

L'ORTHOGRAPHE DU FRANÇAIS : UN OBJET D'APPRENTISSAGE PROBLÉMATIQUE

Ce chapitre a pour objectif d'exposer le système orthographique de la langue française et d'expliquer les raisons de ses difficultés. A travers un aperçu de l'évolution de l'orthographe française au fil des siècles, nous souhaitons montrer « dans quelle mesure que la complexité orthographique du français est en grande partie due à son histoire un peu particulière » (Jaffré, 2009). Ensuite, grâce à la description de la mixité du système orthographique, tissé selon le principe phonographique et le principe sémiographique, ce chapitre tente de montrer le fonctionnement de ces deux principes dans le système graphique du français. Ces deux principes sont par ailleurs à « l'origine d'une certaine opacité et donc complexité de l'orthographe française » (Willemin, 2009). Le système graphique du français se caractérise alors par « des zones de fragilité orthographique » (Pérez & Garcia-Debanc, 2016). Donc dans l'étape suivante, nous présenterons comment orthographient les apprenants de français et aborderons la question des erreurs orthographiques.

1.1. Aspects sociolinguistiques : les Français et leur orthographe

Si l'orthographe française est « motivée par des structures linguistiques originales », elle fait également « l'objet de tractations, de débats, donc l'histoire conserve des traces » (Jaffré & Brissaud, 2006). Donc il n'est pas inutile de mettre l'évolution de l'orthographe française en perspective dans la présente étude et comme Jaffré l'évoque, « la cause de la complexité orthographique du français est d'abord à rechercher dans l'histoire elle-même » (2009), une suite de décisions contradictoires et de conflits entre « Anciens » et « Modernes », entre les défenseurs « d'une orthographe au service de la

lisibilité » et ceux « d'une écriture proche de la parole » (Willemin, 2009).

1.1.1. Éléments d'histoire du système de l'orthographe française

1.1.1.1. Les débuts de l'orthographe française

L'histoire de l'orthographe du français est étroitement associée à l'histoire de sa langue. Comme Jaffré l'indique (2009), la langue française « n'est pas un long fleuve tranquille » et on ne pourrait pas dire tout simplement que cette langue dérive du latin classique. Dès la fin du II^e siècle av. J.-C., les Romains s'emparent progressivement de la Gaule, couvrant les territoires actuels de la France. A la suite de la conquête de la Gaule, la République puis l'Empire romain modifient substantiellement les peuples conquis : ils ignorent les langues celtiques de ces peuples et font en sorte que le latin devienne indispensable. Au tournant du V^e siècle, l'unilinguisme latin est *de facto* atteint. Parallèlement au latin classique réservé à l'administration, à l'aristocratie et aux écoles, s'est développé un latin populaire, parlé à l'origine par les soldats romains. Ce sont de ces formes langagières populaires qu'émergeront les langues romanes.

Dès la période romaine, dans les espaces ruraux existe déjà une césure omniprésente entre le latin et les parlers locaux, situation trouvant son origine dans la diversité des substrats en présence. Cette distinction est bien plus forte et ancrée que dans d'autres pays de langue romane, contribuant de ce fait à isoler les langues gallo-romaines des autres langues romanes.

A partir du III^e siècle, la situation linguistique tend à se complexifier dans le sillage des invasions germaniques et de l'effondrement progressif de l'Empire romain d'Occident. Le latin n'est dès lors plus utilisé que pour les écrits et au latin populaire est substitué de proche en proche une langue nouvelle : le roman, comprenant diverses variétés. En envahissant le nord de la Gaule, les Francs s'approprient le gallo-roman à leur façon : la germanisation des parlers romans. Par l'ajout d'éléments germaniques, de nombreux Francs adoptent la langue des vaincus. La population parle à l'époque « *rustica romana lingua* » (langue romane rustique), sujette à des variations importantes

selon les régions. Le latin classique, en revanche, « n'en reste pas moins la langue de la loi, de l'administration, de l'Église, et plus généralement de l'élite sociale » (Fayol & Jaffré, 2008 : 104). Toute la Gaule connaît ainsi une situation de diglossie pendant une longue période. En outre différencié des autres langues romanes, dérivant d'un dialecte dominant, ce gallo-roman est « une forme composite faite de divers dialectes – normand, picard, champenois, etc. – que le XIX^e siècle nommera francien » (Fayol & Jaffré, 2008 : 104). Appréhendée en tant que langue véhicule de communication, le gallo-roman s'étend bientôt en pays d'oïl, puis en pays d'oc.

Selon Catach (2003 : 9), « le latin et le français ont vécu durant de nombreux siècles en état de véritable symbiose, ils étaient sentis comme une seule et même langue : on lisait le latin à la française, on écrivait le français à la latine ». Les *Serments de Strasbourg*, premiers documents de l'histoire du français, nous permettent de comprendre que les scribes « tentent d'appliquer l'orthographe latine à la langue vulgaire, à l'aide d'un alphabet déjà insuffisant pour le latin classique » (Catach, 2001 : 37). Cette symbiose d'une importance capitale caractérise ainsi l'évolution graphique du français. Catach note à ce titre les efforts d'adaptation phonétique et les hésitations du scribe dans les *Serments de Strasbourg* : *c* a deux valeurs selon la voyelle qui suit (*commun, cist*) ; *z* est en relation étymologique avec *e* dans *fazet* < *faciat*. De même, F. Brunot nous montre la sonorisation des explosives *p* et *e* (*savir, podir, auidha, cadhuna*), cependant, le *d* final étymologique dans *quid*, le *ch* dans *Christian*, le double *mm* de *commun*, les deux *l* de *nulla* sont conservés. Le mélange caractéristique d'orthographe étymologique et d'orthographe phonétique est déjà constaté. Au moins, avec quelques tâtonnements, les scribes s'efforcent d'indiquer la prononciation.

Dans cette longue histoire de la naissance de la langue française, on assiste à une réorganisation en profondeur de la structure syllabique. Avec les tendances telles que l'ouverture vocalique devant la consonne, la fermeture en syllabe ouverte, la prosodie du latin recule et s'estompe, ce qu'illustre notamment le déplacement de l'accent sur la dernière syllabe prononcée. Comme l'évolution habituelle est la conservation de la syllabe accentuée et la chute des voyelles finales autre que le *a*, on constate un affaiblissement important des consonnes ou des groupes de phonèmes internes,

aboutissant à un raccourcissement considérable du mot français. En ce qui concerne le placement ignoré ou erroné de l'accent tonique, il est particulièrement important de noter les effets produits par l'influence germanique. En raison des contacts fréquents avec les langues germaniques, le francique notamment, le gallo-romain se modifie davantage que les langues des autres espaces recomposés de l'ancien Empire romain. L'Italie ou l'Espagne connaissent ainsi des influences moindres. Avec leur accent nordique et leur système phonétique, les Francs assimilent de nouvelles tendances linguistiques au gallo-romain ; autrement dit, ils prononcent les voyelles romanes beaucoup plus fortement, ce qui a une incidence non négligeable sur l'orthographe française. De ce strict point de vue phonétique, la cohabitation linguistique du francique et du roman entraîne le déplacement de l'accent tonique, modifiant le timbre des voyelles. De ce fait, « la plupart des mots terminés par une voyelle audible dans d'autres langues romanes le sont en français par un *e* dit muet » (Fayol & Jaffré, 2008 : 105). De même, les consonnes intervocaliques ne sont pas intactes, ce qui contribue à l'usure généralisée de la structure phonétique des mots. Une multiplication des formes homophones est donc imputable à ces phénomènes. Signalons aussi que l'influence germanique est également assez importante sur le plan morphologique : les finales *-and*, *-ard*, *-aud*, *-ais*, *-er* et *-ier* sont d'origine francique, sans oublier un grand nombre de verbes en *-ir* comme *choisir*, *blanchir*.

Tous ces faits de langue constituent des illustrations éloquentes de la germanisation considérable à l'œuvre, dans le sens où le gallo-romain se différencie par bien des aspects des autres langues issues du latin, notamment au sud où les langues occitanes (ainsi que l'italien, l'espagnol, le portugais, etc.) restent plus proches de la langue latine que la langue française. Ces différenciations progressives ne seront pas sans entraîner, bien entendu, des changements majeurs sur l'orthographe de la langue française.

A partir des *Serments de Strasbourg* de 842 et jusqu'au XII^e siècle, on assiste à la naissance d'une littérature en langue vulgaire et d'une transcription proprement française.

Remarquons également qu'il existe une concentration de la diffusion des écritures :

elles sont aux mains des clercs lettrés. Ces écritures, créées par cette classe peu nombreuse, sont la belle minuscule caroline. Ce qui est particulièrement frappant, c'est qu'une langue parfaitement maîtrisée et stabilisée est utilisée dans les écritures, en dépit de la multiplication des dialectes. Catach (2001 : 47) nous présente quelques caractéristiques de cette langue :

- « La formation de nombreuses diphtongues (*ai, ue, ui, eu, ou*, etc.) et triptongues (*ieu, uou*, etc.) ».
- « Des phénomènes morphologiques originaux, comme les marques de séparation systématiques du masculin et du féminin (*mort/morte*) ».
- « La formation de nouveaux modes, temps et flexions verbales : le subjonctif, le futur et le conditionnel ».
- « De nouveaux paradigmes d'articles (*li, la, les, lo*), de pronoms et de mots-outils ».

Catach signale également les défauts de cette orthographe. Par exemple, elle reste dépendante de l'étymologie et a du mal à noter les voyelles. Le système vocalique de l'ancien français, surtout les diphtongues et les triptongues, évolue rapidement. Jusqu'à la fin du XIII^e siècle, l'orthographe s'efforce de suivre ces changements, puis elle se fige. Comme les transformations ne cessent jamais, on assiste à un écart de plus en plus important entre l'écrit et l'oral. En dépit de tous ces défauts, ces écritures constituent les fondations de l'orthographe d'aujourd'hui.

1.1.1.2. Le moyen français

Pendant la période gothique (XIII^e-XVI^e siècle), on assiste à une transition entre la langue ancienne et le français moderne. Les anciennes structures disparaissent alors que les nouvelles évolutions linguistiques se développent, caractérisant ce que l'on appelle le moyen français : affaiblissement de l'ancien système nominal et verbal, transformation de la syntaxe, bouleversement de l'ordre des mots, enrichissement considérable du vocabulaire, et surtout, évolution phonétique rapide. Cette instabilité

phonique, qui est propre au langage français, éloigne plus que jamais le français du latin et des autres langues romanes. Ces multiples transformations, d'une importance capitale, ont une influence sur le français (Catach, 2003 : 14-15) caractérisée par :

- « Un raccourcissement considérable du mot français » : caractère qui est particulièrement important, non seulement par le fait qu'il s'agit généralement des mots les plus utilisés et les plus utiles à la langue, mais aussi parce qu'il crée des risques nombreux d'ambiguïté par la multiplication des homophones.
- « Un relâchement des liens entre le français et le latin ».
- « Une tendance à la coupure entre radicaux et dérivés » : les premiers subissent une usure phonétique intense, alors que les seconds sont issus plus ou moins directement des radicaux latins.

Enfin, il nous paraît essentiel de rappeler les principales tendances de la graphie de l'époque (Catach, 2001 :78-79) :

- Tendance phonétique, d'une façon générale, dans laquelle la graphie de la quasi-totalité des orthographe, si arbitraire qu'elle apparaisse, ne s'éloigne pas de la langue parlée.
- Tendance étymologique, historique, morphologique et sémantique, qui joue un rôle particulièrement important pour la distinction des mots homophones et le problème des monosyllabes.

Après ce rappel des faits linguistiques concernant l'orthographe du XIII^e au XVI^e siècle, nous aborderons l'étude d'un grand tournant de l'orthographe du français.

A partir du XII^e siècle, l'écriture se transforme dans une direction nouvelle, vers la gothique, dans un contexte où s'épanouit la nouvelle culture de l'écrit : avec la laïcisation des clercs, ce ne sont plus seulement des scribes de monastères qui transcrivent les manuscrits, en peu d'exemplaires, mais de nouveaux clercs, qui passent rapidement de quelques centaines à plusieurs milliers ; en France, grâce à l'utilisation

du papier de chiffon, ce qui diminue considérablement les coûts, le commerce des livres devient florissant ; avec la création des universités, la demande des textes s'accroît de jour en jour.

Du XIV^e au XVI^e siècle, l'orthographe « s'éloigne du phonétisme pour prendre un aspect plus idéographique » (Catach, 2003 : 21). Ce premier caractère de l'orthographe du moyen français s'explique par la contradiction de l'écriture : d'un côté, par le besoin de rapidité de l'écriture, les lettres et les mots se rejoignent ; on voit également la multiplication des ligatures, des accolements, des abréviations. De l'autre côté, par l'impératif de lisibilité, on allonge les boucles et les jambages, on emploie de nombreux signes et repères afin de mieux distinguer les traces semblables, les syllabes et les mots différents. Parmi tous ces signes et repères, il faut noter particulièrement les variantes calligraphiques de positions (Catach, 2003 : 21) : le *s* rond final et le *S* long initial ou médian ; le *i* accentué puis pointé et le *I* long ou *J*, notant, suivant leur position, *i*, *y* et *j* actuels ; le *V* initial, notant indifféremment *u* et *v*, le *u* à l'intérieur, notant également *u* et *v*, etc. De ce fait, une conséquence importante se produit : « on prend l'habitude de déchiffrer chaque lettre d'après son contexte et sa position dans le mot » (Catach, 2003 : 21). On se sert d'une série de consonnes comme repère diacritique, en puisant tout naturellement dans le latin. Ainsi, l'étymologie n'est plus une fin, mais un moyen, qui apporte non seulement l'habitude des rapprochements analogiques, mais aussi de divers procédés graphiques permettant de distinguer le sens des mots.

Aux XIV^e – XV^e siècles, une grande différence sur le plan littéraire se présente entre la France et d'autres pays romans : la France est « entièrement soumise au bilinguisme culturel », alors que l'Italie, par exemple, « possède déjà une littérature reconnue, qui sert très tôt de modèle » (Catach, 2003 : 23). Les copistes italiens suivent ce modèle, et les grammairiens s'appuient sur ces traditions populaires littéraires et linguistiques. Étant donné la présence d'une littérature nationale, l'évolution graphique de l'italien est moins rapide et plus régulière que le français et il ne subit pas non plus les modifications profondes que nous venons d'étudier. Dans la période suivante, avec l'apparition de l'imprimerie, et d'une véritable littérature d'ampleur nationale, de considérables transformations vont être apportées à l'orthographe française.

1.1.1.3. Les deux renaissances

Avec l'effondrement des structures de l'ancien français, le français du XVI^e siècle connaît une évolution originale et rapide sous tous ses aspects : phonétique, morphologique, syntaxique et lexicale. N'étant pas en rupture avec le système de son ancêtre, cette langue évolue progressivement vers le français actuel.

A cette époque, centrée sur un retour direct aux textes sacrés, la religion réformée nécessite une lecture et une écriture accessibles et donc un usage des langues vulgaires. Avec le soutien du monde de l'imprimerie et en appui sur le français, les publications dans tous les domaines ont investi dans ce domaine ; on voit ainsi une première vague d'imprimés en langue française. De 1533 à 1564, la production annuelle des livres en français passe de 42 à 527 (Febvre & Martin, 1973 : 442-443). Parallèlement, une lutte pour la fixation du français écrit a lieu dans le monde de l'imprimerie. Dès le premier quart du XVI^e siècle, les imprimeurs français renoncent aux caractères gothiques et adoptent des caractères romains.

Cependant, les progrès sont aussi difficiles que rapides. Comme au XVI^e siècle, l'écriture de manuscrites reste fidèle aux habitudes gothiques et devient plus rapide, plus mauvaise et plus illisible ; le décalage entre les manuscrits et les imprimés ne cesse de grandir. Après quelques tâtonnements, les imprimeurs « renoncent à la copie textuelle des manuscrits, aux excès et aux signes de dégénérescence de l'écriture manuscrite » (Catach, 2003 : 26). Avec cette modernisation des impressions, émerge une orthographe moderne :

« Pour les imprimeurs d'avant-garde, l'orthographe moderne signifie avant tout la modernisation des impressions : caractères romains et italiques, bonne séparation des mots, usage modéré des majuscules, ponctuations plus riches, meilleure correction matérielle, normalisation des usages, suppression des abréviations et variantes arbitraires, mise en place des accents, etc. » (Catach, 2001 :119).

Ces réformes suivent et transforment l'aspect des textes en moins de vingt ans. Ainsi, un accord de normalisation des textes se produit progressivement entre éditeurs. Ce sont d'abord des imprimeurs d'importance secondaire publiant les textes poétiques qui suivent la nouvelle orthographe. Comme les poètes attachent de l'importance à l'aspect extérieur de leurs œuvres et à la facilité de lecture et de prononciation, ils s'adressent aux imprimeurs d'avant-garde, qui acceptent leurs exigences. Ils engagent ensuite les grands imprimeurs dans la voie des réformes. Ces derniers acceptent aussi certaines publications poétiques écrites dans l'orthographe nouvelle. Ainsi, sous l'influence des typographes et des petits ateliers, les grandes maisons s'acheminent vers « l'unité nouvelle de correction et des règles typographiques » (Catach, 2001 : 121).

Ce qui caractérise l'orthographe de la Renaissance, c'est tout d'abord « la beauté des impressions, le choix des caractères et le soin nouveau apporté à la correction » (Catach, 1968 : 250). En reprenant contact avec l'orthographe des chansons de geste, les poètes suppriment beaucoup de lettres adscrites et emploient les accents. Ils cherchent à éliminer également les lettres grecques et doubles, l'emploi de *y* pour *i*, du *z* marquant *e* fermé, du *x* muet final. Ils remplacent *en* prononçant [ã] par *an* dans de nombreux cas. Notons que ces réformes seront reprises par l'Académie au XVIII^e siècle, et qu'elles sont dans la lignée de la véritable tradition du français.

A la fin du XVI^e siècle, malgré les efforts des typographes d'avant-garde, les imprimeurs du roi reprennent les choses en main. De ce fait, le monde de l'édition connaît une crise. La qualité des livres subit une régression si importante que « les erreurs, les coquilles, l'ignorance des règles les plus élémentaires vont de pair avec le retour à l'orthographe la plus archaïque » (Catach, 2003 : 28). Il faudra attendre jusqu'en 1640, fondation de l'Imprimerie royale par Richelieu, pour que le monde de l'édition en France voie la fin de la crise.

A partir de la deuxième partie du XVII^e siècle, on assiste à une amélioration des éditions ainsi qu'au retour des habitudes graphiques de la Renaissance chez les écrivains. Les deux traditions, « modernisme » et « conservatisme » s'opposent et s'équilibrent. Notons que les dictionnaires sont aussi partagés : en 1680, Richelet publie le premier véritable dictionnaire français, dans lequel les marques morphologiques

essentielles sont gardées, de même que les tendances analogiques et étymologiques, mais elles le sont de façon plus modérée. Cependant, Richelet simplifie les consonnes doubles, et supprime les lettres adscrites et grecques. Il emploie largement les accents, aigu, grave et circonflexe. En recueillant l'héritage des écrivains et des réformateurs, les innovations de Richelet sont en accord avec son temps et son milieu.

De nombreux auteurs et éditeurs s'intéressent aux réformes et jouent un rôle non négligeable dans ce domaine. Prenons Corneille comme un exemple : en 1663, il se déclare en faveur de l'utilisation de l'accent grave, ainsi que les lettres *j* et *v*, et préconise l'emploi du *s* au lieu de *z* en tant que signe de pluriel.

Sur le terrain de l'orthographe du XVII^e siècle, la fixation devient au fur et à mesure moins aléatoire, et ainsi, le français « achève de devenir elle-même » (Catach, 2003 : 29) : les diphtongues disparaissent pratiquement et les consonnes finales ne se prononcent plus en général. Le français continue à s'ordonner et à se répandre.

1.1.1.4. Vers l'orthographe d'État

C'est finalement l'Académie française, fondée en 1635, qui est chargée de normaliser l'orthographe française par l'élaboration de son dictionnaire qui fixe l'usage du français. Les premiers académiciens, plutôt conservateurs, préconisent une orthographe plus étymologisante. La prise de position de l'Académie rend vaines toutes les options radicalement phonographiques. En 1694, la première édition de Dictionnaire paraît, près de soixante ans après la création de l'Académie française. Le Dictionnaire souffre d'une trop longue élaboration, alors que l'orthographe du français se trouve dans une époque où tous les aspects de la langue évoluent rapidement. Avec un demi-siècle de stagnation, les décisions prises en 1635, et en 1639, pour la première édition de Dictionnaire, ne sont plus conformes à l'usage de la fin du siècle. « L'Académie a un demi-siècle de retard, comme l'indique Catach, et ce retard se répercutera à chacune des éditions suivantes, jusqu'au XIX^e siècle » (2001 : 169) ; le fond du problème est que la langue, qui est fixée par l'Académie, devient « classique ».

Cependant, il ne faut pas croire, comme le laisse entendre Beaulieu, qu'après un

siècle passé, l'Académie s'en tienne encore à l'orthographe de R. Estienne. Notons les changements étonnants, si l'on compare la première édition de Dictionnaire à celui de R. Estienne ou à celui de Nicot : elle fait tout d'abord la distinction entre *j* et *v*, peu systématique, mais qui permettra « la chute de toute une série de consonnes ou de signes chargés de distinguer les deux graphèmes manquants » (Catach, 2003 : 33) ; elle élimine un grand nombre de consonnes plus ou moins « étymologiques », à la finale surtout ; elle supprime aussi quantité d'anciens hiatus ; elle n'emploie pas l'accent intérieur, et conserve partout le *s* muet, mais elle écrit *baume* (pour *bausme*), *couture* (pour *cousture*), etc. ; le plus frappant, c'est qu'elle remplace *en* par *an* dans beaucoup de mots .

La deuxième édition, également sous l'autorité de Régnier-Desmarais, se situe dans la lignée de la première. De ce fait, le fossé, déjà grand, entre l'Académie et l'opinion continue à s'élargir.

Il faudra attendre la troisième édition pour que l'Académie française fasse preuve d'un esprit novateur : avec l'entrée des philosophes à l'Académie, des réformes de l'orthographe se mettent systématiquement en place (Catach, 2003 : 37) :

- En 1740, presque tous les anciens hiatus ainsi que les consonnes diacritiques internes sont supprimés ; un nouveau système compensatoire d'accentuation est mis en place ; l'usage du yod et de l'orthographe des participes est réglementé.

- En 1762, avec la consolidation des premières réformes, l'emploi de l'accent grave est mis en place ; notons également la restriction et l'ordonnancement de l'accent circonflexe ; les lettres grecques sont simplifiées et les signes du pluriel sont réglementés.

- En 1798, la simplification des lettres grecques continue et de nombreuses variantes concernant les lettres doubles sont introduites.

A l'aide de cette série de réformes réfléchies, réalisée par d'Olivet, la troisième édition de l'Académie instaure en France une orthographe moderne, qui « se libère de l'influence latine, ou plus exactement qui opte pour une position intermédiaire, en

mêlant de façon parfois hétérogène, tradition et modernité » (Fayol & Jaffré, 2008 : 111). C'est cette orthographe qui devient l'orthographe actuelle. En 1798, la cinquième édition voit le jour, marquant ainsi la fin des initiatives importantes.

1.1.1.5. L'orthographe actuelle

La seconde moitié du XIX^e siècle est marquée par ces nombreuses disputes sur l'orthographe. Les deux courants s'opposent et la querelle prend de l'ampleur. C'est à cette époque que l'État commence à s'intéresser à la question de l'orthographe et publie en 1900 un arrêté du Journal officiel, suivi un an plus tard de la fameuse Circulaire qui propose une simplification de la syntaxe française. Cependant, malgré tous les efforts, aucune décision ferme n'est prise. Il faudra attendre jusqu'en 1952 pour qu'un projet significatif de réforme voie le jour : Aristide Beslais, nommé président d'une Commission ministérielle d'études orthographiques, publie le premier rapport, mais échoue très vite. En 1960, la deuxième Commission Beslais se crée. En dressant un « catalogue plutôt complet des points noirs de notre orthographe qui sont régulièrement repris, en totalité ou en partie, par les projets de simplification » (Fayol & Jaffré, 2008 : 113), cette commission essaie de concilier l'efficacité, la modération et la simplicité de l'orthographe. En 1990, afin d'introduire de la cohérence dans l'orthographe française, les Rectifications orthographiques sont publiées au Journal officiel. Cette fois, le projet est bien accueilli, à la fois par l'Académie et le Conseil supérieur de la langue française. A la fin de l'année 1990, malheureusement, une nouvelle vague de critiques se déclenche, dans le but de discréditer ces Rectifications.

Finalement, toutes ces réformes tentées depuis le XIX^e siècle ont régulièrement échoué, alors que l'orthographe des autres langues européennes, comme l'espagnol, l'allemand et l'italien, est judicieusement simplifiée. Par ailleurs, la démocratisation des usages, et l'avènement de l'école obligatoire mettent en évidence le fait que l'orthographe n'est plus capable à répondre aux besoins d'une société multiforme et étendue. L'orthographe du français, comme l'indiquent Fayol et Jaffré, « aux mains d'une élite masculine, s'est avérée difficile sinon impossible à réformer » (2008 : 114).

1.1.2. Les Français et leurs jugements sur l'orthographe

En rappelant comment s'est construite l'orthographe du français au fil des siècles, nous souhaitons mettre en évidence le fait que ce système graphique se trouve dans « une position médiane, en mêlant de façon parfois hétérogène tradition et modernité » (Jaffré, 2009). Les travaux linguistiques et psycholinguistique d'aujourd'hui nous permettent de considérer que l'orthographe française est l'un des systèmes graphiques les plus complexes à apprendre et à utiliser, question que nous aborderons dans la partie suivante. Néanmoins, nous vivons dans un contexte social où la demande orthographique explose. Par conséquent, la question de la simplification de l'orthographe française se pose avec acuité.

Aujourd'hui, l'orthographe française et son apprentissage suscitent tant de débats et différents chercheurs traitant l'orthographe semblent soutenir l'idée de la réforme (voir par exemple Manesse & Cogis, 2007 ; Fayol & Jaffré, 2008) : il est urgent de « donner aux citoyens une orthographe plus accessible et mieux adaptée à leurs besoins réels » (Jaffré, 2009). Dans la présente étude, nous ne résoudrons pas les problèmes posés par la réforme orthographique, mais nous souhaitons profiter des réflexions que ce débat suscite également au niveau de l'enseignement-apprentissage de l'orthographe française, à savoir, comment enseigner l'orthographe du français dans les milieux scolaires et comment réagissent les apprenants de français, confrontés à ce système graphique en quelque sorte opaque. Tout ceci est dans le but de pouvoir aborder ensuite, dans le chapitre suivant, la question de l'apprentissage de l'orthographe.

1.1.2.1. Questions de norme

Dans l'évolution historique de l'orthographe du français, on assiste à un débat toujours agité, sur la crise qui traverse la langue française : c'est la question de la norme du français. Les conceptions de la norme du français connaissent quatre périodes, depuis que son existence est reconnue officiellement par l'ordonnance de Villers-Cotterets en 1539.

La première période est définie, à la fois « par la conception très religieuse que l'on avait d'une langue (reflet divin de la pensée) et par le souvenir de la rude bataille qui est nécessaire à la reconnaissance du français, langue populaire contre le latin ». De ce fait, le français, langue nouvellement victorieuse, est pure, sacrée et intangible. La norme dans cette période est unique, et les différences sont des impuretés à rejeter. Cette opinion est affirmée par les grammairiens des XVI^e et XVII^e siècles (Malherbe, Vaugelas, etc.). La maîtrise de la langue, c'est connaître le bon usage de la langue, qui est une notion quasi sacrée, à respecter inconditionnellement.

Avec l'arrivée des travaux sur la linguistique, les conceptions de la norme du français sont un peu plus ouvertes et plus conformes à la réalité du fonctionnement social. Dès les années 50, l'idée de l'existence des variations et de la pluralité des normes apparaît. Elle est toujours liée à la notion de hiérarchie de ces normes. Autrement dit, bien que des variétés des normes existent, elles ne sont pas d'égale valeur : il existe trois niveaux de langue : bas (ou familier), courant et soutenu (littéraire, le modèle). A ce titre, l'apprentissage de la langue a pour objectif de faire connaître les niveaux de langue et faire utiliser le meilleur en toute circonstance aux locuteurs.

A partir des années 70, une pédagogie pluri-normative est adaptée dans le milieu scolaire. Selon elle, les variétés sont d'égale valeur et sont classées dans des registres, qui dépendent des situations de communication. Les locuteurs doivent connaître divers registres de langage et choisir entre eux en fonction de la situation de communication. En réalité, cette pédagogie n'a que peu de sens, puisque les choix langagiers dépendent non seulement des données objectives (statut social du destinataire, lieu de la communication, etc.), mais aussi du projet de communication, de ses enjeux et de l'effet que le locuteur veut produire.

Actuellement, on parle de pédagogie pluri-normaliste : c'est-à-dire, l'apprentissage de la langue consiste à prendre conscience des enjeux véritables d'une situation de communication, et à associer les choix langagiers au projet de communication, à l'effet à produire et au résultat attendu.

1.1.2.2. L'enseignement de l'orthographe

Avec une complexité indiscutable, la maîtrise de l'orthographe en France est un signe de distinction : une bonne orthographe « incarne le respect du bien commun, à l'égal de la politesse, du bon goût ou d'autres qualités du vivre-ensemble » (Manesse & Cogis, 2007 : 23). En revanche, son ignorance est liée aux jugements bien au-delà de l'orthographe à proprement parler. Une personne qui a une mauvaise orthographe « court le risque de n'être pas estimée fiable » (Manesse & Cogis, 2007 : 23).

En plus de l'intérêt qu'elle suscite dans l'opinion publique, l'orthographe est aussi l'affaire du milieu scolaire. Sans doute est-elle le domaine le plus emblématique de l'école et après un siècle et demi d'école obligatoire, l'enseignement de l'orthographe évolue progressivement. A la fin du XXI^e siècle, les inspirateurs de la rénovation pédagogique lancent publiquement une campagne de réforme de l'enseignement du français. De ce fait, les priorités pédagogiques sont désormais inversées. L'orthographe et la grammaire sont invitées à passer au second rang de l'enseignement, et l'accent est mis sur la compréhension des textes, car il s'agit de « fournir le jugement, le raisonnement, le cœur et même le goût de l'élève » (Chervel & Manesse, 1989 : 144).

Cependant, l'enseignement de l'orthographe et de la grammaire « ne disparaît pas des préoccupations officielles et encore moins des pratiques des maîtres » (Chervel & Manesse, 1989 : 147). Il subit toutefois une évolution décisive dans son esprit : une conception nouvelle s'impose sur l'enseignement de l'orthographe et de la grammaire. Intégré à l'ensemble de l'enseignement du français, l'apprentissage de l'orthographe et des règles de grammaire commence par les textes que l'on lit. Et la mise en orthographe change aussi de nature. Les automatismes, les mécanismes et la mémorisation, sans disparaître totalement, s'effacent progressivement derrière des pratiques plus souples (Chervel & Manesse, 1989 : 150-151). Jusqu'à la fin du XX^e siècle, l'enseignement du français n'a pas à subir de profondes modifications. Avec le prolongement de la scolarité, l'acquisition de l'orthographe s'étale sur 8 ans. Autrement dit, on retire l'exclusivité de l'enseignement de l'orthographe à l'école primaire et le collège contribue aussi à la consolidation de la connaissance de l'orthographe lexicale et

grammaticale. On assiste ainsi à une modification importante de l'enseignement du français : l'enseignement de l'orthographe et de la grammaire occupe « une place centrale, mais non plus la place unique » (Chervel & Manesse, 1989 : 153).

1.1.2.3. Les réformes de l'orthographe

L'orthographe du français demeure longtemps « Janus à deux visages » (Fayol & Jaffré, 2008 : 18), l'un tourné vers le passé, enraciné dans l'étymologie latine, l'autre qui assure la communication écrite du plus grand nombre. Et les deux courants, l'un conservateur et l'autre réformateur, continuent à s'affronter.

L'un résiste aux changements de l'orthographe, en s'appuyant sur l'idée de l'immobilisme orthographique : certains partisans sont immodérément attachés à leur écriture nationale, ce qui peut être interprété comme une sorte de « chauvinisme ». Un certain nombre d'entre eux confond l'orthographe et la langue : « la question orthographique, dit Burney, toute secondaire qu'elle semble, présente des aspects philosophiques, politiques et sociaux dont les deux parties en présence ne sont pas toujours bien conscientes : l'Académie française et la bourgeoisie traditionnaliste sont opposées dans leur ensemble à la réforme et à l'orthographe de la cuisinière... ». Parmi les arguments des conservateurs, on voit également « la crainte d'occulter la lecture de notre littérature nationale », comme l'explique Catach (2003 : 90), c'est l'attachement « aux règles poétiques de la prosodie et de la métrique, à la physionomie des mots, pour tout dire, à la tradition » qui compte. Autrement dit, les conservateurs vouent un attachement très fort, perçu d'un point de vue esthétique, à l'orthographe. Selon eux, elle est « sentie comme ayant sa beauté, son charme propre dans sa complexité même » (Cibois, 1991). Notons aussi la domination personnelle. Il n'est pas facile de pénétrer dans notre système de l'orthographe ; cependant, lorsqu'on le possède, on découvre un monde riche de sens et d'analogies. Ces partisans s'interrogent récemment sur l'ordre juridique : qui peut réformer l'orthographe ? A cette question, l'écrivain Jean d'Ormesson répond fièrement en 1989 : « Ce n'est ni le gouvernement, ni les syndicats, ni un comité, ni même l'Académie. C'est le peuple ». Il faut cependant bien noter que

la France est un pays où « le moindre écart d'orthographe populaire est durement sanctionné » et où « l'État a réglé au cours des siècles les problèmes de langue, surtout écrite, de façon strictement centralisée » (Catach, 2003 : 91).

Cependant, l'autre courant veut changer l'orthographe du français. De nombreux phonétistes, de Meigret à A. Martinet, en passant par Brunot et Cl. Blanche-Benveniste insistent sur le fait que les modifications orthographiques ne doivent pas s'arrêter en chemin. Les partisans de ce courant indiquent également les positions intermédiaires des conservateurs, en montrant qu'en pratique, même les plus conservateurs reconnaissent « la nécessité d'effectuer quelques concessions aux plaintes exprimées par l'opinion publique » (Catach, 2003 : 91). Ils oublient cependant que, en théorie, tout le monde est d'accord pour faire une réforme, mais dans la pratique, les gens hésitent : quelle réforme faudrait-il faire et par où commencerait-elle ?

Une centaine d'interviews ont été réalisées, entre 1987 et 1989, par les auteurs d'*Orthographe mon amour* (Millet, Lucci & Billiez, 1990). Ils ont réalisé une enquête auprès d'enseignants, de secrétaires, d'apprentis, de gens du livre, et d'élèves. Afin d'avoir une vision globale des jugements portés par le public, leurs questions concernaient « la crise de l'orthographe », « les souvenirs d'apprentissage », les « difficultés » et les « fautes », ainsi que le thème de la réforme. Les résultats ont été intéressants : de nombreuses personnes ont du mal à définir l'orthographe (« l'orthographe pour moi, dit l'un... c'est comme ça ») ; elles ne semblent pas hostiles à l'orthographe, au contraire, elles portent plutôt une attitude réservée, contradictoire, avec cette consolation que « personne n'est infaillible ». De nombreux enquêtés trouvent l'orthographe française belle et avouent qu'ils ont du plaisir à écrire. En ce qui concerne la réforme, certains déclarent avec lucidité qu'il faudrait simplifier le système de l'orthographe, surtout les accents, les accords des verbes pronominaux et des consonnes doubles. Cependant, ils reculent devant une application et apparemment ne semblent pas y songer. Parmi tous les interrogés, il faut bien noter l'avis des parents d'élèves. Dans l'intérêt de leurs enfants, ils sont à la fois pour et contre la réforme : une scolarité mieux réussie, oui ; une baisse de niveau, non. Ils sentent que ces exigences orthographiques démesurées et sans réels profits constituent un obstacle aux progrès de

leurs enfants. Inversement, ils refusent de renoncer à cette orthographe, considérée comme une valeur historique et culturelle, même au prix de difficultés.

Quant aux enfants, les trois-quarts des interrogés disent qu'ils vivent dans les affres au moment de la dictée : « ça m'agace », « ça fait mauvaise impression », « c'est pas que j'ai honte, mais... ça me gêne ». Les adolescents, surtout les apprentis, selon les auteurs, sont plus contestataires. Certains d'entre eux déclarent ouvertement : « c'est idiot », « c'est nul », « on se casse la tête pour rien ». Les autres, plus nombreux, trouvent l'orthographe « normale », mais celle du français est trop compliquée. Notons que c'est surtout l'apprentissage qui leur pose problème : « c'est un dédale », « c'est fait pour induire en erreur ». Cependant, les jeunes sont perplexes quand on leur parle de réforme. Selon eux, c'est trop tard pour la réapprendre. Et ils rejettent l'écriture phonétique : « c'est plus compliqué ». Le plus étonnant, c'est que ce sont les élèves en échec scolaire qui se montrent parfois les plus conservateurs : ils disent qu'ils « en ont bavé », mais plus important, ils ne savent rien de la langue, seulement quelques bribes. C'est justement à cause de cette ignorance qu'ils sont en situation d'échec. Après réflexion, les trois-quarts des jeunes interrogés acceptent les modifications de l'orthographe comme la suppression des consonnes doubles, du *h* des mots grecs, mais n'acceptent pas, par exemple, les atteintes grammaticales, comme suppression du *s* du pluriel. Comme les grands, les jeunes pensent que c'est inutile d'y songer. L'orthographe, pour eux, c'est trop surveillé, trop important, et donc impossible. Comme le soulignent les auteurs, en conclusion : « La confusion entre orthographe et langue apparaît comme le nœud du processus de fétichisation de l'orthographe ».

En 1989, la revue *Lire* lance également une enquête publique sur un échantillon national concernant le problème de la réforme orthographique. On trouve la présence d'un socle commun entre les favorables et les hostiles à l'orthographe (Cobois, 1991) : l'orthographe fait partie de notre patrimoine (86% le pense dans l'échantillon national), elle est difficile (73%) mais sans être un casse-tête inutile (83%), en matière de réforme on pourrait envisager d'en retoucher quelques bizarreries (76%). On peut s'apercevoir que, en respectant l'orthographe, les Français prennent en même temps conscience de ses difficultés et des améliorations marginales.

Cobois note également que l'on est hostile à la « réforme » mais pas à des « modifications ». Selon lui, c'est parce que l'idée de réforme est associée à une réforme radicale, phonétique, utopique, impossible, même quand on la souhaite. De ce fait, on comprend mieux que le Conseil supérieur de la langue française publie au JO en 1990 un rapport, qui abandonne le terme de « réforme » et porte dans son titre le terme de « rectifications ».

Afin de juger de ces modifications proposées par le Conseil, une enquête a été faite en 1990 auprès d'une centaine d'étudiants. Malgré l'existence d'une opposition entre l'orthographe d'enracinement et l'orthographe fonctionnelle, pour la majorité des répondants, la clé pour résoudre le problème de l'orthographe est d'articuler ensemble l'enracinement et la fonctionnalité, c'est-à-dire qu'il faut respecter les racines culturelles de l'orthographe et à la fois, sans étouffer la nécessaire adaptation des outils culturels au présent (Cobois, 1991).

En concluant, l'orthographe du français est « le produit d'une histoire séculaire...qui en codifie les principes, au terme d'une bataille remportée par le camp de la tradition et du conservatisme » (Jaffré, 2009). Au fil des siècles, les événements linguistiques, que nous avons évoqués auparavant, ont une forte influence sur la phonétique et la morphologie du français, puis dans le même mouvement son orthographe. La langue française est par conséquent isolée des autres langues romanes et son orthographe se caractérise par « une grammaire écrite qui n'est pas relayée par la phonographie » (Jaffré, 2009), dont la présence rend ce système graphique compliqué à apprendre et à utiliser. En outre, avec la révolution électronique, les besoins de la communication écrite augmentent énormément et la place de l'écrit est en train de se transformer à l'époque actuelle. Parallèlement avec les besoins de scripteurs, la question de la simplification de l'orthographe française se pose avec acuité et l'orthographe, ainsi que son enseignement-apprentissage, suscitent tant de débats aujourd'hui. Afin de comprendre les problèmes posés par l'apprentissage de l'orthographe française, il nous semble indispensable de mettre en évidence, dans la partie suivante, les caractéristiques de ce système graphique.

1.2. Aspect linguistique : les spécificités de l'orthographe française

Toute écriture entretient une relation étroite avec la langue et se caractérise généralement par deux principes majeurs, qualifiés respectivement de phonographique et de sémiographique. Relevant d'une sphère alphabétique, l'orthographe française dispose d'une solide base phonographique. Cependant, pour des raisons historiques, comme nous l'avons évoqué auparavant, l'orthographe du français est moins régulière que celle des autres langues romanes venues du latin. Mais elle n'en reste pas moins rentable, parce qu'elle est également dotée de l'une des plus complexes sémiographies qui soient. Nous essaierons donc de décrire ce système linguistique mixte en présentant respectivement ces deux dimensions : phonographique et sémiographique.

1.2.1. Un système mixte

1.2.1.1. Présentation générale du plurisystème

Il est bien connu que l'orthographe française se présente sous la forme d'une collection de règles et d'exceptions. Les études conçues par Catach et son équipe de HESO (Catach, 1980) font, cependant, une description du système linguistique du français en mettant l'accent sur la cohérence, c'est-à-dire en montrant que, malgré sa complexité, le système graphique français est bien structuré et fait l'objet d'une description raisonnée.

Cette approche scientifique met en évidence trois éléments très importants : le phonème, le morphème et le lexème (mot), et ils entrent eux-mêmes en combinaison syntaxique (la phrase). Le système orthographique français, constitué essentiellement de ces trois catégories, est donc un plurisystème ou un ensemble de sous-systèmes (Catach, 2003 : 53-66) :

Le centre de ce plurisystème est constitué des phonogrammes, c'est-à-dire des graphèmes qui correspondent directement aux phonèmes. Cette zone du système comprend elle-même un noyau essentiel, les archigraphèmes ainsi qu'une partie annexe,

les graphèmes positionnels, dont nous parlerons en détail plus loin. Ensuite sera abordée la zone des morphogrammes : désinences, flexions verbales, préfixes, suffixes, etc. Prononcés ou non, ces morphogrammes sont considérés comme des marques de série ou de sens. La troisième zone est constituée par des logogrammes, qui jouent un rôle annexe. Les logogrammes servent dans le système à distinguer les homonymes, qui sont en petit nombre par rapport à tous les lexèmes qui existent dans le lexique et le discours, mais qui troublent considérablement les mécanismes d'interprétation. Enfin, la dernière zone est celle des lettres étymologiques ou historiques. Si le critère fonctionnel ne permet pas de justifier de toutes les lettres utilisées dans le système graphique du français, c'est à cause de l'existence de ce reliquat. Ainsi, il est à noter le maintien des lettres grecques dans les mots comme *théâtre*, *rhume*, ou des lettres latines comme le *p* de *prompt* et de *sculpteur*, ainsi que les consonnes doubles qui font appel soit à la connaissance du latin (*appeler*, *effarer*, *offenser*), soit à celle de l'histoire de la langue (dans les mots *sonner*, *donner*, *honneur*, etc., les doubles consonnes notent l'ancienne nasalisation).

Le premier sous-système (phonogrammique) est au centre de tout le système et est relativement stable ; le deuxième (morphogrammique) couvre d'assez larges zones de la chaîne écrite ; le troisième (logogrammique) ainsi que les lettres étymologiques et historiques constituent une zone partielle et marginale.

1.2.1.2. La phonographie : représenter les sons de la langue

1.2.1.2.1. Les sons du français

Les sons produits par la voix sont en nombre infini, comme ceux de l'enfant dans son babil. Cependant, au fur et à mesure qu'il apprend à parler, l'homme devient incapable d'articuler les sons qui n'appartiennent pas à son système linguistique et il perd ainsi une « aptitude naturelle » au profit d'une « aptitude culturelle » (Jakobson, 1963). Les sons produits par la voix n'ont aucune valeur linguistique, mais dès lors qu'ils sont sélectionnés par une langue, ils acquièrent une valeur phonétique, et deviennent ce que l'on appelle le phonème.

Le phonème est la plus petite unité de signifiant phonique dont la variation modifie la valeur du signe à l'oral. La langue française standard dispose de 36 phonèmes : 17 consonnes, 16 voyelles et 3 semi-consonnes. Ce chiffre peut varier selon les régions et l'âge des locuteurs.

1.2.1.2.2. L'écriture des phonèmes : principes généraux

Si le phonème se manifeste à l'orale, le graphème quant à lui se manifeste à l'écrit. Le graphème est donc la plus petite unité de signifiant graphique destiné à transcrire le phonème. Avec des correspondances entre phonèmes et graphèmes, l'orthographe pourrait se structurer et permettre aussi à la compétence orthographique de se construire. Quant aux phonogrammes, ils sont les graphèmes qui correspondent directement aux phonèmes. Considérés comme le centre du plurisystème du graphique français, les phonogrammes comprennent un noyau d'archigraphèmes. Comme les graphèmes de base du français, les archigraphèmes sont les « formes maximales stables » (Catach, 2003 : 62), c'est-à-dire que parmi tous les phonogrammes, leur occurrence est la plus élevée, et on les trouve dans toutes positions, en toute place du mot. Par exemple, les phonogrammes *e*, *é*, *è* et *ê* ont comme archigraphème le *e* ; et le *e*, lui, a les *é*, *è* et *ê* comme graphèmes positionnels. Ceci indique les valeurs diverses que l'archigraphème *e* peut prendre en telle ou telle position. Il y a 33 archigraphèmes en français, qui sont A – E – I – O – U – EU – OU – AN – IN – ON – UN – ILL – Y – OI – OIN – P – B – T – D – C – G – F – V – S – Z – X – CH – J – L – R – M – N – GN.

1.2.1.2.3. L'écriture des phonèmes : problèmes

En français, la complexité de l'écriture des phonèmes s'explique par le fait que la correspondance entre les phonèmes et les graphèmes n'est pas biunivoque, c'est-à-dire que comme le nombre de phonèmes et celui des graphèmes n'est pas identique, leurs relations manifestent certaines distorsions. Dans le cas particulier des phonogrammes, la difficulté vient du fait qu'un phonème dispose de phonogrammes concurrents entre lesquels le scripteur doit choisir. Ce phénomène linguistique tient essentiellement, comme l'expliquent Fayol et Jaffré, à la genèse des orthographes : « avec le temps,

certaines distinctions phoniques peuvent disparaître, en créant un surplus graphique » (2008 : 86). C'est ainsi qu'est née la polyvalence graphique, notion sur laquelle nous reviendrons plus loin. En français par exemple, *o*, *au* et *eau* sont des phonogrammes concurrents, qui jouent des rôles différents. Et pour sélectionner le bon phonogramme, le scripteur doit avoir recours à divers procédés de transcription.

Montonen (2001) montre que le français dispose de trois types de notations pour transcrire les sons français : le signe simple, comme le *p* et le *a* dans le mot *papa* ; le signe complexe, par exemple, dans les mots *feuille*, *feuillage* et *feuille*, le digramme *eu*, le trigramme *ill* et le quadrigramme *ille*, étant tous trois des signes complexes ; le signe accompagné d'un élément auxiliaire, c'est-à-dire accent, cédille ou lettre diacritique, comme dans les mots *à*, *ça* et *cette*. Ces éléments auxiliaires servent à unir ou disjoindre les graphèmes complexes. Par exemple, dans le mot *algue*, le *gu* est un graphème complexe, mais dans le mot *aiguë*, *gu* correspond à deux graphèmes. Pour bien transcrire ces graphèmes, des lois de position doivent être prises en compte. Prenons l'exemple des mots *égaler* et *effort* : on écrit *égaler* avec accent pour préciser la prononciation du *e*, alors que dans le mot *effort*, les lettres qui suivent le *e* indiquent la prononciation et rendent l'accent inutile. Parmi les lois de position, les lois d'accentuation sont les plus régulières, alors que les autres lois sont, de natures diverses, soit inhérentes au système, soit aléatoires, produits du temps et de l'usage.

La sélection des phonogrammes ne répond pas seulement au niveau linguistique, mais aussi au lexique. Si l'on reprend l'exemple de *o*, *au* et *eau*, on voit que *o*, comme l'archigraphème, est la forme la plus neutre : le timbre est soit ouvert, comme dans le mot *vol*, soit fermé, comme dans le mot *voleur* ; et il est également présent dans toutes les positions du mot, *oreille*, *doré*, *métro*, etc. Quant aux phonogrammes *au* et *eau*, ils sont considérés comme des formes marquées, car ils apportent un surcroît d'information. La lettre *a* dans *au* et *eau* a généralement une cause étymologique : « elle rappelle un ancien *a* qui se retrouve dans des mots appartenant à une même famille et qui partagent une même racine » (Fayol & Jaffré, 2008 : 116). Ainsi, plus la polyvalence graphique est élevée, plus la sélection d'un bon phonogramme implique un traitement non phonographique, qui répond à une compétence orthographique supérieure chez le

scripteur.

1.2.1.3. La sémiographie : donner accès au sens

Comme nous l'avons indiqué auparavant, la phonographie fournit les mécanismes de base au bon fonctionnement de l'orthographe. Cependant, l'orthographe ne se résume pas à une simple notation des sons du langage et elle représente également le sens linguistique. Cette mécanique sert à construire un autre principe de l'écriture, le principe sémiographique, qui est tout aussi essentiel que le principe phonographique. Pour se doter d'un sens linguistique, l'orthographe doit reconfigurer les mécanismes phonographiques afin de leur donner une forme fonctionnelle. Il nous semble nécessaire d'introduire ici une notion importante, le potentiel sémiographique des langues : « La raison d'être de l'orthographe est de donner à voir le sens linguistique, la base phonographique s'y prête plus au moins bien selon qu'elle rend visible ou non les unités lexicales et grammaticales d'une langue » (Fayol & Jaffré, 2008 : 94). De ce fait, les orthographes, comme celle du turc ou du coréen, qui transparaissent essentiellement dans la phonographie, disposent d'un potentiel sémiographique très élevé. Cependant, pour certaines langues qui présentent une homophonie importante, source majeure d'une ambiguïté graphique, la base phonographique seule ne permet plus d'assurer une bonne représentation sémiographique. En l'occurrence, l'orthographe doit avoir recours aux procédés complémentaires, par exemple l'ajout d'éléments spécifiques : lettres silencieuses des orthographes alphabétiques ou éléments sémantiques de l'orthographe du chinois.

La sémiographie n'est donc pas homogène et présente différentes facettes : elle est située sur un plan qui « englobe l'écriture d'unités lexicales, ou logographie, et l'écriture d'unités grammaticales, ou morphologie » (Jaffré, 2002). Les formes du principe sémiographique repose sur deux tendances majeures :

« La première, la moins complexe, se limite à la réanalyse des phonogrammes, qui sont associées pour constituer des entités plus larges,

sans que des adjonctions soient nécessaires (sémiographies mineures). La deuxième tendance s'efforce en revanche de compenser les insuffisances sémiographiques de la base phonographique, et de ses formes homophones, en ajoutant les éléments spécifiquement graphiques (sémiographies majeures) » (Fayol & Jaffré, 2008 : 97).

Dans le cas du français, l'orthographe n'échappe, comme toutes les autres orthographe, n'échappe pas à une réanalyse sémiographique de la phonographie. Si l'on prend l'exemple du mot *nouveau*, on voit qu'un trigramme *eau* correspond, d'un point de vue phonographique, au phonème /o/. Mais d'un point de vue morphologique, « une relation de type *nouveau/nouvelle* révèle une nouvelle fonction de ce trigramme, qui motive le *e* de *eau* » (Jaffré, 2003). On constate ainsi l'existence, dans l'orthographe française, de phonogrammes à fonction morphologique, que nous appelons morphogrammes. Ces unités très accessibles constituent les représentations lexicales mais aussi l'interface entre phonographie et sémiographie. A ce titre, elles bénéficient « d'un double traitement potentiel puisque leur présence peut être détectée par la mise en relation avec un phonème ou par une analyse morphologique » (Jaffré, 2003). Mais la sémiographie peut se manifester d'une autre façon plus directe, en utilisant des marques qui échappent à la phonologie. C'est sur ce versant des représentations grammaticales que les choses se compliquent, concernant principalement les morphogrammes, notion que nous développerons plus loin.

1.2.1.4. La morphographie : signaler les rapports entre lexèmes

La morphographie du français est l'aspect orthographique qui correspond à la morphologie écrite. Théoriquement, toute écriture comprend « des aspects qui s'écartent d'une stricte correspondance avec des unités phoniques pour prendre en compte les signes linguistiques » (Jaffré, 2003). Et le français offre à cette dimension sémiographique un exemple tout à fait différent.

Comme nous l'avons abordé auparavant, aucune écriture n'échappe à une

réanalyse sémiographique de la phonographie. Et la discordance entre le nombre de phonèmes et de graphèmes du système de la langue française conduit à une polyvalence graphique motivée par la morphologie. Cependant Fayol et Jaffré montrent que « la complexité sémiographique va bien au-delà de la polyvalence phonographique pour recourir à des graphies propres qui peuvent être des désinences verbales (le *t* de *il sait*), des marques de nombre (le *s* de *deux maisons*), ou de genre (le *e* de *amie*) ; des marques d'accord (*les petits chats jouent*), etc. » (2008 : 116). On voit ainsi que la morphologie écrite du français contient des unités qui échappent à toute phonographie et qui jouent très souvent un rôle grammatical essentiel. Bien que ces morphogrammes ne jouissent d'aucune autonomie sémiographique, ils deviennent des unités à part entière dans le système orthographique. Si l'on prend l'exemple de l'ensemble des désinences verbales, elles concernent, le plus souvent, les vestiges de l'ancien système graphique qui sont disparus aujourd'hui : le *t* de *il veut*, le *s* de *tu parles*, etc. sont des éléments qui font double emploi avec les pronoms *il*, *tu*, etc. On peut citer également l'exemple du *s* du pluriel dans *les petits chats* ; ce morphogramme, dépourvu de correspondances phoniques, signale les rapports entre les trois lexèmes.

1.2.2. Un système qui est source de difficultés pour les scripteurs

1.2.2.1. La polyvalence du français : un compromis entre sémiographie et phonographie

La phonographie a pour finalité de « fournir aux orthographes un matériau de base systématique, économique et récurrent, à partir duquel elles façonnent des signes linguistiques conformes aux exigences sémiologiques de l'écrit » (Fayol & Jaffré, 2008 :100). Pour bien comprendre le compromis entre sémiographie et phonographie de l'orthographe, il nous semble nécessaire d'expliquer une notion importante, qui est la polyvalence graphique. Étant donné que le nombre de graphèmes est supérieur à celui des phonèmes, la quasi-totalité des orthographes alphabétiques présente plus ou moins une polyvalence graphique. Les psycholinguistes se sont particulièrement intéressés à

ce phénomène, en distinguant des orthographe transparentes (finnois, grec, italien, espagnol, allemand) et des orthographe opaques (portugais, français, norvégien, suédois, néerlandais, danois et anglais).

Dans le cas particulier du français, la polyvalence graphique contribue, d'un côté, à complexifier la phonographie, en obligeant le scripteur à faire une sélection parmi les phonogrammes concurrents et d'un autre côté, elle se trouve investie de fonctions morphologiques ou distinctives, c'est-à-dire qu'elle « constitue un avantage sémiographique en fournissant un matériau visuellement fonctionnel » (Fayol & Jaffré, 2008 : 97). Cependant, dans le système du français, la structure phonologique n'arrive pas à rendre compte de la totalité des variations lexicales et morphologiques. L'orthographe française doit donc inventer des procédés spécifiquement graphiques pour tenter de compenser les insuffisances phonologiques. Le plus souvent, ces procédés se présentent sous des formes diverses qui sont tributaires de l'histoire de l'orthographe, ce qui constitue une source majeure des difficultés du scripteur. C'est ce qui caractérise l'orthographe française, un des systèmes sémiologiques les plus spécifiques et les plus opaques.

1.2.2.2. L'opacité du français

Pour les scripteurs, la difficulté d'acquisition de l'orthographe sera fonction du degré d'opacité ou de transparence de la langue, qui représente le degré de correspondance entre l'orthographe et la phonologie de cette langue, c'est-à-dire entre la manière dont elle s'écrit et la manière dont elle se prononce. Les orthographe qui ont plus de chances d'être transparentes sont celles qui sont proches du modèle idéal où la correspondance entre graphèmes et phonèmes est univoque, ou qui sont récentes. Mentionnons ici le cas du turc, dont l'orthographe est transparente grâce à la notation systématique des variantes morphologiques.

Le français, en revanche, se situe parmi les systèmes les plus opaques en orthographe. Si les correspondances graphèmes-phonèmes sont régulières et fiables à environ 95% (dans le sens de la lecture), elles ne le sont qu'à moins de 70% dans le

sens phonèmes-graphèmes (dans le sens de l'écriture) (Peereman & Content, 1999 ; Ziegler, Jacobs & Stone, 1996). Véronis (1988) a estimé que, à partir de simulations basées sur les correspondances phonèmes-graphèmes, seulement 52.7% des formes graphiques du français pouvaient être prédites à partir de leur forme phonologique.

En général, l'opacité du français est liée à plusieurs difficultés de transcription : tout d'abord, c'est l'existence des allographes, c'est-à-dire des graphèmes différents pour transcrire un même phonème. Nous reviendrons sur ce point plus loin. Deuxièmement, le français est difficile à transcrire parce qu'il existe de nombreuses consonnes doubles et le plus souvent, ces consonnes sont absentes au niveau phonologique, comme dans le mot *flamme*. Signalons également l'existence non négligeable de nombreux homophones, hétérographes, comme *sain* et *saint*, ou homographes, comme *rame* de bateau, de métro et de papier. Une autre source d'ambiguïté du système orthographique français se situe au niveau morphosyntaxique, comme les marques du nombre, qui sont le plus souvent, sans contrepartie phonologique.

1.2.2.3. Le degré de régularité

Il n'existe aucune communauté linguistique dans laquelle l'apprentissage de l'orthographe repose sur la mémorisation pure et simple des signes graphiques. Pour être facile à apprendre, toute orthographe doit disposer d'une base phonographique avec des correspondances systématiques entre phonèmes et graphèmes. Généralement, le principe phonographique se manifeste sous différentes formes selon les langues. La différence la plus notable concerne le degré de régularité. Et comme il existe une correspondance systématique entre les unités phoniques et les unités graphiques, la régularité sera « d'autant plus grande que le nombre respectif de ces unités sera proche » (Fayol & Jaffré, 2008 : 86). Une écriture serait absolument régulière si chacun des phonèmes dont elle disposait correspondait à une graphie propre, lettre ou caractère, et si cette graphie était différente de toutes les autres.

Cependant, pour des raisons historiques, ce cas idéal ne se présente quasiment

jamais et à ce titre, la relation entre ces deux unités présente donc certaines distorsions. Lorsque le nombre de phonèmes est supérieur à celui des graphèmes, cette écriture serait, théoriquement, plus facile à encoder qu'à décoder, c'est-à-dire qu'elle serait plus facile à écrire qu'à lire. Il en va autrement pour l'écriture ayant à sa disposition un nombre de graphèmes supérieur à celui des phonèmes. Le décodage serait alors plus aisé que l'encodage, et la lecture serait plus facile que l'écriture. C'est en général ce qui se passe avec les langues alphabétiques. Pour des raisons qui tiennent à la genèse et à l'évolution historique des orthographe, certaines distinctions phoniques peuvent disparaître et un surplus graphique est ainsi créé.

La langue française standard dispose d'une trentaine de phonèmes et de plus de cent graphèmes. Cette distorsion entre le nombre de phonèmes et de graphèmes est révélatrice de la complexité de la phonographie du français et elle se fait au détriment du scripteur, en l'obligeant à sélectionner le bon phonogramme entre plusieurs phonogrammes concurrents. Comme ces phonogrammes concurrents jouent des rôles distincts, la sélection demande au scripteur des traitements supérieurs non phonographiques.

L'équipe HESO a réalisé un sondage statistique en 1972 pour avoir une évaluation chiffrée des phonogrammes du français. Tous les phonogrammes ont été relevés et classés à partir des cinq textes. On a alors observé la prédominance quasi absolue d'un certain nombre d'unités. Pour les voyelles, on parvenait aux chiffres suivants (Catach, 2003 : 63) :

- A : 92% des transcriptions des phonèmes [a] ;
- E : 78% des transcriptions du [e] ; 61% des transcriptions du [ɛ] ;
- I : 98% des transcriptions du [i] ;
- O : 75% des transcriptions du [o] ou du [ɔ] ;
- U : 100% des transcriptions du [y] ;
- EU : 93% des transcriptions du [ø] ou du [œ] ;
- OU : 98% des transcriptions du [u] ;
- ON : 92.8% des transcriptions du [ɔ̃] ;

- UN : 97% des transcriptions du [œ̃], etc.

Dans l'ensemble des textes, le rendement des phonogrammes atteint 82,51%. En revanche, il existe indubitablement des difficultés concernant quelques phonèmes. Si l'on prend l'exemple du phonème [ã], il est noté indifféremment par les graphèmes *an* et *en*, et [ɛ̃], noté par *in* (44%), mais aussi par *ain* (21%) et *en* (23%). Mais avec les lois de position, on pourrait essayer de dissiper l'ambiguïté de ce dernier graphème, lorsqu'il transcrit [ɛ̃], il est le plus souvent précédé de *i*, *e*, *y*, comme dans *chien*, *lycéen*, *moyen*.

Cependant, la correspondance est rompue ici dans les deux sens dès lors que plusieurs phonèmes sont transcrits de façon semblable, et que le même graphème correspond à plusieurs phonèmes, ou au phonème zéro (Catach, 2003 : 64).

Notons que le réseau des graphèmes, même pour cette partie centrale de l'orthographe qui transcrit les sons, est complexe. Pour les scripteurs, il faudrait connaître non seulement les principales représentations graphiques des phonèmes, mais aussi les lois qui régissent leur emploi. De plus, l'habitude de tenir compte des lettres qui précèdent et suivent le graphème, de sa position dans la syllabe, dans le mot, dans le groupe de mots, etc. est également nécessaire.

1.3. Aspects psycholinguistiques : orthographier le français

Après avoir décrit le système orthographique français et son évolution, il nous semble crucial de poser une question sur la gestion des connaissances orthographiques : comment orthographions-nous ?

1.3.1. La place de la composante orthographique dans l'activité de production de texte

Comme « toute connaissance ne vaut que par les réponses qu'elle produit » (Ducard, Honvault & Jaffré, 1995 : 97), l'orthographe a pour objectif majeur de donner forme

graphique à un message. En l'occurrence, la production écrite implique le contrôle de cette forme. La compétence orthographique fonctionne donc dans la production écrite. Comme le soulignent Allal et ses collaborateurs, la situation de production écrite sert « de point de départ et de cadre à la construction et à l'exploitation de compétences orthographiques en cours d'écriture et lors de la révision différée du texte » (Allal, Rouiller, Saada-Robert & Wegmuller, 1999).

1.3.1.1. L'activité de production de texte : présentation générale

Selon les modèles classiques de la production textuelle, il existe un découpage linéaire du processus d'écriture en phases distinctes : planification, rédaction, révision (voir par exemple, Murray, 1978). Cependant, pour les modèles hiérarchiques, proposés par Hayes et Flower en 1980, la révision est plutôt une opération ou un ensemble d'opérations susceptibles d'intervenir à n'importe quelle étape de la production d'un texte. Witte souligne également que la révision n'intervient pas uniquement dans les transformations apportées lors de la relecture d'un brouillon (révision différée), mais aussi pendant la planification du texte (révision prétextuelle), ainsi que tout au long de la rédaction du brouillon (révision on-line intégrée à l'écriture) (1985). Sur la base des modèles hiérarchiques, les chercheurs étudient les opérations intervenant dans la révision (voir par exemple, Bereiter & Scardamalia, 1987 ; Hayes, Flower, Schriver, Stratman & Carey, 1987) :

1. La détection d'éléments textuels qui « posent problème » ;
2. L'identification (ou le diagnostic) de la nature du problème ;
3. Les démarches résultant des opérations précédentes, à savoir :
 - la décision de ne rien modifier
 - la modification du texte existant (corrections d'erreurs et changements visant une légère amélioration)
 - la réécriture (impliquant des transformations/changements plus profondes/profonds du texte)

Généralement, les travaux de révision peuvent porter sur toutes les composantes d'un texte : super- et macrostructure, contenu sémantique, éléments lexicaux, structures syntaxiques, organisateurs textuels, aspects de mise en texte (linéarisation), y compris transcription orthographique (Allal, Rouiller, Saada-Robert & Wegmuller, 1999). Selon Allal *et al.*, la révision prétextuelle met le plus souvent l'accent sur les composantes notamment sémantique et macrostructurale. En l'occurrence, une révision prétextuelle au niveau orthographique est une réflexion du métalangage, telle qu'une modification du champ sémantique, ou une décision de changer l'entrée dans le texte afin d'éviter certaines difficultés orthographiques. Dans la révision on-line, l'orthographe est un élément crucial dont il faut tenir compte : le scripteur relit constamment ce qu'il est en train d'écrire et y apporte des corrections. Et la révision différée a pour objectif d'assurer la conformité de l'orthographe aux conventions de la langue écrite.

1.3.1.2. De l'orthographe des mots isolés à l'orthographe des mots en contexte

Pour mieux comprendre le fonctionnement de la compétence orthographique, il nous semble utile de disposer au préalable de certaines informations sur l'architecture mentale. S'interrogeant sur le mode d'organisation des informations, certaines approches sont localisatrices. Elles regroupent les connaissances selon leur degré de spécialisation et font de la structure mentale une juxtaposition de domaines autonomes et la connaissance humaine est ainsi, organisée en modules ; d'autres optent pour l'interactivité et privilégient les relations entre savoirs, car toute connaissance implique des processus multiples et les associations de ces processus sont organisées d'une façon structurale, ou cognitive, ou encore métacognitive, et elles sont centrées à la fois sur le contenu des savoirs spécialisés et la gestion mentale de ces savoirs. L'activité mentale de l'adulte peut se décrire comme « la combinaison de modules spécialisés et de réseaux qui relient ces modules » (Ducard, Honvault & Jaffré, 1995 : 100).

Il s'ensuit que le cerveau n'est pas seulement une machine à emmagasiner des

connaissances et qu'il permet aussi à l'homme de réaliser la gestion des connaissances. Pendant l'apprentissage, si l'on veut acquérir une certaine compétence, il est indispensable d'associer l'étude de la structure du savoir et les activités qui mettent en œuvre cette structure. Au niveau orthographique, il faut donc bien distinguer les savoirs orthographiques et leur utilisation. Étant donné que l'automatisation d'un savoir implique que l'on le mette au service d'un but qui le dépasse, disposer des connaissances orthographiques ne vise pas à orthographier un mot isolé, mais à orthographier des mots dans leur contexte. C'est-à-dire que les connaissances orthographiques sont au service d'une production écrite. A ce titre, les déficits dans la production graphique ne correspondent pas nécessairement aux déficits des connaissances orthographiques et ils correspondent aussi aux déficits des compétences orthographiques. Et la compétence orthographique à son tour repose à la fois sur l'automatisation et sur l'efficacité des savoirs linguistiques et la capacité à gérer ces savoirs.

Comme Jaffré (1995) le décrit, la gestion des savoirs combine deux processus : l'activation et le contrôle. Dès qu'un besoin se fait sentir, l'activation, processus complexe, permet d'aller chercher l'information, à l'oral comme à l'écrit. Et l'évocation d'un mot donné entraîne « celle de mots ayant avec le précédent un rapport analogique que les similitudes soient phoniques, graphiques, sémantiques, etc. » (Ducard, Honvault & Jaffré, 1995 : 100). Après, une phase de contrôle sera donc nécessaire pour faire le choix le plus approprié parmi toutes les réponses concurrentes que l'activation fait émerger. Dans la dimension orthographique, le processus de contrôle fait le bon choix par l'analyse d'un ensemble d'informations linguistiques. C'est pourquoi l'accord graphique d'un mot dépend de « son appartenance à une classe morphologique et de sa distribution syntaxique » (Ducard, Honvault & Jaffré, 1995 : 101).

1.3.2. Procédures et stratégies pour orthographier

Comment procédons-nous lorsque nous devons transcrire un mot, en passant de sa

forme orale à sa production écrite correcte ? Il existe deux grandes catégories de circonstances où les situations de transcription sont analysées intuitivement : premièrement, s'il s'agit d'orthographier un mot connu et déjà rencontré à plusieurs reprises (en lecture ou en écriture), le scripteur est capable de récupérer la forme orthographique de ce mot directement en mémoire. D'une manière générale, une telle transcription se fait sans erreur¹. Notons que « la connaissance ainsi mobilisée est spécifique au mot et ne se généralise donc pas » (Fayol & Jaffré, 2008 : 138). Deuxièmement, le scripteur entend une forme phonologique nouvelle. En l'occurrence, la solution la plus usuelle pour le scripteur est de conserver temporairement cette forme phonologique en mémoire. Ensuite, il s'agit de décomposer la forme phonologique en segments et de faire la transcription segment par segment, en donnant à chacun une lettre ou un graphème, sans perdre de vue la forme phonologique d'ensemble ; autrement dit le scripteur transcrit le mot par association entre phonèmes et graphèmes.

Au cours des trois dernières décennies, les neuropsychologues ont analysé de nombreux cas de patients cérébrolésés et ont proposé deux modèles concernant le traitement des orthographes alphabétiques. D'après le premier modèle, dit voie d'adressage ou voie lexicale, les items à lire ou à écrire seraient « directement récupérés dans un lexique orthographique, sorte de dictionnaire mental qui stockerait les mots connus, réguliers ou irréguliers » (Fayol & Jaffré, 2008 : 143). Pour ce qui est de l'autre voie, dite phonologique, voie indirecte ou encore voie d'assemblage, les items feraient l'objet de conversion entre phonèmes et graphèmes.

1.3.2.1. L'accès lexical

Dans la mesure où le lien entre la forme du mot et sa signification est arbitraire, la nécessité d'un dictionnaire mental ou lexique mental, qui correspond à « l'ensemble des mots déjà connus et susceptibles d'être reconnus ou produits » (Fayol & Jaffré, 2008 : 143) semble être évidente ; mais comment est-il organisé et construit ?

¹ Lorsque le scripteur commet une erreur, il est en mesure de la repérer rapidement et la corriger lui-même.

D'énormes recherches ont été conduites ces dernières décennies afin de déterminer les caractéristiques de ce lexique mental. Segui Juan (1992) montre que l'existence de certaines capacités élémentaires chez les locuteurs d'une langue permettent d'affirmer que l'organisation du lexique mental est multidimensionnelle. Par exemple, un locuteur français peut accéder à son lexique et le consulter à partir de différentes informations linguistiques.

A partir de l'observation des patients cérébrolésés qui souffrent d'atteintes spécifiques de l'orthographe, on émet l'hypothèse de l'existence de quatre lexiques (Fayol & Jaffré, 2008 : 144) : le lexique phonologique d'entrée (en audition), le lexique phonologique de sortie (en production de la parole), le lexique orthographe d'entrée (en lecture), le lexique orthographe de sortie (en production orthographique). Maintenant, dès lors que les troubles de lecture sont généralement associés à des troubles d'écriture qui présentent de fortes similitudes, il est pertinent de supposer qu'il existe deux lexiques, phonologique et orthographique, ou un seul lexique qui fusionne les différents types d'informations relatives aux mots (Fayol & Jaffré, 2008 : 144). Néanmoins, certains modèles, dits connexionnistes, défendent l'idée qu'il n'existe pas de lexique mental : le traitement lexical repose sur les « associations plus ou moins fréquentes entre des séquences sous-lexicales, suites de sons ou de lettres de taille inférieure au mot » (Fayol & Jaffré, 2008 : 145).

Comment départager ces différentes idées concernant le lexique mental ? À travers l'observation des patients de dysgraphie phonologique, qui sont capables de lire ou d'écrire des mots connus mais pas des mots qu'ils n'ont jamais rencontrés, il est facile de considérer l'existence d'au moins un lexique mental dans lequel les items seraient stockés afin que ces patients puissent les récupérer directement. L'hypothèse qu'il existe un seul lexique permet difficilement d'expliquer le fait qu'un grand nombre d'individus sont capables de lire de nombreux mots qu'ils ne parviennent pas à orthographier correctement. En revanche, cette hypothèse de lexique unique a l'avantage de rendre compte de ce que la lecture de mots orthographiés de façon erronée, conduit à la survenue d'erreurs, même chez les lecteurs qui sont bons en orthographe. L'idée qu'il n'existe pas de lexique mental (ni unique, ni multiple) est également

plausible si l'on prend en compte de la fragilité des représentations orthographiques stockées en mémoire. Plusieurs recherches montrent que des adultes voient « leurs performances orthographiques influencées positivement ou négativement par la présentation de formes orthographiques respectivement correctes ou erronées » (Fayol & Jaffré, 2008 : 146). Cependant, dans l'état actuel de nos connaissances, il reste bien difficile de trancher entre ces différentes théories du lexique mental et des recherches supplémentaires devraient être entreprises.

1.3.2.2. L'accès phonologique

Lorsque les items à orthographier ne peuvent être récupérés directement en mémoire, il faut recourir aux conversions entre phonèmes et graphèmes. C'est ce que l'on appelle un déchiffrage pour la lecture et une transcription phonétique pour l'écriture. Bien que cette voie phonologique soit toujours possible, elle présente des difficultés spécifiques au niveau de son fonctionnement. C'est une procédure coûteuse en attention puisque les formes phonologiques à transcrire doivent être temporairement stockées dans la mémoire tout au long de la transcription. Soulignons aussi que ces mémoires pourraient « se trouver spécifiquement affectées par des troubles mais aussi présenter des difficultés spécifiques de fonctionnement chez certains individus » (Fayol & Jaffré, 2008 : 147). Par exemple, certains individus éprouvent des difficultés à maintenir en mémoire temporaire des séquences phonologiques longues.

Il existe trois types d'opérations que subissent les entités stockées temporairement en mémoire (Fayol & Jaffré, 2008 : 148-149) : premièrement, l'item d'entrée devrait être segmenté en unités, et cette segmentation pourrait être syllabique, phonémique, voire en unités moins familières comme l'attaque (début d'une syllabe) ou la rime (fin d'une syllabe). Deuxièmement, dès que la segmentation d'entrée est effectuée, la correspondance des segments d'entrée avec des unités de la modalité de sortie doit être mise en œuvre. C'est ce que l'on appelle l'opération d'appariement et elle dépend des règles d'associations entre phonèmes et graphèmes mais aussi des probabilités de leurs mises en relation. Enfin, à la suite de l'opération d'appariement, les segments cibles de

la modalité de sortie seraient assemblés pour former l'item de sortie. Notons que chacune de ces opérations (segmentation, appariement, assemblage) pourrait être affectée par des troubles.

En l'état actuel, les chercheurs pensent que l'évolution de la compétence en lecture et en écriture est de moins en moins influencée par la phonologie, au fur et à mesure que le niveau de maîtrise de l'écrit augmente (Fayol & Jaffré, 2008 : 149). La voie phonographique ne serait qu'exceptionnelle et serait écartée au profit d'autres stratégies, dont nous parlerons dans la partie suivante.

1.3.2.3. D'autres stratégies pour orthographier

Entre la voie lexicale et phonographique, plusieurs stratégies sont envisageables. Tout d'abord, nous pourrions recourir à une analogie orthographique lorsque la forme sonore d'un mot à transcrire ressemble à une autre que nous avons déjà rencontrée et transcrite. Si nous devons transcrire un nom nouveau dont la forme sonore est [kobar], nous pourrions l'écrire de la manière phonologique la plus simple *cobar*, ou nous pourrions aussi l'écrire *cobard*, par analogie avec *foulard*, *dard*, *lard*... Dans d'autres situations, nous pourrions saisir les informations qui impliquent le sens du mot à transcrire. Si le nom [kobar] est présenté dans un contexte qui se réfère à un métier et que nous apprenons que la pratique de cette activité est le verbe [kobarte], nous pourrions bien évidemment transcrire *cobart*. L'ajout du *t* permet d'établir une relation de sens avec le verbe, ou d'autres mots de la même famille morphologique comme *corbartage*, *cobarteur*... Il s'agit toujours d'une analogie, mais cette fois, la transcription repose sur « des dérivations associées au sens des mots ou sur la morphologie dérivationnelle » (Fayol & Jaffré, 2008 : 139).

Cependant, la transcription des seules formes phonologiques, particulièrement en français, ne suffit pas à produire une orthographe correcte. Cela s'explique par le fait que, en français, le marquage du nombre et dans une moindre mesure du genre, oblige à mobiliser des marques, *-s* pour les noms et les adjectifs, *-nt* pour les verbes, qui n'ont pas de correspondance phonologique. Et ces marques de morphologie dite flexionnelle

sont très fréquentes en français.

1.3.2.4. La variété des procédures individuelles

En résumé, une production orthographique pourrait s'effectuer en mobilisant différentes stratégies. Comme le remarquent Rittle-Johnson et Siegler (1999), l'orthographe n'est pas un domaine de connaissances dans lequel une quelconque stratégie garantit la réussite puisque chaque stratégie présente non seulement intérêt, mais aussi risques et limites. Notons que « ni le recours aux correspondances phonèmes-graphèmes, du fait de l'existence de nombreux mots irréguliers, ni la récupération directe en mémoire, en raison de la fragilité des représentations orthographiques, n'assure la réussite » (Fayol & Jaffré, 2008 : 152).

Pour bien produire l'orthographe, un éventail de procédures serait nécessaire : la remémoration de la forme orthographique des mots, l'analogie orthographique, le recours aux règles, l'emploi des correspondances entre phonèmes et graphèmes, l'emploi du dictionnaire, etc. La réussite en orthographe demande au scripteur non seulement la connaissance de ces stratégies, mais aussi la capacité à les mobiliser et à les mettre en œuvre avec discernement. Ainsi, nous pourrions comprendre les variations de performance des mêmes individus, puisque, dans une situation donnée, la performance en orthographe dépend « à la fois de la tâche, des contraintes incluant l'état de fatigue et la pression sociale, et des connaissances préalables, y compris relativement à ses propres capacités » (Fayol & Jaffré, 2008 : 141).

1.3.3. Les erreurs et sources d'erreur dans l'orthographe du français

1.3.3.1. Les types d'erreurs

En ce qui concerne l'orthographe, lire et écrire ne posent pas les mêmes difficultés, puisque ces deux activités présentent d'importantes différences : d'abord, par rapport aux configurations écrites (lettres, mots et signes de ponctuation), qui sont, en lecture,

concrètes, permanentes et donc faciles à définir, les configurations sonores à transcrire sont « plus variables d'une région, d'un individu et d'un moment à l'autre, moins concrètes, plus évanescences, et donc plus difficiles à isoler, discriminer et identifier » (Fayol & Jaffré, 2008 : 123). De ce fait, la performance orthographique en écriture varie selon l'accent ou le dialecte qu'on a.

Ensuite, la lecture, c'est aller d'un nombre donné de graphèmes ou de lettres à un nombre plus limité de phonèmes. En revanche, en écriture, il s'agit d'aller d'un nombre limité de phonèmes à un nombre plus élevé de graphèmes. Par exemple, en français, on utilise environ 130 unités de lettres pour transcrire une trentaine de phonèmes. Ainsi se créent les homophones, mots qui se prononcent de manière équivalente, mais qui se transcrivent différemment. En production orthographique, ces homophones induisent des difficultés spécifiques, étant donné que le scripteur ne peut pas s'appuyer tout simplement sur la forme phonologique pour transcrire la forme orthographique.

Enfin, selon l'individu ou le dialecte, la prononciation présente une importante variabilité. En général, nous acceptons ces variations, souvent inaperçues, sans problèmes. Mais la production orthographique présente une tendance inverse, qui est rigide et supporte très peu de variations.

Ces différences entre lire et écrire rendent la production orthographique particulièrement complexe et la dernière nécessite donc « une vigilance continue de part de ceux qui rédigent et l'existence à l'école d'un curriculum, c'est-à-dire un programme spécifique pour l'instruction » (Fayol & Jaffré, 2008 : 125). C'est pourquoi même les plus experts en écriture, en dépit de leurs nombreuses compétences, ne sont pas à l'abri de faire des erreurs. Notons que ces erreurs nous informent de façon utile sur « le fonctionnement de l'expertise en orthographe, en illustrant la distinction entre les savoirs orthographiques et leur gestion » (Ducard, Honvault & Jaffré, 1995 : 100). Pour bien analyser ces erreurs, il nous semble nécessaire de disposer préalablement des connaissances sur la classification des erreurs. Nous avons eu recours ici à celle de Bexis Köhler (1993), une des versions issues des recherches de Catach (1986) (cité par Ducard, Honvault & Jaffré, 1995). Cette typologie nous permet « un affinement sensible, fondé linguistiquement, de la description des erreurs orthographiques »

(Masseron & Lusste-Chaa, 2008). Ce classement d'erreurs présente une exhaustivité : tous les types d'erreurs orthographiques sont représentés. Au total, il regroupe sept catégories d'erreurs, déterminées à partir des fonctions des graphèmes (Weck & Fayol, 2009) :

- Les erreurs phonétiques : omissions (**mercrdi* pour *mercredi*), confusions (**ballan* pour *ballon*), inversions (**quarte/quatre*) et adjonctions (**quartre/quatre*) de graphèmes ;
- Les erreurs phonogrammiques qui résultent d'une non connaissance du code graphique du français : certaines erreurs altèrent la valeur phonique du mot (**assiette* pour *assiette* ; **gider* pour *guider*), d'autres non (**aciette* pour *assiette* ; **anfant* pour *enfant*) ;
- Les erreurs morphogrammiques grammaticales qui résultent du non-respect des morphogrammes grammaticaux : marques du genre et du nombre (*petit* pour *petite* ; **chevaus* pour *chevaux*), et terminaisons des verbes (**j'é* pour *j'ai* ; **ils chan- tais* pour *ils chantaient*) ;
- Les erreurs morphogrammiques lexicales qui résultent du non-respect des morphogrammes lexicaux : en particulier, préfixes et suffixes (**lentemant* pour *lentement*), lettres justifiables par dérivation (**enfan* pour *enfant* ; **petid* pour *petit*) ;
- Les erreurs logogrammiques ont trait essentiellement à la figure des mots et permettent surtout de distinguer les homophones (*sait* pour *c'est* ; *a* pour *à* ; *ver* pour *verre*) ;
- Les erreurs concernant les graphèmes non fonctionnels : essentiellement consonnes doubles (**dificile* pour *difficile*) et lettres finales (**paraplui* pour *parapluie* ; **longtemp* pour *longtemps*) ;
- Les erreurs idéogrammiques concernant surtout les traits d'union (**vingt quatre* pour *vingt-quatre*) et les majuscules (hormis les majuscules liées à la ponctuation).

1.3.3.2. Les problèmes attentionnels

Les recherches sur les performances orthographiques des adultes montrent que les

adultes n'échappent pas à la difficulté orthographique. Des chercheurs ont analysé les erreurs orthographiques recensées dans 420 pages d'écrits ordinaires en essayant de montrer qu'il existe dans un système linguistique donné, indépendamment des scripteurs et des situations, des zones de fragilité spécialement favorables à l'occurrence de fautes orthographiques (Fayol & Jaffré, 2008 : 127-130). Les données recueillies par les chercheurs mettent en évidence l'existence de zones à risque particulièrement propices à l'apparition d'erreurs : la marque du pluriel, les variations de doubles consonnes, la présence d'homophones, la morphologie du verbe. Néanmoins, les données montrent que certaines situations sont plus favorables que d'autres à la survenue de fautes : les situations formelles sont associées le plus souvent à des erreurs sur la morphologie verbale, alors que les situations informelles présentent plus d'erreurs sur l'accord en nombre des noms. Signalons également que certaines populations produisent plus d'erreurs que d'autres : plus le niveau de formation scolaire est élevé, plus la fréquence d'erreurs diminue, sans toutefois disparaître. Les chercheurs indiquent aussi qu'il existe des interactions entre ces différents facteurs : par exemple, les différentes situations n'ont pas les mêmes incidences sur les différents individus, instruits ou non.

Pour approfondir les études sur la performance orthographique des adultes, les chercheurs se concentrent sur la question des accords (Fayol & Jaffré, 2008 : 157-166). La gestion des accords pose particulièrement problème puisque la plupart d'entre eux ne se prononcent pas à l'oral. Notons que trois catégories d'erreurs retiennent l'attention des chercheurs, ce qui leur permet de mieux comprendre le processus qui gère l'accord chez les adultes.

D'abord, ce sont les erreurs qui consistent en l'omission de la flexion : *les jeune(s) ne font plus attention ; les étudiants entre (nt) dans la salle*. Ceci s'observe souvent chez ceux qui utilisent les claviers. Selon les chercheurs, il s'agit d'un problème d'attention. L'orthographe du français est un système dont les marques d'accord sont le plus souvent silencieuses, la gestion d'accord requiert donc toujours beaucoup d'attention. Si une activité (par exemple, taper sur le clavier d'ordinateur, en nous obligeant à tourner le regard régulièrement du clavier vers l'écran, diminue ainsi notre

attention portée au texte au cours de la réalisation) capte l'essentiel voire la totalité de l'attention, il n'en reste plus ou plus assez pour la gestion d'accords. Lorsque la surcharge d'attention est maximale, les erreurs surviennent.

Ensuite, ce sont des erreurs qui consistent à accorder le verbe non pas avec le sujet, mais avec le mot, nom ou pronom, qui le précède immédiatement : *les chiens du voisin arrive (nt) ; la mère des enfants s'inquiètent*. Les spécialistes les appellent une « erreur d'attraction ». Lorsque l'on focalise l'attention sur le verbe, les adultes sont capables de repérer et corriger ce type d'erreur. Cependant, elle passe le plus souvent inaperçue. Il faut noter qu'il est facile d'induire ces erreurs chez les adultes qui sont habitués à rédiger sans faute. Les chercheurs leur demandent de transcrire une phrase, telle que *le chien des voisins arrive*, sous la dictée et de réaliser une seconde tâche en transcrivant. En l'occurrence, le nombre d'erreurs s'élève fortement, alors qu'aucune erreur ne survient avec *le chien du voisin arrive* ou *les chiens des voisins arrivent* si les adultes les transcrivent sans une seconde tâche.

Enfin, ce sont les erreurs constituées par la substitution d'un mot par un autre, qui a la même prononciation mais une orthographe différente : *Le vagabond a des poches. Il les fouille(s)*. Notons que ces erreurs ne surviennent qu'avec certains verbes qui ont un homophone nominal : *il juge/le juge ; il fouille/la fouille*, etc. Les analyses montrent que le pourcentage d'erreurs est le plus élevé lorsque l'homophone nominal est plus fréquent que le verbe : *Le postier prend des paquets. Il les timbre(s)*. Concernant le mécanisme des erreurs de ce type, l'interprétation la plus plausible revient à considérer qu'au lieu de recourir à la règle ou au calcul de l'accord avec le sujet, les adultes récupèrent directement dans leur mémoire certains mots déjà tout fléchis. Il s'agit des mots qui portent souvent « la flexion nominale pour les verbes ayant un homophone nominal, surtout si cet homophone nominal est plus fréquent, et donc dominant » (Fayol & Jaffré, 2008 : 161). Ce mécanisme est adaptatif : il permet de maximiser les réussites en orthographe en minimisant le coût en temps et en attention. Néanmoins, il n'évite pas toutes les erreurs.

En résumé, toutes ces analyses nous permettent de mieux comprendre les mécanismes de la gestion de l'accord. D'abord, il faut noter que les adultes disposent

des règles d'accord de l'adjectif, du nom et du verbe. Cependant, la plupart du temps, ils ne s'y réfèrent pas, à cause du coût élevé en temps et en attention. Ensuite, pour effectuer les accords, ils mobilisent soit « une procédure simplifiée, vraisemblablement automatisée, l'accord avec le mot qui précède », soit « une procédure de récupération en mémoire d'une forme toute fléchie » (Fayol & Jaffré, 2008 : 165). Ces procédures, rapides et peu coûteuses en attention, réussissent le plus souvent. Néanmoins, comme nous l'avons dit plus haut, elles n'évitent pas toutes les erreurs. Mais si les adultes disposent de suffisamment de temps et d'attention pour mobiliser la règle explicite et procéder au calcul d'accord, ils sont en mesure de corriger l'erreur, même avant de l'avoir transcrite. Évidemment, le coût en attention serait très élevé, voire aux dépens d'une autre activité.

Il s'agit donc de distinguer entre les savoirs orthographiques et leur gestion. Les adultes sont généralement capables d'accorder les mots mais, dans un texte donné, à un moment particulier de la production écrite, ils ne sont pas en mesure de contrôler les effets du contexte syntaxique. En l'occurrence, ils doivent mobiliser d'autres compétences qui sont coûteuses en attention. Par ailleurs, la mise en œuvre de ces compétences exige des conditions particulières : « l'exercice de la compétence orthographique a besoin d'un certain espace mental et plus la part de calcul est grande, plus cet espace doit être important » (Ducard, Honvault & Jaffré, 1995 : 100). Ces problèmes, propres aux adultes, nous donnent un avant-goût de ceux que nous allons rencontrer avec des jeunes adultes chinois d'acquisition de l'orthographe française.