

# *OrthoFonic* : un projet de didacticiel pour l'apprentissage de l'orthographe française

Claire FONDET & Fabrice JEJCIC

UMR 8589, CNRS-LAMOP, Paris 1 Panthéon Sorbonne  
Histoire des systèmes graphiques du français et de ses variétés  
Laboratoire de médiévistique occidentale de Paris

*OrthoFonic* is a lexical database of 5,500 words and 8,500 forms which is based on the *Listes orthographiques de base* (Basic Spelling Lists, 1984, 4000 forms). This list takes into account the revised spellings of the “Rectifications orthographiques” of 1990. To these Spelling Lists, which are in turn taken essentially from the *Frequency Dictionary of French Words* by A. Juilland *et al.* (1970), have been added the corresponding *Alfonic* phonological transcriptions and the IPA phonetic transcriptions as used in the *Petit Robert*. Thus, *OrthoFonic* could become the motor for a computerised teaching program constructed around the orthographical, phonological, phonetic and grammatical information associated with the most frequent lexical items of written French.

Creating a software program based on this high frequency lexis would make it possible to do away with forms that are rare or unused, and could encourage the acquisition of the written forms that are the most frequent in actual use.

## Introduction

Chacun sait que la plupart des erreurs orthographiques, chez l'enfant comme chez l'adulte, sont commises sur les mots les plus ordinaires et les plus fréquents, plutôt que sur les mots rares (Ters, 1973: 60), cf. les exercices grammaticaux habituels concernant les homonymes monosyllabiques tels que: *a/à*, *ou/où*, etc. Ces observations plaident, avec évidence, en faveur d'une didactique de l'acquisition de l'écrit construite sur les formes les plus récurrentes de la langue.

Le projet *OrthoFonic* (combinaison de *orthographe* et *Alfonic*) ambitionne de renouveler l'enseignement du français courant en partant de formes vedettes et de formes fléchies classées selon un critère de fréquence. Les deux partenaires, le CNRS-LAMOP et l'Association RAPHAEL, recherchent actuellement un financement auprès de la région Île-de-France dans le cadre d'un PICRI<sup>1</sup>, pour la mise au point d'un didacticiel d'apprentissage de l'écrit et de l'oral du français d'usage courant.

---

<sup>1</sup> RAPHAEL = Recherche pour l'Application de la PHonologie aux Apprentissages de l'Écriture et de la Lecture, PICRI = Partenariats Institutions-Citoyens pour la Recherche et l'Innovation. Le financement serait destiné à la mise au point informatique du didacticiel.

Ce projet pourrait intéresser à la fois le public scolaire (niveau élémentaire et début du secondaire), les personnes en difficulté (enfants et adultes éprouvant des gênes langagières, qu'elles soient orales ou écrites), les enfants scolarisés nouvellement arrivés en France (français langue seconde), les personnes en situation d'apprentissage du français langue étrangère (oral et écrit) et les migrants en situation d'alphabétisation dans les pays francophones<sup>2</sup>.

## 1. Un vocabulaire scientifiquement construit

Les *Listes orthographiques de base* (1620 mots, 4000 formes)<sup>3</sup> étaient, d'une certaine manière, et dit en toute modestie, en avance sur le temps quant à leur utilisation. À l'époque de leur élaboration sous la direction de Nina Catach, au début des années quatre-vingt et publiées en 1984, il n'existait pas encore de moyens techniques pour les utiliser à l'école autrement que sur un support papier. L'informatique était centralisée, les programmes et les données fonctionnaient avec des cartes perforées et, en dehors des laboratoires scientifiques, leur exploitation informatique n'était pas possible. Chaque tri, des quelque 4000 cartes comportant les mots de ces listes, coûtait l'équivalent de 2,60 € et l'on recevait le lendemain à Paris les tris imprimés des résultats du supercalculateur d'Orsay qui étaient livrés par navette. Aujourd'hui, la base complète, avec deux fois plus de données et trois fois plus de champs, nous livre un tri, le temps de taper sur la touche *entrée* du clavier d'ordinateur. Ce changement technologique et la banalisation de l'usage du micro-ordinateur ont motivé l'élaboration du projet actuel.

La base de données lexicales *OrthoFonic* (5500 mots, 8500 formes), fondée pour une grande part sur les *Listes orthographiques de base* (version 1994: 5000 mots, 8000 formes), a la particularité de contenir les formes fléchies les plus fréquentes et les mots les plus fréquents (*a*, *ai*, *avait* et *ont* sont les formes fléchies les plus fréquentes du mot *avoir*, forme vedette du verbe, ou lemme, au sens d'entrée de dictionnaire). D'autre part les *Listes orthographiques de base* (LOB), dont le *Frequency Dictionary of French Words* (FDFW), 5082 mots (Juilland et al., 1970), avait servi de fondement, avaient été complétées par les principales listes de vocabulaire du français: *L'élaboration du français élémentaire*, 1063 mots (Gougenheim et al. 1956, 1971), *Le vocabulaire orthographique de base*, 7691 mots (Ters et al., 1968), *L'échelle Dubois-Buyse*, 3725 mots (Ters et al., 1969) et, enfin, le *Dictionnaire des fréquences du Trésor de la langue française*, 71415 mots

<sup>2</sup> La didactique du français langue seconde (FLS) se distingue de celle du français langue étrangère (FLE) et du français langue maternelle (FLM). Sans entrer dans le débat de cette classification, disons simplement que la didactique doit être adaptée à la situation.

<sup>3</sup> Catach N., Jecic F. et équipe HESO (1984): *Les listes orthographiques de base du français*, Paris, F. Nathan, 160 p.

(document interne du CNRS sous forme de volumes polycopiés). Toutes ces listes, construites à partir d'enquêtes ou de corpus informatisés, font apparaître une grande concentration du vocabulaire d'usage très courant: elles ont en commun la quasi-totalité des 1280 premiers mots du *Trésor de la langue française* dont la fréquence, rapportée au nombre d'occurrences, est supérieure à 5000 (*LOB*, 1984: 17).

## 2. Un vocabulaire performant, de haute fréquence d'usage

Ce travail de synthèse sur les listes de fréquence assure un haut rendement au vocabulaire sélectionné. D'après nos études, dont les résultats sont comparables à la plupart des autres études de fréquence du vocabulaire, nous obtenons, pour les 5082 mots du *FDFW*, les chiffres suivants:

- les 10 premiers mots couvrent près de 35 % de la fréquence d'usage<sup>4</sup>;
- les 100 premiers mots couvrent près de 68 % de la fréquence d'usage;
- les 1000 premiers mots couvrent près de 89 % de la fréquence d'usage.

Parallèlement à cette observation, le *Dictionnaire des fréquences du Trésor de la langue française* fait apparaître que les 50 premiers mots, auxquels on ajoute les formes des verbes *être* et *avoir*, représentent plus de 35 millions d'occurrences, soit 50% de la fréquence (*LOB*, 1984: 16).

Cette "performance lexicale" nous a conduits à penser que nous sommes ici en présence d'une forme de "noyau" de la langue qu'il convient de privilégier dans l'enseignement.

## 3. Un vocabulaire conforme à la norme orthographique à enseigner

Le lexique intègre les *Rectifications orthographiques de 1990*, publiées dans les *Documents administratifs* du *Journal officiel* du 6 décembre 1990. Ces rectifications, approuvées par l'Académie française, constituent, en principe, la nouvelle norme orthographique à enseigner (Bulletins Officiels de l'Éducation nationale: n° 5, du 12 avril 2007 et n° 6 du 28 août 2008). Dans la 9<sup>e</sup> édition en cours de son *Dictionnaire*, l'Académie rappelle que les nouvelles variantes cohabitent aux côtés des variantes graphiques traditionnelles pratiquées par les générations antérieures et: "[qu'] aucune de deux graphies ne peut être tenue pour fautive".

<sup>4</sup> Le coefficient d'usage est la mise en rapport complexe de la fréquence (nombre d'occurrences) et de la dispersion (distribution des occurrences suivant les différents genres de textes: romans, théâtre, essais, presse, textes scientifiques et techniques), cf. *FDFW*, *Introduction*, p. XXX-XXXIX (genres de textes) et p. LXVII-LXXI (usage).

Parmi les dictionnaires actuels, *Hachette*, depuis son édition de 2002, et le *Petit Larousse*, depuis son édition de 2012 (publiée en 2011), intègrent totalement les *Rectifications orthographiques de 1990*. Le *Petit Robert* les insère en partie au fur et à mesure des millésimes.

Voici quelques exemples de mots rectifiés présents dans *OrthoFonic*:

- *abime, accroitre, aout, boîte, bruler, couter, croute, diner, etc.* (suppression de l'accent circonflexe sur *i* et *u*), idem pour *apparaître* et les autres verbes en *-aitre*;
- un *après-midi* / des *après-midis*, un *porte-avion* / des *porte-avions* (marque du pluriel systématisée sur le second élément du composé à trait d'union);
- *à priori* (francisation de l'emprunt latin avec accent grave sur la préposition);
- *assoir* (correction de l'anomalie graphique par suppression du *-e-* parasite);
- *cèdera* (accent grave conforme à la prononciation pour le verbe *céder* et les autres verbes du même type);
- *asséner, évènement* (alignement de l'accentuation sur la prononciation);
- francisation des emprunts (accentuation, soudure de certains composés et marques de nombre), qu'ils soient latins (*référendum, arborétums*), italiens (*spaghettis, à capella*), anglais (*weekend, cowboys*), ou appartiennent à d'autres langues encore.

Les *Rectifications orthographiques de 1990*, par les modifications apportées, représentent donc une rationalisation du système graphique.

#### 4. Une structure lexicale élaborée pour la didactique

*OrthoFonic* présente une structure hiérarchisée composée de 6 listes complémentaires de vocabulaire (voir Fig. 1, ci-dessous).

Les listes LB $\emptyset$  (LB "zéro"), LB1, LB2 et LB3 – qui composent les *LOB* 8000 formes – fonctionnent sur des indices de fréquence décroissante dont la distribution est globalisée à l'intérieur de quatre classes de fréquences. Ainsi la liste LB $\emptyset$  contient-elle les formes lexicales les plus fréquentes et la liste LB3 les formes lexicales les moins fréquentes.

La liste LB4 comprend le vocabulaire le plus fréquent en usage dans la presse (Skupien-Dekens C., Kamber A. & Dubois M., 2010: Avant-propos) et un lexique de la vie pratique issu de méthodes d'alphabétisation. La liste LB4 ne fonctionne pas sur des indices de fréquence calculés, mais intègre un lexique pratique courant, absent des *LOB*.

La liste LB5 est prévue pour accueillir des lexiques personnels, des vocabulaires spécifiques aux utilisateurs ou aux groupes d'utilisateurs.

Cet ensemble de listes peut être représenté par le tableau suivant:

N° liste / nb formes	N° d'usage FDFW	Fréquence ou type
LB0 - 157 formes	inférieur à 100	supérieure à 1000
LB1 - 403 formes	entre 100 et 250	supérieure à 30
LB2 - 3459 formes	inférieur à 1280	supérieure ou = à 5
LB3 - 4000 formes environ	supérieur à 1280	inférieure à 5
	Lexiques complémentaires	Types vocabulaires
LB4 - 500 formes environ	lexiques de la vie pratique	méthodes d'alphabétisation vocabulaire de la presse
LB5 - nb de formes ouvert	lexique personnel / du groupe	vocabulaires des usagers

Fig. 1: Hiérarchie de la structure lexicale

#### 4.1 La fréquence du vocabulaire, listes LB0 à LB3

La notion de fréquence d'usage est tout particulièrement productive pour la didactique. Si nous prenons pour exemple le verbe *être* qui, toutes formes conjuguées réunies, arrive au 4<sup>e</sup> rang de la fréquence d'usage, ceci indique que ce verbe est le 4<sup>e</sup> mot le plus courant. Ce mot vedette représente toute une série de formes verbales; ses formes conjuguées apparaissent dans les trois premières listes:

- *est, était et sont*, dont la fréquence est supérieure à 1000, font partie de LB0;
- *es, étaient, étions*, et encore quelques autres formes dont la fréquence est supérieure à 30, appartiennent à la liste LB1;
- *fussent, serions, soyez*, et encore quelques autres formes dont la fréquence est supérieure ou égale à 5, relèvent de la liste LB2.

Toutes ces formes fléchies ont comme dénominateur commun le numéro du rang d'usage qui est le 4.

#### 4.2 Les lexiques complémentaires, liste LB4

La liste LB4, qui n'est pas construite sur des calculs de fréquence, est cependant élaborée à partir du vocabulaire le plus fréquent de la presse écrite et à partir de lexiques de méthodes d'alphabétisation. À ce jour, cette liste comprend 500 mots supplémentaires, permettant d'actualiser les premières listes LB0 à LB3.

Les tableaux, ci-dessous (Fig. 2 et 3), présentent:

- colonne 1: la forme orthographique;

- colonnes 2, 3 et 4: la nature grammaticale, le genre et le nombre;
- colonne 5: la provenance de la forme: code PR = lexique de la presse, code AL = méthodes d'alphabétisation;
- colonne 6: la transcription phonologique en Alfonic;
- colonne 7: la transcription phonétique en alphabet phonétique international (API). Pour ces deux dernières, cf. § 4.6.

1 - Orthographe	2 - NG	3 - Ge	4 - Nb	5 - L	6 - Alfonic	7 - API
bancaire	AJ	S		PR	bäcer	bãkɛR
banlieue	NO	F	S	PR	bälix	bãljø
bicyclette	NO	F	S	PR	bisiclet	bisiklɛt
bloquer	VE			PR	bloce	bloke
boursier	NO	M	S	PR	bwrsie	bursje
caméra	NO	F	S	PR	camera	kamera
canal	NO	M	S	PR	canal	kanal

1 - Orthographe	2 - NG	3 - Ge	4 - Nb	5 - L	6 - Alfonic	7 - API
biscotte	NO	F	S	AL	biscot	biskɔt
boisson	NO	F	S	AL	bwasö	bwasõ
boiter	VE			AL	bwate	bwate
bonbon	NO	M	S	AL	böbö	bõbõ
bougie	NO	F	S	AL	bwji	buzi
bouillon	NO	M	S	AL	bwyö	bu jõ
boulangerie	NO	F	S	AL	bwläxri	bulãʒɛRI

Fig. 2 et 3: Liste LB4: extraits du vocabulaire le plus fréquent de la presse et extraits du vocabulaire de méthodes d'alphabétisation.

### 4.3 *Le lexique utilisateurs, la liste LB5 ouverte*

La liste LB5 est une partie ouverte destinée à accueillir le lexique personnel de l'utilisateur ou du groupe. Ainsi, par exemple, dans le cadre d'activités spécifiques de la classe, les élèves pourraient intégrer dans cette liste un lexique spécialisé comme celui des sciences de la nature, celui de la géographie ou d'une autre discipline dont les vocables n'appartiennent pas à la langue courante. Et, pourquoi pas aussi des mots du vocabulaire personnel de l'apprenant?

### 4.4 *Le codage orthographique*

Une des difficultés de l'orthographe française provient des lettres muettes et des liaisons obligatoires ou facultatives. C'est pourquoi, en 1984, on avait choisi un codage pour marquer les différents types de lettres

muettes<sup>5</sup>, et un autre codage pour indiquer les liaisons, lorsque la forme considérée est placée devant un mot à initiale vocalique. Il s'agit essentiellement des lettres finales *n, r, s, t* (LOB, 1984: 20-23)<sup>6</sup>. Le tableau ci-dessous illustre ce système et nous donne le nombre de formes relevant de chaque code:

Codage	Type	Exemple	Nb formes
• code 7	lettres lexicales	<i>doigt77</i>	1122
• code 8	lettres grammaticales	<i>aimées88</i>	3116
• code =	lettres susceptibles de liaison	<i>chers8=</i>	373

Le code 7 indique des lettres muettes de type lexical, le code 8 les lettres muettes de type grammatical et le signe = indique les liaisons potentielles (laissées à l'appréciation de l'enseignant). Dans l'exemple *doigt77*, où ni le *-g-* ni le *-t* ne se prononcent, les codages du *-g-* par 7 et du *-t* par 7 établissent un lien avec les mots de la même famille: *doigté* et l'étymologie latine *digitum*. Par contre, ces deux lettres sont prononcées dans la forme dérivée *digital*. Cela correspond à la fonction dérivative de ces graphèmes qu'il est intéressant de transmettre dans le cadre d'un apprentissage. Dans l'exemple *aimées88*, où les deux dernières lettres ne sont pas prononcées, le codage du *-e-*, marque de féminin par 8, et le codage du *-s*, marque de pluriel par 8, renseignent sur la nature grammaticale de ces graphies. Car le *-e* est souvent une façon de marquer le féminin d'une forme, et le *-s* est très fréquemment la marque du pluriel des adjectifs et des noms, bien qu'il indique aussi la marque verbale de la 2<sup>e</sup> personne du singulier. Dans l'exemple *chers8=*, les codes 8 et = signifient respectivement que l'adjectif est au masculin pluriel (code 8) et que le *-s* peut être prononcé devant un mot commençant par une voyelle, comme dans la séquence: *chers8= amis*. Ce système de codage est une aide précieuse à la compréhension et à la maîtrise d'une partie difficile du système grapho-phonique du français.

#### 4.5 Le codage grammatical

Le codage grammatical concerne la nature grammaticale des formes (parties du discours), le genre, le nombre, et indique les mots vedettes (cf. § 1).

<sup>5</sup> Les chiffres 7 et 8 utilisés dans ce codage sont liés au fait qu'à l'époque le codage informatique était limité aux lettres majuscules sans accent. Ainsi, afin de pouvoir traiter les diacritiques, le code 1 représentait l'accent aigu (ex.: E1=é), le code 2 l'accent grave, le code 3 l'accent circonflexe, le code 4 le tréma, le code 5 la cédille. De plus le chiffre 6 codait le nom propre (ex.: PARIS6).

<sup>6</sup> La sélection de ces cas de liaison a reposé uniquement sur le sentiment linguistique des auteurs à l'époque.

- Codage du champ "nature grammaticale"; les codages suivants sont utilisés:

Code	Valeur	CJ	conjonction	NU	numéral
AJ	adjectif	IN	interjection	PN	pronom
AR	article	LO	locution	PS	préposition
AV	adverbe	NO	nom	VE	verbe

- Codage du champ "genre"; les codages suivants sont utilisés:

Code	Valeur	Remarque:
F	féminin	l'absence des codes de notation du genre, M ou F, indique qu'il s'agit d'une forme épïcène
M	masculin	

- Codage du champ "nombre"; les codages suivants sont utilisés:

Code	Valeur	Remarque:
S	singulier	l'absence des codes de notation du nombre, S ou P, indique qu'il s'agit d'une forme invariable
P	pluriel	

- Codage du champ "vedette"; le codage suivant est utilisé:

Code	Valeur	Remarque:
*	forme vedette	la présence de l'astérisque dans le champ indique une forme vedette, type entrée de dictionnaire (ex.: <i>*être</i> );
	forme fléchie	l'absence de l'astérisque dans le champ indique une forme fléchie, réalisée dans le discours (ex.: <i>était</i> 8=)

#### 4.6 Les codages phoniques

Pour le versant oral de la langue, *OrthoFonic* présente deux codages phoniques: l'API et l'Alfonic. L'API fut inventé pour noter toutes les langues du monde, en 1888, par Paul Passy. Ce codage est plus particulièrement adapté à l'apprentissage du français langue étrangère. En revanche pour les élèves francophones, André Martinet, à la demande d'enseignants dans les années soixante-dix, mit au point le code phonologique dénommé Alfonic, selon les oppositions pertinentes du français oral. Le but était d'enseigner l'écrit à partir de l'oral. L'utilisation du clavier de la machine à écrire et la proximité avec la graphie du français ont été privilégiées. Ce code était considéré comme un outil que les élèves devaient s'approprier depuis la dernière année de Maternelle jusqu'au Cours Préparatoire. Dans la perspective de l'acquisition de l'orthographe, l'emploi de l'outil Alfonic qui est conçu comme une écriture de transition, s'est révélé très efficace, conduisant à la mise en place d'une conscience critique de la norme orthographique (Vial J. et Villard J., 1986: 1-2).

Peu d'écoles pratiquent l'Alfonic aujourd'hui, mais un didacticiel a été élaboré en application du projet européen *Je parle donc j'écris* (cédérom et



livret produits en 2005). Ce logiciel est accessible gratuitement (voir sites internet). L'Association RAPHAEL poursuit ce projet.

Certains caractères de l'Alfonic sont différents de ceux de l'API (quelques consonnes et la plupart des voyelles). Pour les consonnes il n'y a pas de différence de signes entre les deux codes. Inversement, certaines voyelles sont notées différemment. L'Alfonic suit la théorie de la phonologie présentée dans les *Éléments de linguistique générale* d'André Martinet. Sur le plan phonologique, il n'y a pas de différence entre les deux codes en ce qui concerne le système consonantique. La question se pose essentiellement pour la notation de certaines oppositions vocaliques (loi de position). Il s'agit des voyelles orales d'ouverture moyenne correspondant aux archigraphèmes É, È, EU, O.

(1) en Alfonic: /è/ ~ /e/, en API: [ɛ] ~ [e]; (2) en Alfonic: /o/ ~ /ô/, en API: [ɔ] ~ [o]; (3) en Alfonic /x/ ~ /x^/, en API: [œ] ~ [ø].

- 1) Opposition /è/ ~ /e/, en API [ɛ] ~ [e]: si nous prenons la forme verbale *prêtait* (707), celle-ci est transcrite ici en Alfonic: /pretè/ et en API: [pɾɛtɛ]. La différence réside ici dans la voyelle de la première syllabe. Car, du point de vue phonologique, il n'existe pas en français d'opposition /e/ ~ /è/ en syllabe intérieure, c'est-à-dire non finale, même si l'orthographe autorise une prononciation ouverte de la voyelle en question. On considère en effet que dans ce cas /e/ et /è/ sont en distribution complémentaire ([e] en syllabe ouverte, [ɛ] en syllabe fermée). Le fait de prononcer /è/ en syllabe ouverte est ici une variante libre, à la différence de *pré* / *prêt* (Alfonic /pre/ ~ /prè/, API [pɾɛ] ~ [pɾɛ]). On trouvera donc en API [ɛ] en toute position, contre /è/ seulement en finale absolue en Alfonic. Autres exemples: *perdu*, Alfonic \perdu\, API [pɛɾdy]; *paisible*, Alfonic \pezibl\, API [pɛzibl]; *prévu*, Alfonic \prevu\, API [pɾevy].
- 2) Opposition /o/ ~ /ô/, en API: [ɔ] ~ [o]. L'API transcrit partout la différence voyelle ouverte ~ voyelle fermée, tandis que l'Alfonic ne garde ici cette opposition qu'en syllabe finale fermée. Exemples: *notre* = Alfonic: /notr/ et API: [nɔtɾ]; *le nôtre* = Alfonic /lx nôtr/ et API [lɔ notɾ]. Et par ailleurs *porter* donne en Alfonic \porte\ et en API [pɔɾte]; *poser* donne \poze\ en Alfonic et en API [poze].
- 3) Même chose pour l'opposition /x/ ~ /x^/, en API: [œ] ~ [ø]. L'Alfonic ne garde cette opposition que dans le cas de la voyelle en syllabe finale fermée. Cette opposition est bien représentée, car même s'il n'existe que de rares paires minimales (ex.: *jeune* ~ *jeûne*), le suffixe *-euse* est productif, ex.: *heureuse*.

Entre l'API et l'Alfonic, tous deux présents dans *OrthoFonic*, quelle notation faudra-t-il privilégier? Telle est la question qui se pose à l'enseignant. Cela dépendra de la situation pédagogique. L'API peut paraître plus "juste" pour noter les manières différentes de prononcer l'archigraphème *EU* dans *fleur*

et dans *feu*. L'expérience montre en effet que, dans les cas de neutralisation de voyelles, les enfants adoptent spontanément une démarche plus phonétique que phonologique<sup>7</sup>.

En fait, la différence entre les deux codes n'a pas été d'avantage accentuée. C'est ainsi que l'effacement du *schwa* intérieur et du *-e* final – ex.: *bib(e)ron*, *ball(e)* – ont été harmonisés dans les deux types de notation.

Quel que soit le code préféré, l'étude de la face phonique du vocabulaire fera prendre conscience de la variation dans l'oral et permettra de développer un esprit de tolérance à ce sujet. Certes, la base *OrthoFonic* peut paraître rigide, mais constitue un point de départ pour alimenter la réflexion.

Dans l'avenir, sur le plan technique, le codage en API pourrait permettre, via une interface appropriée, la reproduction sonore des vocables, utile en FLE et FLS, ainsi qu'en FLM, pour lever l'ambigüité graphique des formes homographes hétérophones (Jejcic 1987: 37-39), ex.: *les poules du couvent couvent*.

## 5. La base de données *OrthoFonic*

Dans sa configuration actuelle, la base de données comprend environ 8500 enregistrements et 16 champs par enregistrement. Le tableau ci-après (Fig. 4) présente la base et les fonctions des divers champs.

Nom du champ	Exemple	Fonction
Nusage	0072	<i>numéro d'usage</i> , établi d'après le rang d'usage du <i>FDFW</i> . Plus ce chiffre est petit, plus la forme est fréquente. L'adjectif <i>petit</i> est au 72 <sup>e</sup> rang, ce qui signifie qu'il est le 72 <sup>e</sup> mot (au sens de lemme), le plus utilisé
Ved	*	la présence du signe "*" étoile, indique qu'il s'agit d'une forme vedette, ex.: * <i>petit7</i> , avec étoile, est une forme vedette; <i>petits78=</i> , sans étoile, est une forme fléchie de l'adjectif <i>petit7</i>
Formes_C	<i>petit7=</i>	<i>formes des mots avec lettres muettes et liaisons codées</i> : ici le code "7" indique une lettre muette de type lexical, dérivative ou non dérivative selon le cas, ex.: <i>petit7</i> donne <i>petite</i> ; le code "8" indique une lettre muette de type grammatical, pluriel, ex.: <i>petits78=</i> pluriel; le code "=" indique une liaison possible devant initiale vocalique, ex.: <i>petits78= enfants</i> , chaîne graphique prononcée [pətizãfã] avec la liaison
Alfonic	pxti	transcription phonologique en caractères Alfonic
Nat	AJ	<i>nature grammaticale</i> , 10 catégories: AJ: adjectif, AR: article, AV: adverbe, CJ: conjonction, IN: interjection, LO: locution,

<sup>7</sup> Martinet, A. (1989: 12). Il s'agit de la neutralisation d'une voyelle dont l'ouverture ou la fermeture dépend du contexte comme dans le cas de *père*, autrefois prononcé [pɛrə]. Selon l'article cité, les enfants avaient tendance à noter *fête* en Alfonic: \fèt\ au lieu de \fet\.

		NO: nom, NU: numéral, PN: pronom, PS: préposition, VE: verbe
Ge	M	<i>genre</i> . Remarque: l'absence de code M ou F indique une forme épicène
Nb	S	<i>nombre</i> . Remarque: l'absence de code S ou P indique une forme invariable
Nl	0	<i>numéro de la Liste orthographique de base</i> : plus le numéro est petit, plus la fréquence de la forme est élevée. Les formes sont regroupées par "fourchette" de fréquence dans 4 listes: liste 0 [liste "zéro"] = mots très fréquents, liste 1 [liste "un"] = mots un peu moins fréquents, liste 2 = encore un peu moins fréquents, liste 3 = mots les moins fréquents; liste 4 = liste complémentaire ne fonctionnant pas sur des critères de fréquence, liste 5 = liste ouverte complétée par l'utilisateur
Sl	FD, DB, GG, HE, TL, AL, PR, _	<i>sous-liste</i> , indique la provenance de la forme ou son attestation dans une liste de vocabulaire: <i>FD</i> , la forme provient du <i>FDFW</i> ; <i>DB</i> , la forme provient de l' <i>Échelle Dubois-Buyse</i> ; <i>GG</i> , la forme provient du <i>Français fondamental</i> de G. Gougenheim; <i>HE</i> , la forme est codée par <i>HESO</i> (équipe de recherche du CNRS, 1973-1998), <i>TL</i> , la forme provient du <i>Dictionnaire de fréquence du Trésor de la langue française</i> (Institut National de la Langue Française-CNRS, aujourd'hui ATILF); <i>AL</i> vocables issus de méthodes d'alphabétisation; <i>PR</i> mots issus du vocabulaire fréquent de la presse; <i>absence de marque</i> , la forme provient de la liste ouverte par l'utilisateur
Formes	petit	<i>formes orthographiques</i> des mots (sans le codage des lettres muettes)
Inverse	titep	<i>formes inverses</i> des mots sans le codage des lettres muettes: les mots sont écrits "à l'envers", la dernière lettre du mot a le rang un et la première lettre du mot le dernier rang, ex.: <i>titep</i> est l'inverse de la forme <i>petit</i> - ceci permet de regrouper et trier des formes de mots qui ont des finales identiques
Inverse_C	7titep	<i>formes inverses</i> des mots avec le codage des lettres muettes: ex: <i>7titep</i> est l'inverse de la forme <i>petit7</i> - ceci permet de regrouper et trier soit des formes de mots qui ont des finales et des codes identiques (toutes les formes terminées par un -t codé 7), soit toutes les formes terminées par un code 7, indépendamment de la lettre qui précède le code
Syllab	pe tit	<i>syllabisation graphique de la forme</i> , selon les règles orthotypographiques de coupure des mots. La coupure syllabique est indiquée par un blanc graphique interne au mot: <i>pe_tit</i>
Phonetique	PàTI	<i>transcription phonétique</i> de la forme en code machine dont la correspondance en API est [pəti]
Syll_Phon	Pà/TI	<i>transcription phonétique</i> de la forme en code machine avec indication par un / "slash" de la coupure syllabique et dont la correspondance en API est [pə/ti]
Phon_Inv	ITàP	transcription phonétique inverse des formes, ex.: [itəp] est l'inverse de la forme [pəti], ceci permet de regrouper toutes les finales à prononciation identique

Fig. 4: Structure de la base de données *OrthoFonic*.

## 6. L'exploitation didactique de la base de données *OrthoFonic*

Les exemples qui suivent représentent quelques pistes d'utilisations diverses de la base sur les graphies des finales:

- les oppositions finales codées / finales non codées (§ 6.1.1)
- les finales codées de liaisons potentielles (§ 6.1.2)
- les finales codées de formes homophones (§ 6.1.3)
- les adjectifs épïcènes (§ 6.1.4)
- les noms féminins terminés par *-t* et les noms masculins terminés par *-e* (§ 6.1.5)

### 6.1 *Les recherches sur les graphies des finales*

Nous l'avons déjà mentionné plus haut: une grande partie des difficultés de l'orthographe provient des mots familiers et des lettres finales muettes. À ce dernier point en particulier correspondent des fonctionnements lexicaux et grammaticaux divers.

Une observation sur les graphies finales les plus fréquentes nous donne le tableau ci-dessous (Fig. 5), les graphies les plus fréquentes étant placées à gauche:

Nombre de finales par type de lettre								
-s	-r	-t	-n	-e	-x	-z	-d	-p
1600	1102	1053	463	414	142	87	68	11

Fig. 5: Recherche sur les lettres finales les plus fréquentes

Trois ensembles se distinguent nettement. Le premier, avec les finales *-s*, *-r* et *-t*, marque la prédominance du *-s*. La polyvalence du *-s* est distribuée entre la marque générale du pluriel des adjectifs et des noms, mais aussi la marque de seconde personne du singulier et de la première personne du pluriel du paradigme verbal, et souvent encore il peut être un marqueur graphique de l'adverbe et de la préposition, etc. Des remarques similaires s'appliquent au second ensemble, avec les finales *-n*, *-e*, et *-x*, et au troisième ensemble avec les finales en *-z*, *-d* et *-p*.

Ce type d'investigation sur les mêmes finales peut être poursuivi sur l'absence et/ou la présence du codage graphique des finales.

#### 6.1.1 Les oppositions finales codées / finales non codées

L'absence de code en finale indique que la lettre est prononcée. Dans le cas contraire, elle est codée 7 ou 8, ce que montrent les exemples qui suivent:

-s est prononcé dans *autobus*, il est muet dans *abus7* où le code, de type lexical, indique une dérivation possible comme dans *abuser*, *abusif*, *abusivement*; il est aussi muet dans *abeilles8* où le code, de type grammatical, indique le pluriel du nom *abeilles*. Le tableau ci-dessous (Fig. 6) donne une vue synthétique de ces fonctionnements, leur nombre d'occurrences (abréviation: nb.) et quelques exemples.

finale	nb.	exemples	finale	nb.	exemples	finale	nb.	exemple
-s	12	<i>autobus</i>	-s7	156	<i>abus7</i>	-s8	1229	<i>abeilles8</i>
-r	329	<i>aboutir</i>	-r7	66	<i>acier7</i>	-r8	705	<i>chanter8</i>
-t	27	<i>brut</i>	-t7	641	<i>abondant7</i>	-t8	299	<i>fait8</i>
-n	446	<i>abandon</i>						
			-e7	122	<i>entrée7</i>	-e8	251	<i>jolie8</i>
						-es78	41	<i>pensées78</i>
			-x7	43	<i>boiteux7</i>	-x8	51	<i>animaux8</i>
-z	1	<i>gaz</i>	-z7	2	<i>nez7</i>	-z8	75	<i>aimez8</i>
-d	1	<i>sud</i>	-d7	60	<i>abord7</i>			
-p	2	<i>cap</i>	-p7	8	<i>beaucoup7</i>			

Fig. 6: Oppositions entre les lettres finales les plus fréquentes

Dans l'exemple *abandon*, le -n est censé être prononcé, mais en réalité il est le second élément du graphème *on*, notant la voyelle nasale. La finale -x, marque lexicale dérivative dans *boiteux7*, est une marque de pluriel dans *animaux8*. Contrairement au -s final, très récurrent, le -x final est une graphie moins fréquente mais présente sur des noms et des adjectifs très courants, dont la dérivation est systématiquement réalisée par l'alternance de -x en -s-, *boiteux7* / *boiteuse*, *heureux7* / *heureuse*, etc. La finale -z correspond au second élément graphique du graphème *ez*, marque systématique de la seconde personne du pluriel des verbes. Par contre, dans *nez*, *chez* et *rez-de-chaussée*, le z a été conservé pour des raisons de distinction, notamment d'homonymes pour *nez* (cf. *DHOF*: 707).

### 6.1.2 Les finales codées de liaisons potentielles

Les liaisons, devant un mot à initiale vocalique, paraissent évidentes pour quelques mots-outils d'usage très courant: *pas7=*, *deux7=*, *est8=*, *grand7=*, *tes8=*, etc., voir tableau (Fig. 7).

let.	nb.	exemples	let.	nb.	exemples	let.	nb.	exemple
-s=	14	<i>en bas= âge</i>	-s7=	21	<i>après7=</i>	-s8=	168	<i>aimables8=</i>
			-x7=	42	<i>couteux7=</i>	-x8=	6	<i>aux8=</i>
			-z7=	2	<i>assez7=</i>	-z8=	7	<i>allez8=</i>
-t=	19	<i>bout= à bout7</i>	-t7=	23	<i>avant7=</i>	-t8=	44	<i>allait8=</i>
-d=	2	<i>de pied= en cap</i>	-d7=	5	<i>grand7=</i>			
			-p7=	1	<i>trop7=</i>			
			-r7=	2	<i>dernier7=</i>			

Fig. 7: Locutions courantes et mots fréquents produisant la liaison

Certes ces liaisons ne sont pas à mettre toutes sur le même plan: seules deux d'entre elles sont obligatoires ici. L'économie de la base *OrthoFonic* ne tient pas compte de cette différence. Cependant, le codage de certaines liaisons moins évidentes permet une réflexion sur les registres de langue, ce qui est le cas de l'adjectif *couteux7=* dont la liaison suppose un style soutenu.

D'autre part, environ 80 locutions courantes, ou expressions figées (*de pied= en cap*), ont été intégrées aux listes. Leur intérêt est d'apporter à l'enseignement de l'orthographe des compléments d'information sur des liaisons internes anciennes qui souvent ne se font plus par ailleurs.

### 6.1.3 Les finales codées de formes homophones

La distinction apportée par le système de codage est une des clés montrant les diverses fonctions des marques morphologiques de l'orthographe. La requête qui demande à la base de lister tous les mots prononcés en [fɛ], nous donne à l'écran le résultat suivant:

FORMES-C	NAT	GE	NB	SL
fais8=	VE		S	1
fait7	VE	M	S	1
fait8=	VE		S	1
faits78	VE	M	P	2
fait7=	NO	M	S	1/2
faits78	NO	M	P	1
fait7	AJ	M	S	3
faits78	AJ	M	P	3

Fig. 8: Listage de tous les mots qui se prononcent en [fɛ]

Les mots listés nous montrent l'opposition entre les verbes, les noms et les adjectifs ainsi que l'opposition entre les morphogrammes lexicaux et les morphogrammes grammaticaux. Cette différenciation entre les codes 7 et 8 est illustrée par les occurrences d'homographes *fait7=* et *fait8=*. Dans le

premier exemple, le code 7 indique une lettre muette de type lexical qui est rattachée à la catégorie du nom masculin singulier, le code =, qui s'applique à la même lettre, signale que le -t final est prononcé devant initiale vocalique, comme dans la locution "prendre fait et cause". En revanche, dans le second exemple, le code 8, désigne une lettre muette de type grammatical qui dépend de la catégorie du verbe au singulier. Le code = signale une liaison facultative, ex.: "il fait encore beau". Il est possible d'analyser ainsi chacune des occurrences listées.

Il nous semble que ce processus de découverte des sous-systèmes graphiques de l'orthographe contribue à une acquisition solide des connaissances nécessaires à la maîtrise de l'écrit.

Parallèlement au travail d'analyse graphique, l'observation peut aussi porter sur la fréquence des formes relevant des différentes catégories grammaticales: les formes [fe] du verbe et du nom relèvent toutes des listes LB1 et LB2 (code SL, informations de la dernière colonne à droite), contrairement aux deux dernières occurrences qui appartiennent à la liste LB3.

#### 6.1.4 Les adjectifs épïcènes

Si la conscience linguistique paraît bien établie sur la séparation des genres – féminin et masculin – pour les apprenants en FLM, une autre piste de réflexion peut être amorcée sur l'absence de marque de genre, lorsque la forme graphique se termine en -e. C'est le cas des adjectifs épïcènes dont voici quelques exemples.

FORMES-C	NAT	GE	NB	SL
difficile	AJ		S	2
difficiles8	AJ		P	2
pauvre	AJ		S	2
pauvres8=	AJ		P	2
facile	AJ		S	2
faciles8	AJ		P	2
nécessaire	AJ		S	2
nécessaires8	AJ		P	2
politique	AJ		S	3
politiques8	AJ		P	3

Fig. 9: Extraits du listage des adjectifs dont la colonne genre (GE) est vide

L'exercice de la commutation "une femme *pauvre* / un homme *pauvre*" valide l'absence de genre marqué. L'observation de la fréquence des formes indique que l'adjectif *politique*, liste LB3, est moins fréquent que les autres, liste LB2, ce que confirme aussi le sentiment linguistique de tout un chacun.

D'autre part, ces investigations peuvent être étendues à la sphère du nom. La notion de nom épïcène a facilité les changements récents liés à la féminisation des noms de métiers, titres, grades et fonctions (*FJTN*: 1999). Par exemple, *professeur* et *ministre* sont épïcènes phonétiquement parlant, mais graphiquement le genre féminin est indiqué par l'ajout d'un *-e*, marque de féminin, à *professeure* et d'une marque syntaxique par le déterminant féminin *une/la* devant le vocable *ministre*.

### 6.1.5 Les noms féminins terminés par -t et les noms masculins terminés par -e

On peut s'intéresser à un paradoxe: certains noms féminins sont terminés par -t, d'autres, masculins, le sont par -e. On constate que les premiers sont beaucoup moins nombreux que les seconds. On peut se poser encore d'autres questions (étymologie, difficulté de mémorisation du genre des noms masculins terminés par -e, par exemple le genre de *solde* qui a deux entrées dans le dictionnaire).

190	part7	698	plupart7	2335	minuit7
208	nuit7	1035	dent7		
345	mort7	1094	forêt7		

Fig. 10: Listes noms féminins terminés par -t

70	homme	678	verre	1250	ministère
268	compte	695	membre	1302	fleuve
271	exemple	702	théâtre	1344	capitaine
277	père	705	système	1503	mystère
300	livre	709	genre	1532	décembre
324	service	758	arbre	1563	kilomètre
329	maitre	787	terme	1586	programme
337	problème	789	pote	1644	musée
343	nombre	815	passage	1646	septembre
352	caractère	822	centre	1677	ange
354	être	859	commerce	1703	poste
356	ordre	884	crime	1781	comte
358	siècle	895	rêve	1795	souffle
361	peuple	899	ouvrage	1814	charme

Fig. 11: Extraits de la liste de noms masculins terminés par -e

Ce ne sont-là que quelques exemples de recherches que les apprenants, quels que soient leur niveau et leur âge, pourront effectuer selon tel ou tel critère. Autre possibilité: le maître ou le formateur proposera en vrac une liste de mots à classer sans autre consigne, pour amener l'élève à découvrir la structure de la langue. Si cet exercice est orienté vers une solution unique, la présence d'un intrus la rendra plus ludique. La base *OrthoFonic* est un outil dont chacun pourra découvrir les possibilités.



## 7. Les objectifs du projet

Le succès de la candidature PICRI serait le début d'un partenariat réunissant l'ensemble des acteurs pour la mise au point du didacticiel en question. Ce partenariat impliquerait, outre le CNRS-LAMOP et l'association RAPHAEL déjà cités, d'une part des étudiants inscrits en Master 2 de traitement automatique des langues, d'autre part des enseignants et leurs élèves. Les étudiants auraient à déployer une interface logicielle dans un langage le plus naturel possible pour les futurs utilisateurs. Avec le groupe de professeurs et directeurs d'écoles, serait établie une typologie des requêtes fondée sur leur expérience. On obtiendrait ainsi un outil de travail souple, capable de produire des sélections et des tris adaptés à des séquences didactiques déterminées. La mise au point du logiciel devrait être réalisée en relation avec les praticiens et leurs élèves. Pour ces derniers, la médiation de l'ordinateur, dans un travail en petits groupes, est un élément stimulant dont il faut tenir compte.

Ce programme concernerait des classes de CM1 et CM2 de l'école élémentaire et les classes de collège. Il bénéficierait notamment aux enfants en difficulté pour une meilleure intégration scolaire.

Le champ d'expérimentation, si l'on peut s'exprimer ainsi, concernerait en grande partie des classes situées en ZEP (zone d'éducation prioritaire), dans la commune de Villejuif, où la mixité sociale et les différences socioculturelles sont du plus grand intérêt pour une telle entreprise.

Par ailleurs, pourquoi ne pas envisager par la suite un autre volet destiné à des adultes (souvent parents d'enfants scolarisés), en exploitant les mêmes éléments de la base *OrthoFonic*, afin de développer un module d'alphabétisation?

## Conclusion

Le caractère novateur de ce projet de logiciel interactif et dynamique repose sur:

- la haute fréquence d'usage du vocabulaire proposé;
- le système de codage grammatical des lexèmes et le système de codage orthographique des formes facilitant la maîtrise des diverses fonctions des lettres muettes de l'orthographe;
- l'appropriation systématique de la relation oral - écrit.

*OrthoFonic* vise de la sorte à renouveler les modalités d'acquisition par une approche multidimensionnelle montrant les relations entre système graphique et système phonique, mais aussi les relations à l'intérieur de l'écrit entre les divers sous-systèmes de l'orthographe.

À terme, nous espérons disposer gratuitement d'un tel logiciel, qui favoriserait la démarche heuristique et valoriserait les connaissances de l'apprenant au fur et à mesure que ce dernier consolide ses acquisitions.

## Abréviations et sigles

API = Alphabet Phonétique International

CNRS = Centre National de la Recherche Scientifique

DHOF = Dictionnaire Historique de l'Orthographe Française (cf. Bibliographie: Catach, N. dir.)

FDWF = Frequency Dictionary of French Words (cf. Bibliographie: Juilland, A.)

FLE = Français Langue Étrangère

FLM = Français Langue Maternelle

FLS = Français Langue Seconde

FJTN = Femme, j'écris ton nom... (cf. Bibliographie: Becquer, A.)

LAMOP = Laboratoire de Médiévistique Occidentale de Paris

LOB = Listes Orthographiques de Base (cf. Bibliographie: Catach, N., Jecic, F. & équipe HESO)

PICRI = Partenariats Institutions-Citoyens pour la Recherche et l'Innovation

RAPHAEL = Recherche pour l'Application de la PHonologie aux Apprentissage de l'Écriture et de la Lecture

TLF = Trésor de la Langue Française

UMR = Unité Mixte de Recherche

## Bibliographie de l'article

Association Phonétique Internationale, Le maître phonétique, revue fondée par Paul Passy et éditée de 1904 à 1967.

Becquer, A., Cerquiglini, B., Cholewka, N., Coutier, M., Frécher, J. & Mathieu, M.-J. (1999): *Femme, j'écris ton nom... Guide d'aide à la féminisation des noms de métiers, titres, grades et fonctions*. Paris (CNRS-INaLF, La Documentation française).

Bulletins Officiels de l'Éducation nationale: n° 5, du 12 avril 2007 et n° 6 du 28 août 2008.

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

Catach, N., Jecic F. & équipe HESO (1984): *Les listes orthographiques de base du français*. Paris (Nathan).

Gougenheim, G., Michéa, R., Rivenc, P. & Sauvageot, A. (1956): *L'élaboration du français élémentaire*. Paris (Didier).

Gougenheim, G. (1971): *Le français élémentaire*. In: *Introduction du Grand Larousse de la langue française*, t. 1, p. LXXXII-XC. Paris (Larousse).

Jecic, F. (1987): *Pour un programme minimal de phonétisation automatique du français*. In: *Liaisons-HESO*, 15, 33-59.

Juilland, A. et al. (1970): *Frequency Dictionary of French Words*. The Hague (Mouton).

Martinet, A. (1960): *Éléments de linguistique générale*. Paris (Armand Colin).

Martinet, A. & J. (1980): *Dictionnaire de l'orthographe Alfonic*. Paris (SELAF, Peeters-France).

- Martinet, A. & J., Villard, J., Boyer, D. & Dominici, G. (1983): Vers l'écrit avec Alfonic. Paris (réédition 2006, Association RAPHAEL).
- Martinet, A. (1989): Alfonic au plus près de l'orthographe. In: Liaison Alfonic, fascicule 2, 7-12.
- Les Rectifications de l'orthographe (1990): Journal Officiel de la République française, Documents administratifs, 100 (06.12.1990).
- Skupien-Dekens, C., Kamber, A. & Dubois, M. (2010): Manuel d'orthographe pour le français contemporain. Institut de langue et civilisation françaises (Université de Neuchâtel).
- Ters, F., Mayer, G. & Reichenbach, D. (1968): Vocabulaire orthographique de base. Neuchâtel (Messeiller), Paris (OCDL).
- Ters, F. (1973): Orthographe et vérités. Paris (ESF).
- Ters, F., Mayer G. & Reichenbach D. (1975): L'échelle Dubois-Buyse d'orthographe usuelle française. Neuchâtel (OCDL, Messelie) [3e éd.].
- Trésor de la langue française, Dictionnaire de la langue du XIXe et du XXe siècle (1789-1960), 1971-1994, publié sous la direction de Paul Imbs puis de Bernard Quemada. Paris (CNRS-Klincksieck-Gallimard).
- Vial, J. & Villard, J. (1986): À l'entrée au cours élémentaire. In: Liaison Alfonic, fascicule 1, 1-2.

## Bibliographie complémentaire

- Baddeley, S. & Jecic, F. (2010): Autour des rectifications de l'orthographe, aspects historiques et actuels. De la "droicte escripture" aux Rectifications de 1990. In: Rectification à l'oral et à l'écrit, Paris (coll. L'Homme dans la langue, Ophrys), 191-206.
- Biedermann-Pasques, L. & Jecic, F., eds. (2006): Les rectifications orthographiques de 1990. Analyses des pratiques réelles. In: Cahiers de l'Observatoire des pratiques linguistiques, 1, Orléans, (Délégation générale à la langue française et aux langues de France – Presses Universitaires d'Orléans).
- Biedermann-Pasques, L., & Jecic, F. (2005): "Orthographe". In: Encyclopaedia Universalis, site abonnés internet: <http://www.universalis.fr>.
- Catach N., Gruaz, C. & Duprez, D. (1980, 2e éd. 1984, 3e éd. 1995): L'orthographe française, traité théorique et pratique. Paris (Nathan).
- Dictionnaire de l'Académie française (9e édition en cours), t. 1, A-Enzyme, 1992. Paris (Imprimerie Nationale); t. 2, Écène-Mappemonde, 2000, Paris (Fayard, Imprimerie Nationale); t. 3, Maquereau-Quotité, 2011 Paris (Fayard, Imprimerie Nationale); la suite du Dictionnaire paraît en fascicules, dernier paru en 2011, de *Quadru-* à *Raidisseur*, Paris Imprimerie Nationale, Documents Administratifs du Journal Officiel de la République Française.
- Fondet, C. (1990): Vaincre l'illettrisme. Paris (Édition Science et service Quart Monde).
- Martinet, A. & Walter, H. (1973): Dictionnaire de la prononciation française dans son usage réel. Paris (France-Expansion).
- Martinet, A., Martinet, J. et al. (1983): Vers l'écrit avec Alfonic. Paris (Hachette), rééd. Paris (RAPHAEL, 2006).
- Taulelle, D., Rey-Debove, J., Boumendil-Lucot, A., Lechheb, S. et al. (1989): Le Robert Oral-Écrit. Paris (Dictionnaires Le Robert).

## Sites internet

<<http://www.inforef.be/projets/jeparledoncjecris/index.html>>, accès libre au didacticiel "Je parle donc j'écris".

<<http://www.vjf.cnrs.fr/orthographe>>, accès aux documents officiels concernant les rectifications de l'orthographe française et à la bibliographie thématique.