

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/322770523>

A learning progression based on statistics on the French orthography (from Manulex–Morpho): For typical and atypical beginning readers

Article in ANAE - Approche Neuropsychologique des Apprentissages chez l'Enfant · October 2017

CITATIONS

0

READS

2,177

1 author:



[Liliane Sprenger-Charolles](#)

Aix-Marseille Université

175 PUBLICATIONS 5,061 CITATIONS

SEE PROFILE

Une progression pédagogique construite à partir de statistiques sur l'orthographe du français (d'après Manulex-Morpho) : pour les lecteurs débutants et atypiques

L. SPRENGER-CHAROLLES*,**

* CNRS-Laboratoire de Psychologie cognitive (LPC, UMR 7290), Marseille, France.

** Brain Language Research Institute (BLRI), Aix-Marseille-Université, France.

RÉSUMÉ : Une progression pédagogique construite à partir de statistiques sur l'orthographe du français (d'après Manulex-Morpho) : pour les lecteurs débutants et atypiques

Pour comprendre ce qu'on lit dans une écriture alphabétique, il faut avoir maîtrisé le décodage, c'est-à-dire les correspondances graphème-phonème (CGPh). C'est l'automatisation de cette procédure de lecture qui va permettre au débutant de constituer son lexique orthographique et, grâce à cela, de comprendre progressivement ce qu'il lit comme il comprend ce qu'il entend. Le décodage doit donc être enseigné systématiquement en tenant compte de la fréquence des CGPh dans les mots de la langue cible et, surtout, de leur consistance. En effet, cette dernière a un fort impact sur l'apprentissage de la lecture, les enfants apprenant mieux et plus vite à lire quand l'orthographe de la langue cible est consistante (en espagnol, et même en français, au moins pour la lecture) que lorsqu'elle ne l'est pas (en anglais). Les statistiques sur la fréquence et la consistance des CGP en français utilisées pour construire la progression présentée dans cet article sont issues de Manulex-Morpho, base de données élaborée à partir de Manulex. Cette dernière contient environ 2 000 000 de formes non lemmatisées (c'est-à-dire, tous les mots, y compris ceux avec des marques morphologiques) contenues dans 54 manuels du primaire. Le corpus de Manulex-Morpho comprend les entrées lexicales les plus fréquentes de Manulex, ce qui correspond à 98 % de celles incluses dans les manuels examinés. L'article présente les caractéristiques majeures de la base de données Manulex-Morpho ainsi que la progression proposée. Il est destiné non seulement aux personnes en charge de l'apprentissage de la lecture chez les lecteurs typiques et atypiques, mais aussi à leurs formateurs ainsi qu'aux concepteurs de manuels scolaires.

Mots clés : Orthographe – Correspondance graphème-phonème – Méthode d'apprentissage de la lecture – Morphologie.

SUMMARY: A learning progression based on statistics on the French orthography (from Manulex-Morpho): for typical and atypical beginning readers

In an alphabetic system, a child must have mastered decoding skills (i.e. be able to use grapheme-phoneme correspondences, GPC) to understand what s/he is reading. It is the automatization of this reading procedure that allows beginning readers to build up an orthographic lexicon and gradually understand what they read the same way they understand what they hear. Decoding should therefore be systematically taught based on the frequency of GPC in the words of the target language and, above all, on their consistency. Indeed, the latter has a strong impact on learning to read, children learning better and faster to read when the spelling of the target language is consistent (in Spanish, and even in French, at least for reading), than when it is not (in English). The statistics on the frequency and consistency of the French CGPs used to construct the progression presented in this article are taken from Manulex-Morpho, a database developed from Manulex. The latter includes about 2,000,000 non-lemmatized word forms (i.e., all words, including those with morphological markers) contained in 54 primary textbooks. The Manulex-Morpho corpus includes Manulex's most frequent lexical entries, which corresponds to

Pour citer cet article : SPRENGER-CHAROLLES, L. (2017). Une progression pédagogique construite à partir de statistiques sur l'orthographe du français (d'après Manulex-Morpho) : pour les lecteurs débutants et atypiques. *A.N.A.E.*, 148, 247-256.

98% of those from the textbooks examined. In this paper, Manulex-Morpho's main characteristics are presented along with the proposed learning progression. It is intended not only the staff in charge of reading instruction with typical and atypical readers, but also their mentors and reading methods developers.

Key words: Orthography – Spelling – Grapheme-phoneme-correspondences – Phonics – Morphology.

RESUMEN: *Una progresión pedagógica construida a partir de estadísticas acerca de la ortografía del francés (según Manulex-Morpho): para los lectores principiantes y atípicos*

Para comprender lo que se lee en escritura alfabética, es necesario dominar la descodificación, es decir, las correspondencias grafema-fonema (CGF). La automatización de este procedimiento de lectura permitirá al principiante constituir su léxico ortográfico y, gracias a ello, comprender progresivamente lo que lee al igual que comprende lo que oye. Por consiguiente la descodificación se debe enseñar sistemáticamente teniendo en cuenta la frecuencia de CGF en las palabras de la lengua diana y, sobre todo, de su coherencia. En efecto, este último punto tiene un fuerte impacto en el aprendizaje de la lectura, los niños aprenden a leer mejor y más rápidamente cuando la ortografía de la lengua diana es coherente (como en español, e incluso en francés, al menos en la lectura) que cuando no lo es (como en inglés). Las estadísticas sobre la frecuencia y la coherencia de las CGF en francés utilizadas para construir la progresión que se presenta en este artículo provienen de la base de datos Manulex-Morpho, que se elaboró a partir de la base Manulex. Esta última contiene alrededor de 2.000.000 de formas no lematizadas (es decir, todas las palabras, incluidas aquellas que tienen marcas morfológicas) incluidas en 54 manuales de educación primaria. El cuerpo de la base de datos Manulex-Morpho comprende las entradas lexicales más frecuentes de Manulex, lo que corresponde al 98 % de las entradas incluidas en los manuales examinados. En el artículo se presentan las características principales de la base de datos Manulex-Morpho, así como la progresión propuesta. Está destinado no únicamente a personas responsables del aprendizaje de la lectura de lectores típicos y atípicos, sino también a sus formadores y a creadores de manuales escolares.

Palabras clave: Ortografía – Correspondencia grafema-fonema – Método de aprendizaje de la lectura – Morfología.

Pour comprendre ce qu'on lit dans une écriture alphabétique, il faut avoir maîtrisé le décodage, c'est-à-dire les relations entre les unités de base de l'écrit (les graphèmes : 'f' 'ou' et 'r' dans le mot « four ») et les unités correspondantes de l'oral, les phonèmes¹ (/f/, /u/ et /R/ dans ce même mot, cf. Gentaz, Sprenger-Charolles & Theurel, 2015). C'est l'automatisation de cette procédure de lecture qui va permettre au débutant de constituer son lexique orthographique et, grâce à cela, de comprendre progressivement ce qu'il lit comme il comprend ce qu'il entend. Le décodage doit donc être enseigné systématiquement en tenant compte de la fréquence des correspondances graphème-phonème (CGPh) dans les mots de la langue cible et, surtout, de leur consistance. En effet, cette dernière a un fort impact sur l'apprentissage de la lecture, les enfants apprenant mieux et plus vite à lire quand l'orthographe de la langue cible est consistante (en espagnol, et même en français, au moins pour la lecture) que lorsqu'elle ne l'est pas (en anglais, cf. Share, 2008 ; voir Ziegler & Goswami, 2005 pour une synthèse en anglais et Sprenger-Charolles & Colé, 2013 pour une synthèse en français). Les caractéristiques de l'orthographe du français analysées dans cet article sont issues de Manulex-Morpho (Peereman, Sprenger-Charolles & Messaoud-Galusi, 2013), base de données élaborée à partir de Manulex (Lété, Sprenger-Charolles & Colé, 2004) qui prend en compte la morphologie, principale source de difficultés de notre orthographe (Fayol, 2013).

L'article comporte deux parties. La première présente la base de données Manulex-Morpho ainsi que des statistiques sur l'orthographe du français issues de cette base. La seconde porte sur leur utilisation pour guider l'élaboration d'une progression « idéale » dédiée à l'apprentissage de la lecture, ce qui est essentiel non seulement pour les lecteurs débutants, mais aussi dans le cadre d'un programme de rééducation.

PRÉSENTATION DE LA BASE DE DONNÉES MANULEX-MORPHO

L'originalité de Manulex-Morpho est la prise en compte de la morphologie, principalement celle des flexions. Le corpus comprend les items non lemmatisés (toutes les formes lexicales, y compris celles avec des marques morphologiques) les plus fréquentes contenues dans Manulex. Cette dernière présente la fréquence lexicale d'environ 2 000 000 de formes non lemmatisées contenues dans 54 manuels du primaire. Manulex-Morpho a retenu les formes lexicales les plus fréquentes de Manulex (10 000 sur 50 000), soit 20% des entrées lexicales mais 98 % des entrées textuelles de cette base.

¹ Le phonème est l'unité sonore qui permet de différencier deux mots dans une langue donnée. Ainsi, /b/ et /v/ sont deux phonèmes en français, qui distinguent « bol » de « vol », mais pas en espagnol. En revanche, le 'r' simple et le 'r' roulé sont deux phonèmes en espagnol, qui distinguent « pero » (mais) de « perro » (chien), mais pas en français. C'est au cours de la première année de la vie que notre système phonologique va progressivement se limiter aux phonèmes pertinents pour traiter la langue parlée dans notre environnement linguistique.

Dans Manulex-Morpho, la consistance des CGPh et des CPhG a été calculée à partir de toutes les formes incluses dans le corpus (« chant », mais aussi « chanter » - « chantes » - « chants »). La notation morphologique a été effectuée pour les graphèmes qui notent quatre types de marques. Il s'agit des flexions nominales (genre et nombre : 'e', 's', 'x', cf. « quatre amies » et « les jeux ») et verbales (personne et temps : 'es' et 'ent' de « tu chantes » et « ils chantent », 'a' et 'ait' de « il marcha » et « il marchait »). Il s'agit également du morpho-phonogramme 'ent' des adverbes en 'ment' (rarement, vraiment) ainsi que des consonnes finales qui sont silencieuses dans le mot racine, mais prononcées dans les formes fléchies et/ou dérivées (par exemple, le 'd' de « grand » sur la base duquel sont construits les mots « grande » et « grandeur »).

La base de données Manulex-Morpho fournit la fréquence et la consistance des CGPh et des CPhG rencontrées dans ce corpus ainsi que, pour les mots, en plus de leur fréquence lexicale, leur consistance. Cette dernière a été calculée à partir des CGPh ou des CPhG contenues dans chacun des mots. Plusieurs variantes de Manulex-Morpho sont disponibles. La première concerne l'index de fréquence d'usage des mots utilisés et celui de leur consistance. Ces statistiques, pondérées par les valeurs de fréquence textuelle (rapportées pour un million de mots, à des fins de comparaison avec d'autres corpus), sont données par deux indices : « F » et « U ». Ce dernier prend non seulement en compte la fréquence des mots dans les manuels scolaires du corpus, mais également le nombre de manuels différents dans lesquels ils sont rencontrés. Pour chacune de ces deux variantes, les calculs ont été générés en fonction de la distinction - ou non - de certaines voyelles, celles qui correspondent aux archiphonèmes /A/ (/a/-/A/), /œ/ (/ø/-/œ/), /E/ (/e/-/E/) et /O/ (/o/-/O/), qui sont surtout des variantes contextuelles, dépendant de la structure de la syllabe, fermée ou ouverte, dans laquelle elles se rencontrent (Delattre, 1965)².

Cinq fichiers (.zip Excel) sont disponibles sur le site Manulex-Morpho³. Quatre donnent la fréquence et la consistance des mots de la base : deux en fonction de l'indice de fréquence U et deux en fonction de l'indice F, dans chacun des cas, un avec les voyelles non distinguées et un avec les voyelles distinguées. Le dernier fournit la fréquence et la consistance des CGPh et des CPhG, avec les mêmes variantes.

STATISTIQUES D'ENSEMBLE

Les calculs sont basés sur le nombre de fois qu'un graphème se prononce d'une certaine façon dans trois

² Léon et Léon (1964) conseillent, à des fins pédagogiques, de privilégier un système avec ces 4 voyelles orales. C'est également le choix qu'a fait Catach (1973, pp. 35-36).

³ Accessible sur le site du LPNC, laboratoire CNRS de l'Université de Grenoble : http://lpnc.univ-grenoble-alpes.fr/resources/ronald_peereman/Manulex_morpho/indexfr.html

positions : début, milieu et fin de mot. Par exemple, ‘s’ en fin de mot ne se prononce pas, sauf dans de rares items dans lesquels il peut correspondre à une marque morphologique (ours) ou n’être qu’un simple graphème (bus). Quel que soit le cas de figure, c’est le ‘s’ muet qui prédomine (85.7 % des cas pour celui qui est un morpho-phonogramme et 90.1 % le ‘s’ non motivé). En début de mot, ce graphème se prononce toujours /s/, alors que, en milieu de mot, il se prononce soit /s/ (50.7 % : triste, ourson, danse), soit /z/ (49.3 % : rose, arrosant, arrosoir, chinoise). La seconde prononciation correspond au contexte dans lequel ce graphème se trouve entre deux voyelles ou semi-voyelles⁴, sauf lorsque la voyelle qui précède est une nasale, qui se termine par une consonne graphique (danse).

Il ressort des analyses (cf. *le tableau 1*) une forte asymétrie entre CGPh et CPhG, les premières étant les plus régulières dans tous les cas : place dans le mot, prise en compte ou non de la morphologie. Autre fait marquant : l’asymétrie de consistance entre CGPh et CPhG, très marquée en fin de mot quand la morphologie n’est pas prise en compte (12 %, soit 78 % versus 66 %), diminue fortement quand elle l’est (5 %, soit 80 % versus 75 %). De plus, ce sont surtout les CPhG qui bénéficient de cette prise en compte : le gain est de pratiquement 10 % (de 66 à 75 %), contre 2 % pour les CGPh (de 78 à 80 %).

L’asymétrie entre CGPh et CPhG s’explique par le fait que, si certains graphèmes se lisent toujours de la même façon (comme ‘o’, ‘au’, ‘eau’), pour écrire le phonème correspondant, il faut choisir entre différents allographes. Cette asymétrie est également due, comme attendu, à la morphologie. Pour ces deux raisons, apprendre à lire est plus facile qu’apprendre à écrire en français. C’est ce que plusieurs publications ont montré, certaines contenant des données longitudinales du milieu du CP à la fin du CM1 sans prise en compte de la morphologie (Sprenger-Charolles, Siegel, Béchennec & Serniclaes, 2003) ; d’autres ayant intégré la morphologie (Fayol, 2013).

Les statistiques présentées dans le *tableau 1* indiquent que l’orthographe du français est relativement consistante, tout au moins pour la lecture. Cela est d’autant plus vrai que les données prises en compte pour établir ces statistiques n’ont pas intégré les règles contextuelles (p. ex. pour la prononciation de « s » ou du « c-g »).

AUTRES ANALYSES POUR LES CORRESPONDANCES GRAPHÈME-PHONÈME

Correspondances graphème-phonème les plus fréquentes et consistantes

Le *tableau 2* présente les CGPh les plus fréquentes et les plus consistantes par ordre de fréquence (fréquence textuelle U par million de mots). Les seuils choisis pour dire d’une association qu’elle est fréquente est supérieur ou égal à 5 000 occurrences et, pour la consistance, supérieur ou égal à 95 %.

Toutes les voyelles sont transcrites par au moins un graphème fréquent et consistant, sauf ‘i’, ‘u’ et ‘e’. C’est également le cas pour presque toutes les consonnes, sauf ‘gn’, ‘z’ et ‘k’, qui sont consistants mais rares (fréquence 2.728, 713 et 396), ‘c’ et ‘g’, qui sont fréquents (65 600 et 21 889), mais peu consistants (64.3 % et 48.4 % pour la prononciation /k/ et /g/), tout comme ‘s’ et ‘t’ (fq = 208 403 pour ‘s’, consistance pour ‘s’-/s/ = 32.4 % ; fq = 139.824 pour ‘t’, consistance pour ‘t’-/t/ = 71 %). Seule une des trois semi-voyelles est à la fois fréquente et consistante, le ‘oi’ de « loi », pas le ‘u’ de « lui » ni le « yod » (‘i’ de « ciel », ‘y’ de « paye », ‘ill’ de « paille », ‘ll’ de « fille »). En dehors de ces cas, 3 lettres de l’alphabet ne figurent pas dans le *tableau 2* : ‘w’ (qui est très rare, 50), ‘h’ (qui n’est pas un phonème en français) et ‘x’ qui correspond à plusieurs phonèmes (axe, examen, dix) ou est muet (deux, jeux, tu peux, mieux).

Les lettres géminées

On ne trouve pas de lettres géminées en français pour les voyelles (sauf dans quelques mots étrangers, cf. « foot »). Contrairement aux voyelles, pratiquement toutes les consonnes peuvent être doublées, sauf ‘v’ et ‘z’ (excepté dans quelques mots étrangers comme « pizza » et « puzzle »), ‘j’, ‘q’ et ‘x’. Celles qui sont le plus souvent doublées et qui ont des CGPh consistantes sont ‘s’ (fréquence > 11 800) ainsi que ‘t’, ‘m’, ‘r’ et ‘n’ (fréquence entre 6 800 et 5 280). Le double ‘ll’ est fréquent (13 732) mais peu consistant : il est prononcé /l/ dans 87.7 % des cas (ville) et /j/ dans 12.3 % (fille).

Tableau 1. Consistance des CGPh et CPhG en français en fonction de la position dans le mot et de la prise en compte (ou non) des morpho-phonogrammes (d’après Peereboom *et al.*, 2013).

Position dans le mot	CGPh (Lecture)		CPhG (Écriture)	
	Sans morphologie	Avec morphologie	Sans morphologie	Avec morphologie
Début	96 %	96 %	87 %	87 %
Milieu	74 %	75 %	72 %	72 %
Fin	78 %	80 %	66 %	75 %

⁴ Toutefois, dans cette position, la prononciation /s/ est attestée dans quelques mots composés issus de « soleil » (parasol...).

Tableau 2. CGPh les plus fréquentes (Fq textuelle U > 5.000) et les plus consistantes (> à 95 %).

Voyelles				Consonnes			
CGPh	Exemples	Fréquence	Consistance	CGPh	Exemples	Fréquence	Consistance
Voyelles orales fréquentes et consistantes				Consonnes fréquentes et consistantes			
a=/A/	La-Bal	143.280	99.9	r(r)=/R/	Rire (<i>Barre</i> = 5.720 ; 100 %)	171.890	99.7
ou=/u/	Fou	50.024	96.8	l=/l/	Lilas	160.533	99.7
o=/O/	Do-Bol	49.707	99.9	p=/p/	Pipe	81.615	94.6
é=/E/	Été	38.875	100.0	m(m)=/m/	Maman (<i>Pomme</i> = 6.057 ; 100 %)	58.937	99.9
eu=/E/	Deux-Neuf	23.315	97.6	v=/v/	Vivre	52.447	100.0
				n(n)=/n/	Nominal (<i>Année</i> = 5.284 ; 100 %)	45.517	100.0
Voyelles nasales fréquentes et consistantes				qu=/K/	Quelque	29.668	99.9
on=/ô/	Oncle	33.125	97.2	b=/b/	Barbe	28.564	98.3
an=/â/	Tante	29.147	99.9	f=/f/	Se faufler	27.396	99.2
un=/û/	Un	19.074	100.0	ch=/ʃ/	Cherche	18.614	99.2
in=/î/	Fin	8.310	100.0	j=/ʒ/	Jour-Séjour	16.507	99.9
				d=/d/	Dinde (Grand = 7.580 ; Fq = 7 %)	100.406	93.0
Allographes fréquents et consistants des voyelles				Cas particuliers : les graphèmes 's', 'ss' et 't'			
au=/O/	Auto-Paul	17.003	100.0	s=/s/	Sel-Sol (début de mot)	47.867	100.0
eau=/O/	Eau	6.102	100.0	ss=/ss/	Passé	11.826	100.0
ai=/E/	Fait-Faire	22.816	98.0	s=/s/	Quelle que soit la position	67.552	32.4
et=/E/	Effet	20.290	98.0	s=/z	Rose	14.408	6.9
è=/E/	Lève	11.773	100.0	s=/#	Amis, Tu parles, Gris, Radis	126.443	60.7
ê=/E/	Arrêt-Arrête	5.932	100.0				
Cas particuliers : les graphèmes 'e', 'i' et 'u'				t(t)=/t/	Tarte (<i>attendre</i> = 6843 ; 100 %)	99.257	71.0
e=/ə/	Le-Une (fin de mot)	218.004	96.4	t=/#/	Droit	(37.784)	(27.0)
e=/#	Vie (fin de mot)	8.257	3.6	t=/s/	Addition	(2.783)	(2.0)
i=/i/	Lire (Ciel=24.746)	97.911	79.2	Semi voyelles fréquentes et consistantes			
u=/y/	Lune (Lui=11.318)	51.248	81.9	oi=/wa/	Loi	20.725	100.0

Les lettres muettes

Les lettres muettes sont notées soit par le signe '#' (le 's' de « amis » ou le 'e' de « vue ») soit par 'e' (le 'e' de « gare » ou de « table »). Le *tableau 3* présente les voyelles et consonnes les plus fréquentes qui ne se prononcent pas (fréquence textuelle U > à 5.000).

À la différence du 'e' de « gare » ou « porte », qui peut ou non se prononcer et qui est très fréquent (plus de 266 000 occurrences), le 'e' de « amie » ou de « il joue » ne se prononce jamais (9 700 occurrences). Il est à noter que ce 'e' est un morpho-phonogramme dans environ 56 500 occurrences. Le 'ent', qui est toujours un morpho-phonogramme, a une fréquence bien plus faible (5 554 occurrences). Lorsqu'il s'agit du pluriel d'un verbe, soit il se prononce comme le 'e' de « gare » ou de « table » (ils parlent, ils livrent, 3 754 occurrences), soit il disparaît (comme dans « ils jouent », 1 800 occurrences). En revanche, il se prononce /â/ dans les adverbes en 'ment', qui sont le plus souvent formés à partir d'un adjectif au féminin (actuellement, clairement, longuement...).

Pour les consonnes, les graphèmes muets les plus fréquents sont tous des morpho-phonogrammes. C'est le cas pour le 's' de « petits », « tu chantes » et « gris ». C'est également le cas pour le 't' de « il finit » et, surtout, pour celui de

« toit » ainsi que pour le 'x' de « jeux », « tu peux » et « deux ». C'est enfin le cas du 'd' (cf. « rond » et « il prend »).

En dehors de ces cas, qui concernent presque tous la fin d'un mot (sauf dans quelques items, comme « lesquels »), une autre lettre ne se prononce pratiquement jamais en français : le 'h' (heure, bonheur, Sahara...). Il y a toutefois des exceptions, avec un 'h' aspiré dans « les héros » (pour ne pas confondre avec « zéro »), et « hauteur » (pour ne pas confondre avec « auteur »)...

PRINCIPES POUR ENSEIGNER LES CGPH⁵

La fréquence des mots est un facteur important dans l'apprentissage de la lecture, les mots les plus fréquents étant supposés connus des enfants. Il est à noter que certains mots très fréquents contiennent des CGPh peu régulières. C'est le cas pour de nombreux mots « fonction » (avec, et/ cet, pendant) ainsi que pour certains verbes (il est grand, il a eu...), et certains noms (dix, sept, femme, monsieur...). Ces mots devraient être introduits très tôt dans la mesure

⁵ Ces principes reprennent partiellement ceux présentés dans Dehaene *et al.* (2011), avant la mise en place de la base de données Manulex-Morpho (Peereboom *et al.*, 2013).

Tableau 3. Lettres muettes les plus fréquentes (> à 5000, en gras, fréquence > 15 000).

Exemples (différentes prononciations du graphème)	Total pour le graphème	Total « muet »	Graphème muet 'e': /ə/ ou /#/	Total par catégorie			
				Flexion nominale	Flexion verbale	Support de flexion/dériv	Non motivé
'e' : <u>se</u> <u>l</u> , <u>fi</u> <u>n</u> e, <u>ga</u> <u>r</u> e...	360.358	266.114 9.710	'e'-> ə 'e'-> #	32.522 (<u>brune</u>) 7.464 (<u>amie</u>)	16.518 (je <u>chan</u> te) (je <u>joue</u>)		217.074 (<u>ga</u> re) (<u>lycée</u>)
'ent' : <u>vrai</u> ment, ils <u>par</u> lent (<u>joue</u> nt)	9.758	5.554	'ent'-> ə 'ent'-> #		3.754 (ils <u>vol</u> ent) 1.800 (ils <u>joue</u> nt)		
's' : <u>so</u> <u>l</u> , <u>ami</u> s, <u>gris</u> ...	208.403	126.443	's'-> #	79.198 (<u>ami</u> s)	(tu <u>chan</u> tes)	4.950 (<u>gris</u>)	41.211 (<u>dans</u>)
'x' : <u>axe</u> , <u>taxi</u> , <u>deux</u> ...	15.492	10.955	'x'-> #	5.349 (<u>jeu</u> x)	(tu <u>peu</u> x)	(<u>deux</u>)	(<u>mieu</u> x)
't' : <u>toi</u> t, il <u>fini</u> t...	139.841	37.784	't'-> #		8.429 (il <u>fini</u> t)	19.484 (<u>toi</u> t)	9.871 (<u>habi</u> t)
'd' : <u>dos</u> , <u>rond</u> , <u>froid</u> ...	107.986	7.580	'd'-> #		(il <u>prend</u>)	4.581 (<u>rond</u>)	(<u>fou</u> lard)
'h' : <u>heu</u> re, <u>sah</u> ra...	6.898	6.800	'h'-> #				

où, sans eux, il n'est pas possible de construire des phrases. Toutefois, la progression pédagogique doit surtout tenir compte de la fréquence et de la régularité des CGPh (et non de celle des CPhG qui sont moins consistantes que les CGPh, cf. le *tableau 1*). Il faut également considérer d'autres spécificités de l'orthographe de la langue dans laquelle s'effectue l'apprentissage : pour le français, les lettres muettes.

Dans les premiers mois de l'apprentissage

Dans les premiers mois de la première année du primaire, entre septembre et fin décembre, voire fin janvier, il faudrait présenter aux élèves les graphèmes les plus fréquents qui se lisent presque toujours de la même façon (cf. le *tableau 2*). C'est le cas, par exemple, non seulement pour « b » et « a » (d'où le B.A.BA), mais également pour « ou » et « ch ».

Voyelles et consonnes les plus fréquentes et consistantes

Les voyelles pouvant être introduites en premier sont 'a', 'é', 'o', 'eu' et le 'e' de « le », ainsi que 'i' et 'u'. Ces voyelles sont en effet toutes fréquentes et ont une prononciation consistante, sauf les deux dernières lorsqu'elles sont suivies par une voyelle (ciel, lui). En conséquence, elles ne seront jamais présentées dans ce contexte dans les premières séquences de la progression. On peut introduire par la suite les voyelles nasales fréquentes ayant une prononciation régulière (an, on, in, un) ainsi que les autres orthographes très fréquentes du /E/ ('ai', 'et', 'è' et 'ê'). Pour les consonnes, on peut introduire au début de l'apprentissage les fricatives 'f', 'j', 's', 'ch', 'v'⁶, parce qu'elles se lisent pratiquement toujours de la même façon et que leur reconnaissance en isolat est plus facile que celle des occlusives (t, d...) étant donné qu'elles permettent produire un son continu (fff, sss...). Le 's' ne devrait toutefois être présenté au début de l'apprentissage que lorsqu'il est à l'initiale d'un mot. En effet, c'est seulement dans cette position qu'il se prononce toujours de façon identique. Les autres consonnes fréquentes ayant une

prononciation régulière pouvant être intégrées au début de l'apprentissage sont 'l' et 'r' qui permettent de construire des mots ayant une structure syllabique complexe (bleu, tri, brique, blanche, fleur...) ainsi que, par la suite, 'p', 'b', 'm', 'n', 'd', 't' (sauf pour ce dernier dans des mots comme « addition ») et 'qu' (qui est dans de nombreux mots fonctions : qui, que, quoi, quel...).

Autres principes à prendre en compte

Durant une première période, les mots utilisés devraient être des monosyllabiques composés d'une syllabe simple, avec une voyelle (V) et une consonne (C) à l'initiale (ou, la, sa, jeu, feu, fou), les items CVC étant introduits dans un second temps (bal, bol, par, sol, sur, jour, pour, leur). Par la suite, des mots CCV (tri, blé, bleu) et CCVC (fleur, pleur) peuvent être présentés ainsi que, afin de focaliser l'attention des enfants sur l'importance de l'ordre des graphèmes dans le décodage, des couples CV-VC (tu-ut, fou-ouf) et CVC-CCV (tir-tri, dur-dru).

En outre, en raison de la fréquence des lettres finales muettes en français, les mots utilisés dès les débuts de l'apprentissage de la lecture peuvent se terminer par une lettre muette fréquente, comme 'e' (fée, amie), 's' (jus, souris) ou 't' (debout, front). Dans les premières séquences, il est possible de les signaler par une couleur spéciale.

Les autres principes les plus importants qu'il est souhaitable de respecter au début de l'apprentissage sont qu'il ne faudrait utiliser que des graphèmes déjà vus et ce dans des items courts (une à deux syllabes), qui sont, autant que faire se peut, des mots connus des enfants et qui, en outre, permettent de construire des énoncés. À cette fin, il est nécessaire d'introduire rapidement quelques mots irréguliers très fréquents (articles, pronoms et d'autres mots fonctions : les, des, elle, avec) ainsi que les formes très fréquentes de certains verbes irréguliers (« est » du verbe « être »).

Un exemple

Une progression possible pour un module « de base » pour l'apprentissage de la lecture en français est présentée dans l'*encadré 1* ci-dessous. Quelques illustrations se trouvent dans l'article sur Grapho-Game. Cette progression intègre également une série d'exercices spécifiques permettant d'évaluer l'intégrité du système phonologique des enfants ainsi que de leur système visuel.

⁶ Pas le 'z', qui est rare et en concurrence avec le 's' intervocalique.

Encadré 1. Un exemple de progression possible pour les premières étapes de l'apprentissage (Correspondances graphème-phonème les plus fréquentes et les plus consistantes).

<p>Séquence 1 :</p> <ul style="list-style-type: none"> = Voyelles orales : 'a', 'i', 'o', 'u', 'é', 'ou', 'eu' et le 'e' en fin de mot après une consonne (le, me, rire) = Consonnes continues : 'f', 'j', 'l', 'r' (feu, fou, jeu, jour, le, leur, ré, rire...) et le 's' en début de mots (sol, sale...) = Mots monosyllabiques consonne-voyelle (V et CV : ou, la, fou, feu), ainsi que quelques CVC (sol, four, jour)
<p>Séquence 2 :</p> <ul style="list-style-type: none"> = Les consonnes 'p', 'm', 'd', 'v' (peu, ma, dur, vol...) puis 'b', 'n', 'ch' (bol, neuf, char) et 't' (sauf dans des mots comme 'addition') = Introduction de mots avec des structures syllabiques CCV et CCVC (pli, bleu, pleur, fleur) = Mots très fréquents contenant des CGPh irrégulières (ou pas encore enseignées) permettant de construire des énoncés : <ul style="list-style-type: none"> - articles, pronoms et autres mots fonctions : 'un', 'au', 'et', 'elle', 'les', certains pouvant être introduits à la fin de la séquence 1 - formes très fréquentes des verbes irréguliers : il est, tu es, il fait... = Graphèmes muets fréquents de fin de mots (certains pouvant être introduits dans la séquence 1) : 'e' (amie, fée), 's' (dos), 't' (lit)
<p>Séquence 3 :</p> <ul style="list-style-type: none"> = Les voyelles nasales : 'an' (an, chante...), 'on' (on, bon...), 'in' (pin, lapin...) et 'un' (un, lundi...) = Les autres orthographes fréquentes du 'e' : 'ai', 'et' ainsi que les deux autres /E/ avec accent ('è' et 'é') = La consonne 'qu' (qui, que, quoi, lequel...)
<p>Exercices spécifiques sur les confusions visuelles et phonologiques (/g/ et /z/ n'ont pas été introduits)</p> <ul style="list-style-type: none"> - Confusions phonologiques (différence : 1 trait phonologique) Occlusives: /p/-/t/, /t/-/k/, /t/-/d/; Fricatives: /f/-/s/, /s/-/ʃ/, /ʃ/-/v/, /ʃ/-/ʒ/ - Confusions phonologiques et visuelles (différences : 1 trait phonologique et 1 visuel): /b/-/d/, /p/-/b/ - Confusions visuelles (différence : 1 trait visuel): n-u, p-q, d-q, i-j, v-y, h-n

Dans une seconde étape de l'apprentissage

Dans une seconde étape, il faut présenter les CGPh consistantes qui sont moins fréquentes, ou plus difficiles à lire, que celles incluses dans le module précédent, par exemple, les semi-voyelles 'oi' de « loi » et 'u' de « lui ». Il faudrait aussi intégrer rapidement le 'h', qui, sauf dans quelques cas (liaisons, en particulier) ne se prononce pas. Il est nécessaire d'aborder ensuite les principales difficultés de l'orthographe du français, la plupart posant surtout des problèmes pour l'écriture.

Les graphèmes qui s'écrivent ou se lisent différemment en fonction du contexte

Une première série de problèmes vient des mots contenant des graphèmes ayant, en fonction du contexte, une prononciation différente, par exemple, les deux 's' de « sosie », tout comme les deux 'c' de « cycle » et les deux 'g' de « garage ». Il en est de même pour ceux qui, également en fonction du contexte, doivent s'écrire d'une autre façon, comme le 'n' des voyelles nasales devant 'p' et 'b' (santé, chambre, ampoule ; conte, sombre, compte ; dinde, timbre, impossible...). Dans ces différents cas, l'explicitation de règles peut faciliter l'apprentissage.

Les graphèmes qui ont des allographes (c'est-à-dire plusieurs orthographes)

Une seconde série de problèmes est due au fait que certains phonèmes ont plusieurs orthographes comme /s/ (saucisse), /E/ (é, e, è, ê, ai, et...), /O/ (auto, peau), /i/ (pin, pain, faim, plein). Une autre, de même nature, concerne un phonème plus complexe, le yod, qui peut s'écrire non seulement 'i', comme dans « ciel » mais aussi 'y' (noyer, paye) et 'il(l)' (paille, rail, fille), sans compter que 'ill' précédé par une consonne peut également se prononcer comme dans « ville ».

Les lettres doubles (ou géménées)

La troisième série de problèmes est liée aux lettres doubles, encore appelées « lettres géménées », qui sont quasiment uniquement des consonnes (sauf dans quelques mots étrangers, cf. « football »). Le plus souvent, ce doublage n'est plus motivé, sauf quand ces consonnes géménées sont après 'e', qui se prononce alors /E/ (elle, belle, presse, verre, cette, mettre).

Il est à noter que le double 'nn', comme le double 'mm', indiquait autrefois que la voyelle qui précède était nasalisée (comme encore dans « ennui » et « emmener »). Or depuis plusieurs siècles, la plupart des voyelles ont perdu leur nasalité. En effet, on ne prononce plus « année » comme « an » et la leçon de « grammaire » n'est pas une leçon de « grand-mère ».

Autres problèmes : les lettres muettes de fin de mots et d'autres irrégularités

Une quatrième série de difficultés concerne les lettres muettes de fin de mot, qui sont souvent des marques morphologiques, comme 's' (amis, gris), 't' (petit, il finit), 'x' (jeux, tu peux, deux) et 'd' (grand). C'est la raison pour laquelle un travail spécifique doit être mené autour de la morphologie.

Une dernière série de difficultés est liée à d'autres irrégularités de l'orthographe du français. Par exemple, celle de mots comme « femme », « sept », « monsieur », ou le fait que 'x' ne se lit pas de la même façon dans « taxi », « examen », « dixième » et « deux ». De plus, cette consonne, qui est le plus souvent muette en fin de mots, peut parfois se prononcer (cf. « six », « dix »), tout comme 's', 't' et 'r' (cf. « bus », « net », « mer »).

Comment aborder ces différentes questions ?

Certaines de ces caractéristiques du français peuvent s'expliquer soit par des règles contextuelles (lecture de 's', 'c' et 'g' et écriture du 'n' des voyelles nasales), soit par la morphologie : flexions nominales (amie, amis, jeux), flexions verbales (tu chantes, ils chantent, chanter) et lettres qui sont des supports de dérivation et/ou de flexion ('d' de 'grand', 's' de 'gris', 'x' de deux). Un enseignement explicite peut donc faciliter leur apprentissage.

Par contre, l'apprentissage des autres spécificités de notre orthographe doit passer par la mémorisation d'instances spécifiques, qui peut être facilitée par des régularités statistiques non motivées phonologiquement. C'est le cas pour

les lettres fréquemment géminées ayant des CGPh consistantes (tt, mm, rr, nn) comparativement à celles qui ne sont géminées que rarement (bb, dd). Plusieurs études (cf. Pacton, 2008) ont en effet montré que les enfants sont très tôt sensibles à ces régularités statistiques qui sont purement orthographiques (ce qui n'est pas le cas du double 's').

L'*encadré 2* présente les questions qui devraient être abordées dans les étapes ultérieures de l'apprentissage de la lecture. Comme ce qui a été proposé pour les débuts de cet apprentissage, cette partie s'appuie sur Manulex-Morpho et tient compte de la fréquence et de la consistance des CGPh dans cette base de données, ainsi que de l'apport de la morphologie, principalement flexionnelle. Ces différentes questions devraient pouvoir être prises en compte par les enseignants (ou par les personnels en charge des enfants ayant des difficultés de lecture), en fonction des besoins des enfants.

POUR CONCLURE

La proposition de progression présentée dans cet article est un exemple qui tient compte de la fréquence et de la consistance des CGPh dans le lexique écrit adressé à l'enfant via les manuels scolaires. Les tableaux inclus dans ce texte, qui résument les statistiques de Manulex-Morpho, peuvent aider ceux qui souhaitent adapter cette progression (non seulement les personnels en charge de l'apprentissage de la lecture, mais aussi les formateurs de ces personnels ainsi que les concepteurs de méthodes), le mieux étant bien sûr de consulter la base en elle-même⁷.

Cette progression est focalisée sur l'identification des mots écrits. Cela s'explique par le fait que, pour comprendre ce qu'on lit dans une écriture alphabétique, il faut avoir maîtrisé cette procédure de lecture. Seule son automatisation permet au débutant de comprendre progressivement ce qu'il lit comme il comprend ce qu'il entend (cf. pour des synthèses en français, Dehaene *et al.*, 2011 ; Sprenger-Charolles & Colé, 2013).

En conclusion, il faut préciser que, au début de l'apprentissage de la lecture-écriture, le travail sur la compréhension d'énoncés, comme celui sur leur production, devra s'effectuer principalement à l'oral. En effet, il n'est pas possible de travailler sur la compréhension d'énoncés tant que les enfants n'ont pas bien automatisé les procédures d'identification des mots écrits (Gentaz *et al.*, 2015). Il n'est évidemment pas d'avantage possible de travailler sur la production d'énoncés, étant donné le retard de l'écriture sur la lecture (voir Fayol, 2017).

Le fait de ne travailler la compréhension qu'à l'oral au début de l'apprentissage de la lecture n'est pas un problème, le processus de compréhension étant très largement amodal, c'est-à-dire similaire à l'écrit et à l'oral. Au début de cet apprentissage, comme cela est le cas en maternelle, ce travail s'effectuera à l'aide de textes lus par le maître ou écrits par lui sous la dictée des enfants, par exemple. Il faudra prendre en compte différents types de textes afin de faire comprendre aux enfants qu'on peut lire avec des buts divers : pour le plaisir (lire des contes, des poèmes...),

pour s'informer (lire des chapitres d'un livre d'histoire ou de sciences...), pour réaliser quelque chose (en suivant la notice d'une recette de cuisine ou d'un jeu). Dans ce travail, il faudra mettre l'accent sur les marques qui relient les mots et les phrases d'un texte et, ainsi, facilitent sa compréhension: reprises de noms des personnages (Luc -> il -> ce garçon) ou des lieux (à Paris -> dans cette ville); mots indiquant la cause ou la conséquence d'une action (parce que, puisque, en conséquence), ou l'opposition (mais, pourtant) ; mots indiquant des relations spatiales (à Paris, ici, là, devant, derrière) ou temporelles (hier, aujourd'hui, demain, avant, après), ces dernières pouvant aussi être portées par des flexions verbales dans certaines langues (en français) mais pas dans d'autres (en anglais)⁸. Il sera également important de présenter des textes issus d'univers culturels différents pour sensibiliser très tôt les enfants aux variations culturelles.

REMERCIEMENTS

Je remercie l'équipe française de Grapho-Learn, en particulier Johannes Ziegler, responsable du projet, ainsi que Édouard Alavoine, Jean-Patrice Albrand et Julie Lassault qui ont adapté la progression proposée au logiciel GraphoGame. Mes remerciements s'adressent également à Alain Desrochers pour une lecture attentive d'une première version de ce manuscrit accompagnée de remarques très pertinentes.

⁷ Voir la note 3.

⁸ C'est le cas, par exemple, dans 112 des 222 langues examinées par Dahl et Velupillai (2013) dans l'inventaire du *World Atlas of Language Structures* [WALS]. En outre, dans certaines de ces langues, le futur n'est même pas codé par un auxiliaire du verbe, comme en anglais (*shall, will, going to*) : il est seulement indiqué par une marque temporelle non verbale (aujourd'hui, demain), comme en finnois..

Encadré 2. Questions devant être abordées dans les secondes étapes de l'apprentissage.

<p>• 1 : Les semi-voyelles (le 'oi' de « loi » et le 'u' de « lui »), la consonne muette 'h', plus 'ph'</p>
<p>• 2 : Règles contextuelles : 'n' des voyelles nasales ; consonnes /s/-/z/ et 'c'-'g'</p> <p>2a. Le 'n' des voyelles nasales s'écrit 'm' devant 'p' et 'b' (chambre, ampoule ; sombre, complet ; timbre, impossible...)</p> <p>2b. Le /s/ et le /z/</p> <ul style="list-style-type: none"> - La consonne 's' se lit /s/ en début de mot et quand elle est précédée ou suivie par une consonne (liste, bourse, danse) Entre deux voyelles ou semi-voyelles, elle se lit /z/ (arroser, arrosant, arrosoir, chinoise), sauf lorsqu'elle est précédée par une nasale qui se termine graphiquement par une consonne (danse) Dans ce même contexte, pour se lire /s/, la consonne 's' doit s'écrire 'ss' (brosse, brousse, boisson, poisson) - La consonne 'z' se lit /z/ en début et en milieu de mot (zone, onze) <p>2c. Le 'c' et le 'g' (plus le 'gn')</p> <ul style="list-style-type: none"> - Les 'c' et 'g' durs : devant les voyelles commençant par 'a'-'o'-'u' (a, au, ai, an[m]; o, ou, on[m]; u, un) ou devant une consonne car-cause-caisse-camps, corps-cou-conte, culbute-chacun, classe-crasse ; gare-gant, gout, légume-figure, grand-glace..., - Les 'c' et 'g' doux : devant les voyelles commençant par 'e'-'è'-'ê' et 'i'-'y' (e-é-è-ê, ei, en[m]; i, in[m]; y, yn[m]) celle-cédille-cent-ceinture, cil-cinq, cygne-cymbale ; gel-gèle-genre, gilet, gymnastique - Autres orthographes de 'c' et 'g' : 'ce', 'cu'⁹ et 'ge', 'gu' plus 'ç' : liées aux règles contextuelles précédentes, par exemple, ajout d'un 'e' après 'g' au participe présent des verbes en 'ger' (en mangeant) ou en remplaçant le 'c' par un 'ç' (plaçant)
<p>• 3 : Allographes des voyelles orales</p> <ul style="list-style-type: none"> - La voyelle 'o' s'écrit non seulement 'o', mais aussi 'au', et 'eau' (auto, chapeau...) - Autres allographes de /E/ : 'e' devant une double consonne (belle, cette, presse...) ainsi que dans une syllabe fermée phonologiquement (bel, sel, sec...) ou orthographiquement (les, effet, user, nez) plus 'ei' (reine) - La voyelle 'i' peut aussi s'écrire 'y' (lys) - La voyelle 'eu' peut aussi s'écrire 'oeu' (nœud)
<p>4. Allographes des voyelles nasales (la plupart sont des marques morphologiques)</p> <ul style="list-style-type: none"> - Allographes de /â/ : 'an-t/d' (chan-t[er]), marchan-d[ise]), 'en-t' (den-t[iste]), plus 'an'/'en', (tante, tente) - Allographes de /i/ : 'ain' (bain/baigner), 'aim' (faim/famine), 'ein' (plein/pleine), plus 'in'/'en' (fin, chien, examen, agenda)
<p>5. La semi-voyelle /j/ (appelée 'yod')</p> <ul style="list-style-type: none"> - 'i' : suivie par une autre voyelle orale (ciel, piano) ou par une voyelle nasale (bien, rien, avion...) - 'il(l)' : rail (r + a + yod), fauteuil (t + eu + yod), paille (p + a + yod), feuille (f + eu + yod), nouille (n + ou + yod) - 'y' = 'i' + 'i' : le premier 'i' change la valeur de la voyelle qui précède, le second est un yod : paye (ai + yod), noyer (oi + yod)...
<p>6. Quelques pièges</p> <ul style="list-style-type: none"> - Mots fréquents qui ne se lisent pas comme ils s'écrivent (femme, sept...) De même, la 'fille' ne se lit pas comme la 'ville' et une 'quille' ne se lit pas comme 'tranquille' - Comment se lit la consonne 'x' en milieu de mot (taxi, examen, dixième) et en fin de mots (deux, prix, index, six) - Le 't' de 'tion' se lit /s/ (situation) sauf quelques exceptions dans des noms (question) et pour tous les verbes (nous chantions) - Mots se terminant par les consonnes 's', '(e)t', '(e)r' qui se prononcent (bus, net, mer, cher, fer, fier)... - La consonne 'ch' se lit /k/ dans quelques mots (technique)
<p>Autour de la morphologie (devrait être inséré dans les blocs précédents)</p> <p>Flexions verbales :</p> <ul style="list-style-type: none"> - Présent des verbes du premier groupe (chanter, jouer) et du second groupe (finir) - Présent des verbes irréguliers les plus fréquents (être, avoir, faire) - Les participes présents (en mangeant) et les adverbess en 'ment' (rarement) <p>Flexions nominales : genre (le 'e' de 'finale' et celui de 'amie') et nombre (le 's', puis le 'x').</p> <p>Lettres supports de dérivation / flexion : 's' (gris -> grise; trois -> troisième), 'x' (deux -> deuxième), 'd' (grand -> grande, grand-deur)</p>
<p>Exercices spécifiques sur les confusions visuelles et phonologiques (idem module 1 avec /g/ et /z/)</p> <ul style="list-style-type: none"> - Confusions phonologiques. Occlusives: /p/-/t/, /t/-/k/, /d/-/g/, /t/-/d/, /k/-/g/. Fricatives: /f/-/s/, /s/-/ʃ/, /v/-/z/, /z/-/ʒ/, /ʃ/-/v/, /s/-/z/, /ʃ/-/ʒ/ - Confusions phonologiques et visuelles : /b/-/d/, /p/-/b/ - Confusions visuelles : n-u, p-q, d-q, i-j, v-y, h-n

RÉFÉRENCES

- CATACH, N. (1973). Que faut-il entendre par système graphique du français. *Langue française*, 20, 30-44.
- CATACH, N. (2005). *L'Orthographe française*. Paris : Colin.
- CATACH, N. [Dir.] (1995). *Dictionnaire historique de l'orthographe française*. Paris : Larousse.
- DEHAENE, S. [DIR.], DEHAENE-LAMBERTZ, G., GENTAZ, É., HURON, C. & SPRENGER-CHAROLLES, L. (2011). *Apprendre à lire : des sciences cognitives à la salle de classe*. Paris : Odile Jacob.

- DELATTRE, P. (1965). *Comparing the phonetic features of English, French, German and Spanish*. Heidelberg: Jumiuss Gross Verlag.
- FAYOL, M. (2013). *L'Acquisition de l'écrit*. Paris : PUF, Que sais-je ?
- GENTAZ, É., SPRENGER-CHAROLLES, L. & THEUTEL, A. (2015). Differences in the predictors of reading comprehension in first graders from low socio-economic status families with either good or poor decoding skills. *PLOS ONE*. 10 (3), e0119581.
- LETE, B., SPRENGER-CHAROLLES, L. & COLÉ, P. (2004). Manulex: A lexical database from French readers. *Behavioral Research Methods, Instruments and Computers*, 36 (1), 156-166.
- DAHL, O. & VELUPILLAI, V. (2013). The Future Tense. In: M.S. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/67>, Accessed on 2017-09-28.)
- LEON, P. & LEON, M. (1964). *Introduction à la phonétique corrective*. Paris : Hachette/Larousse.

⁹ En fait, le 'ue' précédé par 'c' (qui se trouve quasi-uniquement dans les mots de la famille de « cueillir »), tout comme certains 'g+ue' (orgueil) est l'ancienne orthographe de 'eu' (Catach, page 1113 du Dictionnaire historique).

- PACTON, S. (2008). L'Apprentissage de l'orthographe du français. In A. Desrochers, F. Martineau, Y.C. Morin (éds), *Normes et pratiques orthographiques*. Ottawa : Éditions David.
- PEEREMAN, R., SPRENGER-CHAROLLES, L. & MESSAOUD-GALUSI, S. (2013). The contribution of morphology to the consistency of spelling-to-sound relations: A quantitative analysis based on French elementary school readers. *L'Année psychologique*, 113 (1), 3-33. http://lpnc.univ-grenoble-alpes.fr/resources/ronald_peereman/Manulex_morpho/indexfr.html
- SHARE, D.L. (2008). On the Anglocentricities of current reading research and practice: the perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, 134, 584-615.
- SPRENGER-CHAROLLES, L., SIEGEL, L., BECHENNEC, D. & SERNICLAES, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading and in spelling: A four years longitudinal study. *Journal of Experimental Child Psychology*, 84, 194-217.
- SPRENGER-CHAROLLES, L. & COLÉ, P. (2013). *Lecture et dyslexie : approche cognitive*. Paris : Dunod.
- ZIEGLER, J. & GOSWAMI, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3-29.