvent sur la redondance des marques du genre, fréquemment absent des langues sources des apprenants (Gonac'h & Mortamet, 2011). Pour ce qui est

des marques du nombre à l'écrit, la thèse d'Ågren (2008) a permis de caractériser trois stades de développement pour les jeunes apprenants de FLE : stade initial (quantifieur, nom, pronom, déterminant), stade intermédiaire (verbe irrégulier, verbe régulier, verbe avec alternance de base verbale) et stade avancé (adjectif). Les erreurs à dominante logogrammique (homonymes) sont moins présentes chez les étudiants allophones de niveau avancé que chez les locuteurs natifs (Gonac'h & Mortamet, 2011 ; Makassikis & Pellat, 2011).

D'un point de vue méthodologique, le système pluri-systémique de Catach (1995) ne permet pas de classer toutes les erreurs relevées chez les apprenants de FLES. Par exemple, des auteurs (Luzzati, 2010 ; Geffroy Konštacký & Nováková, 2011 ; Li, 2021) ont pointé l'existence d'erreurs de type phonologique pour les apprenants de FLES (erreurs dues à un mauvais encodage des suites sonores composant un mot : \*crocodile) alors que d'autres (Gonac'h & Mortamet, 2011) regroupent ces erreurs dans la zone morphogrammique du système. Ces erreurs sont notées comme extra-graphique par Catach (1995, p. 282). De plus, des erreurs de type idéogrammique, qui portent sur la perception de la frontière des mots (soudures ou segmentations erronées), ont été relevées chez de jeunes enfants allophones (Le Bot & Richard, 2011 : \**c'est ai trô beau, je tes bien la bas*) ou chez des scripteurs peu lettrés dialectophones lors de leur correspondance privée pendant la Grande Guerre (Dal Bo, 2019 : \**Tes cheres l'etre je les li et reli je ne sai combien de foi*) ; cependant, ce type d'erreurs est mentionné de manière anecdotique par Catach (1995).

### 3. Objectifs de l'étude

Si le système orthographique français a fait l'objet de nombreuses études qui ont permis d'établir des zones de complexité (description linguistique par Catach, 1995 ; Lété, 2008 ; Peerman & al., 2007 ; Fayol & Jaffré, 2014), ces recherches sont très faiblement mobilisées pour établir les difficultés des apprenants de FLES, qu'ils soient enfants ou adultes. De plus, les études

s'appuient souvent sur des groupes homogènes avec des apprenants partageant les mêmes langues premières et avec des compétences en français plutôt de niveaux intermédiaire ou élevé.

Notre recherche tentera de comprendre les zones de fragilité d'un public adulte plurilingue (n=20) avec des langues premières différentes (n=14) alors qu'il entre dans le système orthographique français (niveau A1). Ainsi, nous chercherons à répondre aux questions suivantes :

- Quelles sont les erreurs orthographiques de débutants en FLES ?
- Ces erreurs reflètent-elles des catégories déjà relevées dans d'autres études en FLES ?
- Ces erreurs sont-elles le reflet des zones de complexité notées par les spécialistes de l'orthographe ?

## 4. Méthodologie

### 4.1 *Le dispositif pédagogique : l'approche neurolinguistique (ANL)*

Notre étude a été menée en contexte écologique : les productions analysées sont issues d'un cours de langue. Ce dernier était de niveau A1.1, c'est-à-dire qu'il s'agissait de proposer un contenu pour des apprenants qui n'avaient jamais suivi de cours de français ou qui souhaitaient réactiver des compétences acquises lors d'une scolarité ancienne.

Le cours était construit selon les principes de l'approche neurolinguistique (Germain, 2017) qui promeut la précision linguistique aussi bien à l'oral qu'à l'écrit (Germain & al., 2004a ; Germain & al., 2004b ; Gettliffe, 2020). En effet, la boucle de la littéracie<sup>1</sup> mobilisée dans cette approche permet de réduire le

---

<sup>1</sup> Etapes de la boucle de la littéracie : Interactions à l'oral avec mémorisation des suites sonores composant des phrases – lecture avec repérage des régularités et des irrégularités graphiques pour transcrire des sons ou des informations grammaticales – écriture de paragraphes – lecture à voix haute des textes produits. L'enseignant reprend systématiquement toute erreur de prononciation ou d'écriture pendant chaque phase.



décalage entre oralité et écriture d'un mot, aussi bien dans le sens lecture-écriture qu'écriture-lecture (Gettliffe, 2023).

Notre étude porte sur les cinq premiers modules d'enseignement du niveau A1.1, à savoir les 20 premières heures.

## 4.2 *Participants*

Les participants (n=20) inscrits au cours de français niveau A1.1 avaient des langues premières très diverses : portugais (3), allemand (2), anglais (2), hongrois (2), polonais (2), chinois (1), croate (1), espagnol (1), grec (1), italien (1), letton (1), roumain (1), tagalog (1) et turc (1). Les non-anglophones utilisaient l'anglais comme langue relais<sup>2</sup> dans un cadre professionnel. La plupart des participants avaient la trentaine, un niveau d'études supérieures important (au moins Bac+3) et vivaient dans un milieu où le français est une des langues officielles du pays.

## 4.3 *Instruments, mesures et analyses*

Afin de vérifier les acquis des apprenants, une dictée qui reprenait en grande partie des mots prononcés et écrits pendant les différentes phases de l'ANL (oral-lecture-écriture) a été proposée à la fin de chaque module (environ 4 heures de cours). L'ensemble des dictées produites par le groupe et rédigées sur ordinateur était affiché sur un écran numérique et chaque apprenant, à tour de rôle, était invité à repérer des erreurs éventuelles et proposer des modifications sur une phrase rédigée par un autre apprenant. Il pouvait s'appuyer sur des propositions graphiques qui se trouvaient dans les dictées de ses pairs. Les autres apprenants pouvaient aussi intervenir pour proposer d'autres modifications. À la fin de chaque tour, l'enseignante confirmait les propositions et pointait les erreurs restantes.

---

<sup>2</sup> Une langue relais est une langue différente de la langue première, apprise par l'apprenant et sur laquelle il peut s'appuyer pour faire des inférences concernant la nouvelle langue qu'il est en train d'apprendre (Defays & Saenen, 2011).



La dictée a été envisagée au sein de l'approche ANL comme un moyen de vérifier la compréhension orale des apprenants mais aussi leurs compétences orthographiques. Les négociations graphiques collectives étaient nécessaires afin que chaque apprenant puisse encoder dans sa mémoire à long terme une représentation correcte des mots et des accords. Cette dernière peut ensuite être récupérée plus tard par la voie directe, qui est plus rapide et plus efficace (Fayol & Jaffré, 2014)<sup>3</sup>. Aucune notation chiffrée n'a été proposée.

Les cinq dictées proposées comptaient entre 56 et 167 phonèmes à transcrire (voir annexe 1). La plupart des mots avaient été présentés dans les modules d'apprentissage mais certains étaient nouveaux de manière à renforcer l'attention. Au total, notre analyse porte sur 78 dictées (certains apprenants ayant été absents pendant l'exercice).

Les erreurs de chaque locuteur ont été codées selon les catégories du plurisystème orthographique proposé par Catach (1995, p. 282-283), à savoir :

- les erreurs à dominante phonogrammique : réalisation graphique erronée du phonème ou de l'archiphonème (*\*Italy*)
- les erreurs à dominante morphogrammique : omission, erreur, ajout d'une marque de genre, de pluriel ou de forme verbale (*\*une ville international*)
- les erreurs à dominante logogrammique : confusion entre deux homophones (*\*sont appartement*)
- les erreurs à dominante non fonctionnelle : oubli (*\*toujour\_*)

Nous avons complété notre grille par la catégorie *Erreurs à dominante idéogrammique* (majuscule, ponctuation, apostrophe, trait d'union) (Catach, 1995) afin de rendre compte de certaines erreurs de soudures ou de segmentations erronées.

---

<sup>3</sup> La voie directe est un processus cognitif qui permet le rappel de l'écriture d'un mot directement de la mémoire à long terme sans passer par sa décomposition phonétique.

Nous avons aussi rebaptisé la catégorie *Erreurs à dominante extragraphique* (Catach, 1995) en *Erreurs à dominante phonémico-graphique* car cette composante phonologique était relevée dans de nombreux travaux en FLES alors qu'elle apparaît comme une composante mineure dans la grille de Catach (1995). Il peut s'agir d'omissions ou d'adjonctions de phonèmes ainsi que de confusions entre voyelles (\*d pour de) ou consonnes (\*noise pour noir).

Des analyses statistiques descriptives ont été réalisées afin de calculer le nombre d'erreurs total, les catégories d'erreurs les plus fréquentes et pour chaque catégorie d'erreurs, les principales difficultés.

## 5. Résultats

### 5.1 Typologie des erreurs (vue générale)

Au total, nous avons relevé 1104 erreurs pour les cinq dictées sur un total de 2988 mots, c'est-à-dire que 37% des mots contenaient une erreur.

La répartition des erreurs dans le graphique ci-dessous indique que les apprenants ont surtout commis des erreurs à dominante phonémico-graphique (59%), suivi d'erreurs à dominante idéogrammique (12%). Les erreurs restantes sont à dominante phonogrammique (10%), morphogrammique (10%), logogrammique (8%) et étymologique (1%).

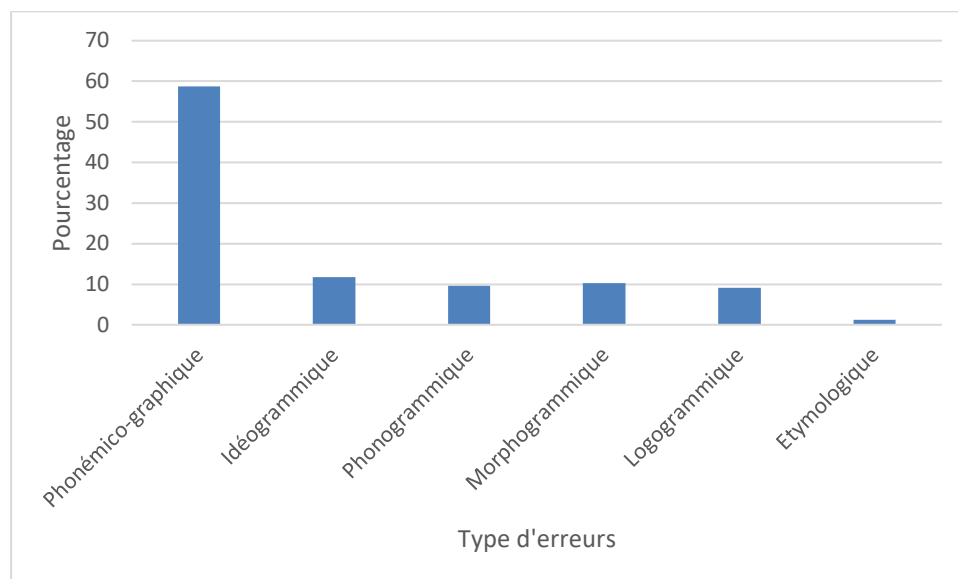


Figure 1 : Répartition des erreurs en orthographe des apprenants débutants en FLES (%)

## 5.2 Typologie des erreurs (vue détaillée)

### 5.2.1 Les erreurs à dominante phonémico-graphique

Les *erreurs à dominante phonémico-graphique* représentent 648 erreurs sur un total de 9634 phonèmes à retranscrire, ce qui signifie que seulement 7% des phonèmes à retranscrire contiennent une erreur de ce type. Ces erreurs se répartissent en trois catégories : omission(s), erreur(s) de représentation phonologique/transcription ou ajout de phonème(s).

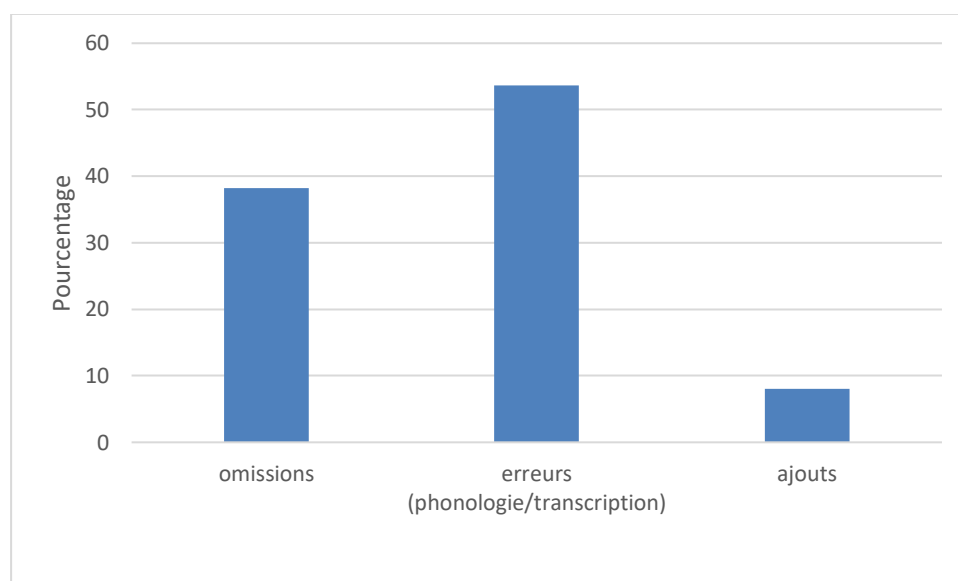


Figure 2 : Type d'erreurs à dominante phonémico-graphique (%)

Une catégorie importante d'omissions (20% des erreurs à dominante phonémico-graphique) porte sur l'oubli de la lettre *e* en fin de mot, graphie générale de la consonne en finale de mot (Catach, 1995, p. 45) (*\*architect*, *\*timid*, *il \*arrêt*, *\*dam* pour *dame*, *\*canadien* pour *canadienne*), de même que l'oubli de la lettre *s* pour former le son [e] : *\*de/des* ; *\*se/ses* ; *\*le/les*.

Pour ce qui est de la deuxième catégorie *d'erreurs à dominante phonémico-graphique*, à savoir les confusions, ces dernières portent à 91% sur de mauvaises représentations ou retranscriptions<sup>4</sup> de *voyelles* ; les sons consonantiques sont presque parfaitement restitués. On note que les apprenants ont le plus de difficultés avec l'archiphonème [œ], avec les bigrammes représentant des voyelles nasales (*in*, *an*, *on*) et le schwa. D'autres spécificités du français posent problème dans une moindre proportion, à savoir le bigramme *ou*, les accents sur la lettre *e* ainsi que la transcription du phonème /o/. Finalement, les phonèmes /a/, /y/ et /i/ sont relativement bien transcrits (moins de 10% d'erreurs).

Phonème	Réalisations erronées (%)	Exemples
[œ]/[Ø]	62,27	*yeaux/yeux ; *restoratore/restaurateur ; *ne/neuf ; *never/9h
[ẽ]	48,08	*italian ; *a studio/un studio ; *ba/bain ; *la fa/la fin ; *ciquante/50
[ə]	44,53	*du Strasbourg ; *promiere/premier ; *lu quartier ; *sol da bains/salle de bain
[ɔ]	41,67	*s'amper/son père ; *cant/compte

<sup>4</sup> Plusieurs hypothèses peuvent expliquer les erreurs phonémico-graphiques dans une dictée: soit l'apprenant ne connaît pas ou ne reconnaît pas un mot et il transcrit les sons qu'il a entendus en faisant des erreurs sur les sons mal perçus ; soit il reconnaît le mot, mais il a encodé une mauvaise version sonore du mot et il retranscrit cette version erronée (cela arrive quand il y a des interférences avec des langues premières proches du français); soit il a une représentation sonore correcte du mot mais il fait une erreur au moment de la transcription graphique en choisissant une ou des mauvaise(s) lettre(s).

[ɑ]	39,07	*sont tro/centre ; *don lu quartier ; *sont/cent ; *ou Italie/en Italie
[u]	25,64	*strasburgo ; *seuvant/souvent ; *beaceau/beaucoup
[e]	24,45	*e ne/est né ; *electrique ; *propietere ; *etage ; *cartie/quartier
[o]	24,29	*souf/sauf ; *ou Canada/au Canada ; *beacoup ; *restoirateur
[ɛ]	13,33	*mere ; *peer/père ; *françoise/français ; *mire/mère
[a]	7,33	*sans salle de bain ; *an Paris/à Paris ; *somer/sa mère
[y]	6,41	*de matain/du matin
[i]	1,71	*t'emide/timide

Figure 3 : Transcriptions incorrectes des phonèmes (%)

Concernant les *consonnes*, il s'agit principalement du phonème /s/ qui est mal orthographié dans sa version du *c cédille* comme dans le mot *français*.

La dernière catégorie *d'erreurs à dominante phonémico-graphique* porte sur les ajouts de lettres qui ne correspondent à aucune réalisation phonétique (\*architecte) ou qui transcrivent des liaisons (\*il est t<sub>l</sub>architecte). L'ajout des lettres *e*, *s* et *r* est particulièrement saillant (38% des ajouts) et participe à faire « sonner » des consonnes finales qui devraient rester muettes (\*anglaise/anglais) ou changer la sonorité du mot (\*des/de).

### 5.2.2 Les erreurs à dominante idéogrammique

Nous poursuivons en détaillant le type d'erreurs à dominante idéogrammique (n=127) qui sont pointées dans le référentiel orthographe FLE (Luzzati, 2010) comme le premier jalon à maîtriser. La ponctuation représente 42% des erreurs avec très souvent l'ajout de marques de ponctuation (point ou virgule) (\**Martine e Italienne, e Canadienne*). Les erreurs concernent aussi l'intervention des lettres majuscules et minuscules (42%) notamment dans la dictée 1 qui fait alterner la nationalité (\**elle est italienne*) et les langues (\**Elle parle Français*). On retrouve encore ici des graphies typiques du système

graphique germanique (\*un Studio ; \*une Ville) ou anglais (...et Il<sup>5</sup> arrêt). L'apostrophe apparaît comme une zone de fragilité avec des réalisations qui révèlent une mauvaise compréhension de sa fonction typographique, à savoir marquer une élision (\*Ilest architecte ; \*Ilem ; \*de'Alsace ; \*l'Il parle ; \*s'amper/\*s'amer pour son père/sa mère ; \*se s'enfants ; t'emide pour timide ; \*Il nest pas ; \*ele sapele). Les autres erreurs portent sur le trait d'union (\*centreville) ou l'invention de marques diacritiques sur les lettres (\*Canadâ).

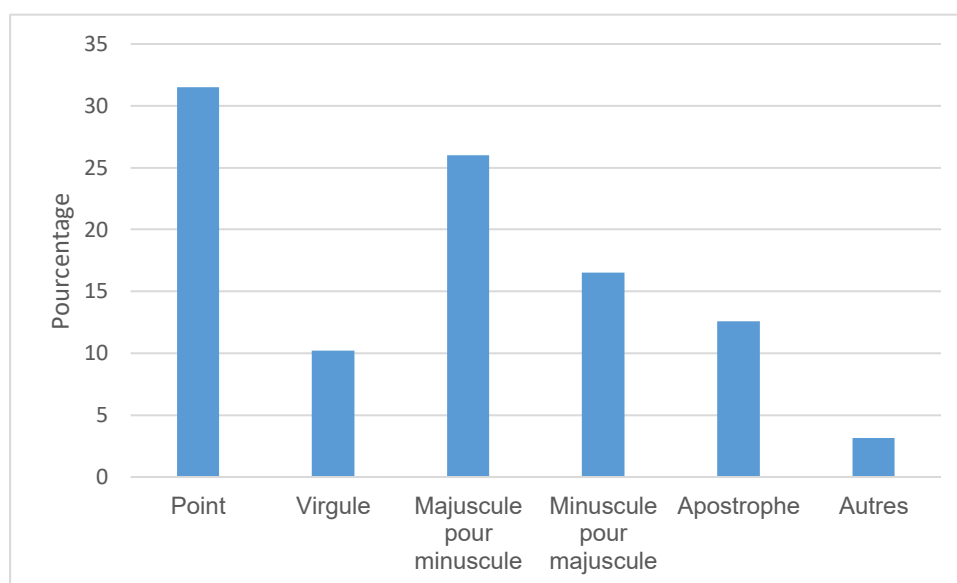


Figure 4 : Détails des erreurs à dominante idéogramme (%)

### 5.2.3 Les erreurs à dominante phonogramme

Les erreurs à dominante phonogramme (n=104) regroupent toutes les transcriptions qui sont correctes d'un point de vue phonétique mais qui ne mobilisent pas la bonne suite de lettres. On remarquera qu'une des premières difficultés concerne l'écriture des pronoms *il/elle* retranscrits \*ille, \*ile, \*ill, \*el, \*e// (17 erreurs)<sup>6</sup>. La transcription de [ɛ] est aussi problématique pour les

<sup>5</sup> En anglais, le pronom *I* (« je ») utilise une lettre majuscule. Par analogie, les étudiants utilisent fréquemment le *I* majuscule pour le pronom *il* en français.

<sup>6</sup> Un relecteur nous a suggéré de considérer ces unités lexicales comme étant des logogrammes. Nous avons conservé notre catégorisation initiale car la classification des erreurs de Catach (1995) impose que le logogramme possède au moins deux sens pour une même suite sonore, ce qui n'est pas le cas pour les pronoms *il* et *elle*. Dans nos exemples, on comprend que les apprenants produisent par analogie *ille* à partir de *elle* ou *el* à partir de *il* en essayant de mettre

apprenants puisqu'on peut utiliser un accent, des bigrammes (ei/ai), des lettres diacritiques (ez, redoublement des consonnes) ou aucune marque pour réaliser ce phonème : \*arrête, \*arrette, \*françaïs, \*italiene, \*canadiene, \*peir, \*mair, \*per, \*mer/mère, \*commerce, \*propriétaire, \*nouvel/nouvelle, \*em/aime (16 erreurs). Les autres difficultés sont assez anecdotiques : doubles consonnes autres que celles servant à transcrire le son [ɛ] (5 erreurs), écriture du [o] (4 erreurs), écriture du [i] dans \*Italy/Italie (3 erreurs).

#### 5.2.4 Les erreurs à dominante morphogrammique

Les erreurs à dominante morphogrammique (n=111) reflètent en très grande majorité l'omission du pluriel (70% des erreurs) (\*des pont) suivie d'erreurs de terminaisons dans les conjugaisons (\*il fais, \*il aimes, \*il ne vois pas) (14%). Le féminin est régulièrement utilisé pour le masculin (\*il est née) et rarement oublié. On remarque une grande variabilité concernant la présence de la marque du pluriel. Tous les apprenants ont écrit *yeux* et *cheveux* avec un x pour le pluriel et seuls 4 d'entre eux ont commis une erreur en oubliant le x ou en ajoutant un s pour *des châteaux*. Les déterminants sont corrects à 82% sur l'ensemble des dictées, les noms à 74% et les adjectifs à 63%. Finalement, on note que le marquage du pluriel se met en place de gauche à droite, à savoir que si le déterminant n'est pas marqué alors le nom et l'adjectif qui suivent ne portent pas non plus de marquage.

#### 5.2.5 Les erreurs à dominante logogrammique

Pour ce qui est des erreurs à dominante logogrammique (homophones) (n=100), les erreurs portent principalement sur la confusion à/a (30% des erreurs), *et/est* (21%) ainsi que *sont/son* (10%) et *ou/où* (6%).

---

de l'ordre dans une orthographe qui leur paraît confuse tout en inventant des formes intermédiaires *ile/ill/ell*.



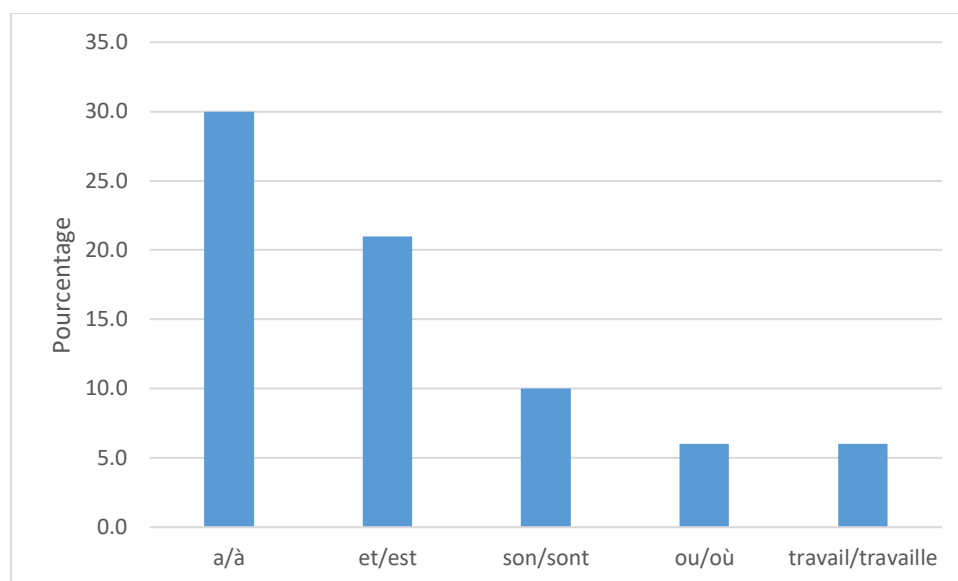


Figure 5 : Répartition des erreurs à dominante logogrammique (%)

### 5.2.6 Les erreurs concernant les lettres historiques ou étymologiques

Finalement, les erreurs à dominante étymologique ne concernent que quelques mots : \*beaucoup, \*il n'est pa, \*toujour, \*prè/près, \*chaufage, \*chateau, \*mil, et représentent très peu d'erreurs (n=14).

## 6. Discussion

### 6.1 *Distinguer erreurs de perception et erreurs de transcription*

Plusieurs études se sont attachées à répertorier les erreurs d'orthographe d'apprenants en français langue étrangère ou seconde; toutefois, ces dernières se sont appuyées principalement sur des corpus d'étudiants avancés (Amokrane, 2011 ; Makassikis & Pellat, 2011 ; Essoh Ndobbo & Ebica Odey, 2015 ; Katoozian, 2013 ; Li, 2018, 2021). Les erreurs relevées étaient surtout à dominante morphogrammique, logogrammique, ou elles portaient sur des lettres historiques ou étymologiques. Notre recherche s'est attachée à relever les erreurs d'apprenants de français débutants ou faux-débutants avec un niveau de littéracie élevé dans leur langue maternelle et une connaissance de l'anglais, soit comme langue première (2 apprenants), soit comme langue relais/professionnelle. Les résultats montrent clairement que la majorité des erreurs (59%) sont de type phonémico-graphique, une catégorie quasiment

absente chez les enfants de langue maternelle qui entrent dans l'écriture et peu relevée chez les étudiants qui possèdent déjà un niveau intermédiaire à avancé en français langue seconde ou étrangère. En revanche, cette catégorie avait déjà été relevée chez des étudiants iraniens de niveau débutant à intermédiaire (Katoozian, 2013) dans des proportions assez similaires (autour de 60%).

Notre corpus regroupant des apprenants avec différentes langues maternelles, les erreurs phonémico-graphiques relevées ont pu faire émerger certaines particularités du système phonologique français qui posent des difficultés aux apprenants, comme la perception de l'archiphonème [OE] et des voyelles nasales, et ceci quelle que soit leur langue première. La dictée reposant en partie sur la perception des sons, il est clair que si le mot n'est pas reconnu (par exemple *neuf heures*), l'apprenant devra s'appuyer uniquement sur une répétition en boucle des suites de sons entendus pour tenter de transcrire le mot. Si sa perception sonore est erronée, la transcription sera alors fautive. En revanche, si le mot est reconnu, c'est-à-dire qu'à une suite sonore un sens est associé, la transcription par voie directe (rappel à partir de la mémoire à long terme sans passer par la transcription son par son) est possible, ce qui peut expliquer que les mots *cheveux* et *yeux* dans la dictée 2 aient été transcrits presque toujours parfaitement alors qu'ils mobilisent un son « difficile », à savoir le /ø/. Comme le relèvent Fayol et Jaffré (2014), la mobilisation de la voie directe est à encourager puisqu'elle permet une plus grande efficacité dans la transcription (précision et vitesse). Cependant, il faut que le mot soit dans un premier temps reconnu (perception), ce qui reste problématique pour les apprenants débutants.

## 6.2 Le code (ortho)graphique du français

Certaines erreurs relevées pointent vers des spécificités du code graphique mobilisé pour écrire le français. Ainsi, l'utilisation de l'apostrophe pour marquer l'élision est un premier élément de difficulté pour les apprenants. Cette marque typographique apparaît très rapidement dans les premiers apprentissages (Je

*m'appelle ...*) et est mobilisée par les apprenants pour transcrire certaines suites sonores : *\*s'amper/\*s'amer* (pour son père/sa mère), *\*Ill'aime*, etc.

Le deuxième élément du code graphique posant problème porte sur l'ensemble des signes diacritiques que ce soient des accents ou des lettres. Dans un premier temps, les apprenants débutants mémorisent les mots sans leur accent (*\*mere*), avec le mauvais accent (*\*il est nè*) ou encore avec des accents fantaisistes (*\*Canadá*) sans forcément faire le lien entre son et transcription, un peu comme si ces accents faisaient partie de l'image du mot. La cédille est aussi un signe diacritique qui n'est pas assimilé par les apprenants débutants alors qu'elle apparaît aussi rapidement avec le mot *français*. Concernant les lettres diacritiques, on peut pointer l'utilisation en français des doubles consonnes<sup>7</sup> pour transcrire le [ɛ], ce qui n'est pas assimilé par les apprenants débutants : *\*Canadiene*, *\*Italiene*, *je m'\*appele*, *\*el*. Un troisième élément perturbateur du code graphique français correspond à l'utilisation en finale de la lettre e pour donner une valeur pleine à la consonne finale (*petit/petite*) ou l'emploi de la lettre s après un e pour changer le timbre de la voyelle en [e] (*de/des ; se/ses ; le/les*).

Les erreurs à dominante phonogrammique, à savoir la réalisation graphique erronée d'un phonème ou d'un archiphonème, sont relativement limitées pour notre niveau débutant car elles sont éclipsées par les erreurs à dominante phonémico-graphique. Ainsi, alors que Catach (1995) relève une parité dans la transcription du son /ã/ par *an* ou *en*, ces erreurs ne sont pas relevées dans notre corpus car les apprenants ont le plus souvent écrit *in* ou *on*. Cette

---

<sup>7</sup> En français, il existe des signes diacritiques (les accents, la cédille) mais aussi des lettres diacritiques comme certaines doubles consonnes qui changent la prononciation de la voyelle précédente. On peut aussi noter l'usage de la lettre –r ou –z pour marquer le son [e] comme dans chanter, le nez. Ces lettres avaient été imaginées pour transcrire le son des voyelles précédentes avant l'apparition des signes diacritiques, qui a eu lieu à peu près au même moment que l'apparition de l'imprimerie. Avant cette époque, il était difficile pour les moines scripteurs d'ajouter des accents car les lignes étaient trop serrées. De fait, l'écriture des sons vocaliques en français est une résultante de signes diacritiques (époque récente) et de lettres diacritiques (époque plus ancienne).

observation avait aussi été relevée par Katoozian (2013). Les consonnes sont transcrites correctement dans la majorité des cas et elles ne semblent pas poser de difficultés aux apprenants débutants dans leur variation phonogrammique. On peut s'étonner que les phonèmes /k/, /s/, /ʒ/ mobilisent presque toujours la bonne variante, comme par exemple *étage*, *jamais*, *chauffage*, *commerce*, *place* ou *quartier*, ces derniers étant cependant proche de leur équivalent en anglais, la langue relais.

Les erreurs à dominante logogrammique ne touchent pas un nombre important de mots dans notre corpus. Luzzati (2010) indique que les apprenants en FLES produisent en général moins ce type d'erreurs que les locuteurs natifs car ils peuvent associer une variante orthographique à sa traduction dans leur langue maternelle. Cependant, il est essentiel de stabiliser rapidement les représentations graphiques liées à chaque sens pour éviter que les apprenants ne privilégient excessivement une seule graphie, comme l'a observé Amokrane (2011).

### 6.3 Les fins de mots

Lété (2008) a relevé que les fins de mots sont plus difficiles à écrire en raison d'une absence de réalisation sonore des dernières consonnes reflétant une morphologie flexionnelle et dérivationnelle en grande partie silencieuse en français. Ågren (2008) a pu établir un ordre acquisitionnel en ce qui concerne la réalisation graphique des pluriels, à savoir que le marquage se met en place d'abord sur les déterminants, puis sur les noms et finalement sur les adjectifs. Notre corpus indique aussi que les chaînes d'accord se mettent en place de gauche à droite : on ne retrouve jamais un adjectif au pluriel si le déterminant et le nom ne sont pas déjà marqués.

Le Bot et Richard (2011) ainsi que Dal Bo (2019) avaient noté que la frontière des mots n'était pas toujours claire pour les personnes entrant dans l'écriture (soudures ou segmentations erronées). Dans notre corpus, il s'agit plutôt d'un effet de réorganisation syllabique du français qui peut prêter à confusion (*// commence à \*neveux ; a \*never/ 9 heures*) ou la présence de liaisons qui font

« sonner » des lettres silencieuses en fin de mots au regard de leur nouvel environnement linguistique (*\*Cette une ville historique ; au \*promiere étage ; Il est \*t'architect*).

## 7. Conclusion

Le propos de notre étude était de répertorier les différents types d'erreurs orthographiques des personnes débutant l'apprentissage du FLES en s'appuyant notamment sur des descriptions linguistiques synchroniques de zones de complexité (Catach, 1995 ; Lété, 2008 ; Fayol & Jaffré, 2014). Nous voulions également mobiliser des recherches en acquisition de l'orthographe en FLES afin de mettre en lumière les particularités d'un public débutant plurilingue avec différentes langues premières. Notre recherche a pu établir des différences majeures entre les apprenants débutants et ceux plus avancés. En effet, si ces derniers produisent plus d'erreurs de type morphogrammique, logogrammique ou sur les lettres non-fonctionnelles, les apprenants de français langue étrangère et seconde débutants doivent se familiariser dans un premier temps à la perception des sons de la langue avant de pouvoir les retranscrire : ainsi, leurs erreurs sont majoritairement de type phonémico-graphique, phonogrammique ou idéogrammique.

Il serait intéressant de compléter cette étude en analysant les difficultés orthographiques d'apprenants débutants lors de la rédaction de productions libres : ceci pourrait permettre de confirmer ou d'infirmer les zones de fragilité orthographique mises en évidence dans notre recherche.

## BIBLIOGRAPHIE

Ågren, M. (2008). *À la recherche de la morphologie silencieuse. Sur le développement du pluriel en FL2* [Thèse de doctorat, Université de Lund]. Études Romanes de Lund. Disponible : <https://portal.research.lu.se/sv/publications/%C3%A0-la-recherche-de-la-morphologie-silencieuse-sur-le-d%C3%A9veloppement>

Amokrane, S. (2011). Du lien entre la maîtrise du système phono logique et les compétences orthographiques. *TRANEL*, 54, 49-61.

Catach, N. (1995). *L'orthographe française*. Armand Colin.

Catach, N. (2011). *L'orthographe*. PUF.



Dal Bo, B. (2019). *Aux frontières de la norme : Usages linguistiques de scripteurs peu lettrés dans des correspondances de la Grande Guerre* [Thèse de doctorat, Université Paul Valéry, Montpellier III - Università degli studi, Gênes, Italie]. HAL. Disponible : <https://theses.hal.science/tel-02519070>

Defays, J.-M. & Saenen, F. (2011). De l'incidence du cotexte sur les choix orthographiques en FLE: Étude de cas. *TRANEL*, 54, 129-146. <https://doi.org/10.26034/tranel.2011.2871>

Essoh Ndobu, E. G. & Ebica Odey, V. (2015). Difficultés et erreurs dans le passage de la phonie à la graphie en dictée de langue française : Cas des apprenants du FLE de l'université de Calabar (Nigéria). *Glottopol*, 26, 157-166.

Fayol, M. & Jaffré, J.-P. (2014). *L'orthographe*. PUF.

Geffroy Konšťáček, D. & Nováková, S. (2011). De la phonologie à la graphie vers leur synergie: Une expérience tchèque. *TRANEL*, 54, 63-72.

Germain, C. (2017). *L'approche neurolinguistique (ANL) : Foire aux questions*. Myosotis Presse.

Germain, C., Netten, J., & Movassat, P. (2004a). L'évaluation de la production orale en français intensif : Critères et résultats. *Revue canadienne des langues vivantes/The Canadian Modern Language Review*, 60(3), 309-333.

Germain, C., Netten, J., & Séguin, S. (2004b). L'évaluation de la production écrite en français intensif. *Revue canadienne des langues vivantes/The Canadian Modern Language Review*, 60(3), 333-354.

Gettliffe, N. (2020). Les recherches portant sur l'Approche Neurolinguistique pour l'enseignement des langues étrangères et secondes : axes actuels et perspectives. *Didactique du FLES*, 1(1), 137-174. Disponible : <https://doi.org/10.57086/dfles.148>

Gettliffe, N. (2023). La boucle de la littéracie de l'ANL permet-elle d'équilibrer le développement de l'aisance à communiquer et de la précision linguistique à l'oral comme à l'écrit ? Une étude de cas de deux apprenants dans un centre d'enseignement du FLS. *Lidil. Revue de linguistique et de didactique des langues*, 68, 1-20. Disponible : <https://doi.org/10.4000/lidil.12168>

Gonac'h, J. & Mortamet, C. (2011). Pratiques orthographiques en français d'étudiants étrangers: Le cas d'étudiants hispano-américains et afghans. *TRANEL*, 54, 113-127.

Jaffré, J.-P. & Pellat, J.-C. (2008). Sémiographie et orthographes : Le cas du français. In C. Brissaud, J.-P. Jaffré, & J.-C. Pellat (éds.), *Nouvelles recherches en orthographe* (pp. 9-30). Lambert-Lucas.

Katoozian, K. (2013). *Analyse des erreurs orthographiques des étudiants iraniens en langue française* [Thèse de doctorat non-publiée]. Université Paris 3.

Le Bot, M.-C. & Richard, E. (2011). Pour une analyse morphologique des productions écrites d'élèves FLS. *TRANEL*, 54, 159-172.

Lété, B. (2008). La consistance phonographique : Une mesure statistique de la complexité orthographique. In C. Brissaud, J.-P. Jaffré & J.-C. Pellat (éds.), *Nouvelles recherches en orthographe* (pp. 85-99). Lambert-Lucas.

Li, Y. (2018). Une erreur peut en cacher une autre : Les apprenants chinois du français L2 face aux problèmes de morphologie verbale. *SHS Web of Conferences*, 46, 07014. Disponible : <https://doi.org/10.1051/shsconf/20184607014>

Li, Y. (2021). L'apprentissage de l'orthographe lexicale par les apprenants chinois de français langue étrangère : Analyse et réflexions didactiques à partir d'un corpus écrit. *Synergies Chine*, 16, 93-105.

Luzzati, D. (2010). *Le français et son orthographe*. Didier.

Makassikis, M. & Pellat, J.-C. (2011). Les étudiants natifs et allophones face à l'orthographe française : Le cas des homonymes. *TRANEL*, 54, 21-48.

Massarenti, L. (1971). *Psychopédagogie des moins doués*. Nathan.

Peerman, R., Lété, B. & Sprenger-Charolles, L. (2007). Manulex-infra : Distributional characteristics of grapheme-phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods*, 39, 593-603.

## ANNEXE 1 : LES DICTÉES

Dictée 1 : Elle s'appelle Martine. Elle habite à Paris dans le quartier de Notre Dame. Son père est né au Canada et il parle français et anglais. Sa mère est née en Italie et elle parle italien. Martine est Italienne et Canadienne.

Dictée 2 : Il s'appelle Martin. Il a les cheveux bruns et les yeux noirs. Il est architecte. Il n'est pas timide.

Dictée 3 : Pierre est restaurateur. Il travaille beaucoup. Souvent, il commence à 9 heures du matin et il arrête à minuit. Il fait une pause vers 15 heures. Il travaille toujours la fin de semaine. Il ne passe pas beaucoup de temps avec ses enfants sauf le mercredi après-midi où il ne travaille jamais.

Dictée 4 : Maxime est propriétaire d'un studio au centre-ville de Strasbourg. Son appartement est au premier étage. Son chauffage est électrique. Il aime son quartier calme avec des maisons typiques d'Alsace.

Dictée 5 : Lausanne est une ville internationale. Elle a beaucoup de commerces dans le centre-ville mais aussi 150 banques dans le quartier de Chauderon. Lausanne compte 137 000 habitants. C'est une ville historique avec une cathédrale, des places, des châteaux et des ponts.