

Manulex-MorphO, une base de données sur l'orthographe du français intégrant les morpho-phonogrammes

Ronald Peereман, Liliane Sprenger-Charolles

DANS **LANGUE FRANÇAISE** 2018/3 (N° 199), PAGES 99 À 109

ÉDITIONS **ARMAND COLIN**

ISSN 0023-8368

ISBN 9782200931841

DOI 10.3917/lf.199.0099

Article disponible en ligne à l'adresse

<https://www.cairn.info/revue-langue-francaise-2018-3-page-99.htm>



CAIRN.INFO
MATIÈRES À RÉFLEXION

Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...

Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour Armand Colin.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Ronald Peereman

Laboratoire de Psychologie et NeuroCognition (UMR 5105) & Université Grenoble Alpes

Liliane Sprenger-Charolles

Aix-Marseille Université & Laboratoire de Psychologie Cognitive (UMR 7290) & Institut Langage, Communication et Cerveau (Marseille)

Manulex-MorphO, une base de données sur l'orthographe du français intégrant les morpho-phonogrammes

La consistance des correspondances graphème-phonème (CGPh) et phonème-graphème (CPhG) est un facteur clé pour l'apprentissage de la lecture-écriture dans une écriture alphabétique (Treiman & Kessler 2014). L'étude de l'acquisition du code alphabétique et le développement de programmes d'apprentissage de la lecture-écriture nécessitent donc de disposer de descriptions quantitatives de ces correspondances dans le lexique écrit adressé à l'enfant. Jusqu'à présent, ces statistiques ont été réalisées indépendamment de l'analyse morphologique. C'est le cas des études de J. C. Ziegler, A. M. Jacobs et G. O. Stone (1996) et de R. Peereman et A. Content (1999), qui n'ont pris en compte que des mots monosyllabiques. C'est également le cas de celles de N. Catach (1973, 2005) et J. Véronis (1986) – les deux développées à partir des mots de l'échelle Dubois-Buyse (Ters, Mayer & Reichenbach 1977) –, ainsi que de celle de R. Peereman, B. Lété et L. Sprenger-Charolles (2007), développée à partir de la base Manulex (Lété, Sprenger-Charolles & Colé 2004).

L'absence de prise en compte de la morphologie pose problème en raison du fait qu'elle différencie fortement l'écrit de l'oral en français. Ainsi l'existence de marques morphologiques écrites, mais silencieuses à l'oral (<s> dans *amis*, *tu chantes*, *gris*), et l'homophonie sans homographie d'autres marques (<ais> et <ait> dans *chantais*, *chantait* ; <ait> et <aid> dans *lait* et *laid*) ont un impact négatif sur l'apprentissage de la lecture et, surtout, sur celui de l'écriture (Fayol 2013).

Afin d'examiner la contribution de l'information morphologique à la consistance des CGPh et CPhG, nous avons développé la base de données Manulex-MorphO (Peereman, Sprenger-Charolles & Messaoud-Galusi 2013). L'article présente d'abord cet outil, puis ce qu'apporte l'indication de la morphologie à la consistance des relations grapho-phonémiques et, enfin, des exemples d'exploitations pour l'enseignement.

1. PRÉSENTATION DE MANULEX-MORPHO

1.1. Corpus et codage morphologique

Le corpus lexical analysé est issu de Manulex (Lété, Sprenger-Charolles & Colé 2004). Manulex comprend 48 900 mots non lemmatisés issus de 54 manuels du primaire et Manulex-MorphO les 9 949 mots les plus fréquents de Manulex, soit 20 % des entrées lexicales de Manulex, mais 98 % des 1.9 million d'occurrences contenues dans les manuels examinés¹. Les indices fréquentiels correspondent à la fréquence par million de mots (indice F), qui peut être pondérée par le nombre de manuels dans lesquels chaque mot est rencontré (indice U)².

Des caractères standards ont été utilisés pour les codes phonétiques. Deux signes codent le <e> muet : l'un pour celui de *table*, qui permet de générer une nouvelle syllabe (Clements 1990) et est obligatoire (/°/), l'autre pour celui de *gare*, qui est optionnel (/°/). Le signe /#/ code les lettres n'ayant pas d'équivalent à l'oral (<e> et <s> de *amies*). Les autres codes spécifiques sont /1/ pour <un>, /5/ pour le <in> de *cinq*, /2/ et /9/ pour respectivement le <eu> de *deux* et celui de *neuf*, /@/ pour *an* et /§/ pour *on*. Seules les marques correspondant à un seul phonème ont été considérées comme étant un morpho-phonogramme, plus précisément : les flexions nominales et verbales, les finales <ent> des adverbes en *-ment* (mais pas *-ment*) et les consonnes support de flexion et/ou dérivation (le <d> de *grand* vs *grande*, *grandeur*). Les associations concernées par le marquage de la morphologie sont au nombre de 7 456, réparties sur 6 377 mots, soit 64 % du corpus lexical.

1.2. Calculs de la fréquence et de la consistance des CGPh-CPhG et des mots

Les CGPh et CPhG sont analysées en fonction, d'une part, de leurs fréquences d'occurrence dans les mots et, d'autre part, de leurs consistances (probabilité d'un graphème à partir d'un phonème et *vice-versa*). La fréquence et la consistance de chaque association ont été évaluées en prenant, ou non, en compte la

1. Soit 81 % des mots de l'échelle de Dubois-Buyse.

2. Les fréquences d'occurrence des mots sont extraites de Manulex, tout comme leurs catégories grammaticales.

fréquence des mots portant l'association (par type, la fréquence lexicale ; par token, la fréquence textuelle).

Par exemple, la fréquence lexicale de l'association <ch>-/k/ est 13 (13 mots contenant cette association) et sa fréquence textuelle de 136, ce qui correspond à la somme des fréquences d'occurrence des 13 mots. Le graphème <ch> est prononcé différemment dans 565 autres mots avec une fréquence cumulée de 18 289. La consistance lexicale (multipliée par 100) de l'association <ch>-/k/ est donc faible (2.25 ; soit $(13 / [565+13]) \times 100$), tout comme sa consistance textuelle (0.74 ; soit $(136 / [18\ 289+136]) \times 100$).

Autre exemple, la consistance des associations G-Ph pour le mot *autre* est supérieure à 97 %, quelle que soit la fréquence choisie (lexicale ou textuelle), sans changements majeurs en fonction de la position. En revanche, celle des associations Ph-G n'est que de 79 % pour la fréquence lexicale, ce qui est dû au fait que /O/ peut s'écrire de différentes façons, cette consistance s'élevant toutefois à 90 % pour la fréquence textuelle, ce qui provient de ce que les mots commençant par <au> sont plus fréquents que ceux commençant par <o> ou <eau>.

1.3. Disponibilité des données issues de Manulex-MorphO

Les fichiers sont disponibles sur le site Manulex-MorphO³. Quatre présentent les consistances et fréquences des associations contenues dans chacun des mots du corpus. Un cinquième fichier donne la fréquence et la consistance des CGPh et des CPhG, avec les mêmes variantes : indices U et F pour la fréquence lexicale et textuelle. Ces dernières informations permettent de déterminer les consistances orthographiques pour des mots absents du corpus lexical ou lors de l'utilisation de mots inventés (des pseudomots). Dans tous les cas, les informations sont procurées avec, ou sans, distinctions entre les voyelles correspondant aux archiphonèmes /A/, /œ/, /E/ et /O/, à la différence des analyses de N. Catach (1973, 2005), qui n'ont porté que sur ces archiphonèmes⁴.

2. CONSISTANCE DES CGPH ET DES CPHG : L'APPORT DE LA MORPHOLOGIE

L'apport de l'information morphologique n'était pas pris en compte dans notre travail initial (Peereman, Lété & Sprenger-Charolles 2007), ainsi que dans la récente étude statistique réalisée par M. Gingras et M. Sénéchal (2017). Dans cette

3. En attente de la mise en place d'un site permettant de consulter la base en ligne, les fichiers sont accessibles au format Excel [http://pnc.univ-grenoble-alpes.fr/resources/ronald_peereman/Manulex_morpho/].

4. Ces archiphonèmes correspondent aux voyelles /a/-/A/, /ø/-/œ/, /e/-/E/ et /o/-/O/, qui sont des variantes contextuelles, dépendant de la structure de la syllabe (fermée ou ouverte), dans laquelle elles se rencontrent (Delattre 1965).

partie, nous examinons l'apport de cette information à la consistance globale des CGPh et des CPhG (§ 2.1), puis dans un cas particulier : celui des lettres muettes de fin de mot (§ 2.2). La dernière section (§ 2.3) concerne l'apport de la morphologie à l'analyse des productions écrites des enfants.

2.1. Impact de la morphologie sur la consistance des CPhG et CGPh

Pour évaluer la contribution de l'information morphologique à la consistance des CGPh et des CPhG, les statistiques ont été réalisées avec et sans indiciage de la morphologie et en tenant compte de la position dans le mot (début, milieu, fin). Il ressort des analyses (Tab. 1) une forte asymétrie entre CGPh et CPhG, les premières étant les plus consistantes. Autre fait marquant : cette asymétrie, très marquée en fin de mot quand la morphologie n'est pas prise en compte, diminue fortement quand elle l'est, et ce sont surtout les CPhG qui en bénéficient : le gain est de pratiquement 10 % (de 66 à 75 %) contre 2 % pour les CGPh (de 78 à 80 %).

Cette asymétrie est due, en plus de la morphologie, au fait que, si certains graphèmes se lisent toujours de la même façon (<f> et <ff> et <ph> se lisent /f/), pour écrire le phonème correspondant, il faut choisir entre ces différents allographes. Pour ces deux raisons, apprendre à lire est plus facile qu'apprendre à écrire en français, en témoignent les résultats relevés aussi bien dans des épreuves prenant en compte uniquement des formes lemmatisées (Sprenger-Charolles *et al.* 2003) que dans celles avec des formes fléchies (Fayol 2013).

Tableau 1 : Consistance (%) des CGPh et CPhG en fonction de la position dans le mot et de la prise en compte des morpho-phonogrammes (fréquence textuelle)

Position dans le mot	CGPH (lecture)		CPhG (écriture)	
	Sans morphologie	Avec morphologie	Sans morphologie	Avec morphologie
Début	96	96	87	87
Interne	74	75	72	72
Fin	78	80	66	75

2.2. Le cas des lettres muettes de fin de mot

Le Tableau 2 présente les graphèmes les plus fréquents (U>10 000) qui ne se prononcent pas en fin de mots (le <e> de *amies* et le <x> de *auxquels* ne sont donc pas pris en compte). Parmi les différents <e> de fin de mot, le plus fréquent est celui de *gare* et *table* (76.9 % des CGPh), qui n'est pas un morphème. En revanche, pour les consonnes, les graphèmes muets les plus fréquents (<s>, <x>, <t>) sont, dans la majorité des cas, des morpho-phonogrammes.

Tableau 2 : Graphèmes muets les plus fréquents (U>10 000)

Graphèmes (avec exemples) et % de ceux qui sont totalement muets (#) suivi par le nombre total de graphèmes par série		Pourcentages par catégorie (total pour chaque graphème [<e>, <s>, <x>, <t>]=100 %)			
		Flexion nominale	Flexion verbale	Support flexion ou dérivation	Non motivé
<e> Il joue, Amie Il chante / Table	#=12 % (226.609)	3 % (amie) 13 % (fine)	<1 % (je joue) 7 % (je parle)		<1 % (lycée) ^a 77 % (gare)
<s> Amis, Gris Ours, Bus	#=96 % (136.930)	58 % (amis)	5 % (tu joues)	4 % (gris) <1 % (ours)	30 % (radis) 3 % (bus)
<x> Jeux, Tu peux Six, Index	#=96 % (11.371)	47 % (jeux)	8 % (tu peux)	37 % (deux) 4 % (six)	5 % (mieux) <1 % (silex)
<t> Toit, Il finit Net	#=93 % (31.684)		26 % (il finit)	49 % (dent) 1 % (net)	23 % (appétit) <1 % (ouest)

a. Le 'e' de « allée » a été considéré comme une marque de flexion nominale.

2.3. Apport de l'indilage morphologique dans l'analyse des productions orthographiques

L'apport des estimations statistiques réalisées dans Manulex-MorphO a été évalué en le comparant avec celui d'estimations n'incluant pas d'indilages morphologiques (Manulex-Infra ; Peereman, Lété & Sprenger-Charolles 2007) pour un ensemble de 3 136 mots écrits par des enfants du CP au CM2 (Pothier & Pothier 2003). Des analyses de régression « par pas » ont été réalisées sur les pourcentages de réponses correctes après prise en compte de la variabilité de la performance en fonction de facteurs tels que la fréquence des mots et leur longueur.

Les résultats indiquent que les estimations de consistance indiquant les morpho-phonogrammes ajoutent de manière significative à l'explication de la variance par rapport aux estimations réalisées indépendamment des indices morphologiques. Pour les 612 mots marqués par des indices morphologiques flexionnels/dérivationnels, l'accroissement du pouvoir explicatif est apparent dès le CE1.

L'apport de ces nouvelles estimations résulte en partie du fait que l'identification d'un phonème comme correspondant à un morpho-phonogramme (ou à une partie) réduit l'incertitude liée au correspondant orthographique (p. ex. <ent> pour les verbes ou pour les adverbes en *-ment*).

3. UTILISATION DE MANULEX-MORPHO POUR ÉLABORER UNE PROGRESSION POUR L'APPRENTISSAGE DE LA LECTURE

Manulex-MorphO peut être utilisé pour élaborer une progression pédagogique centrée sur le décodage et tenant compte de la fréquence et de la consistance des CGPh, ainsi que de celle des mots ⁵. Les données présentées dans les Tableaux 3 et 4 peuvent guider la mise en place de cette progression : elles fournissent en effet un résumé des statistiques issues de Manulex-MorphO. Sont surlignées en gris les CGPh fréquentes (>5000, gras et italique) et très fréquentes (>15 000, gras), qui sont également consistantes (>95 %). Les numéros dans la colonne centrale indiquent l'ordre des CGPh les plus consistantes et fréquentes. Par exemple, les cinq premières sont, pour les voyelles : <a>=/A/, suivi par <ou>=/u/, puis par <o>=/O/, <é>=/E/, <on>=/ô/ ; et, pour les consonnes : <r>=/R/, suivi par <l>=/l/, puis par <p>=/p/, <m>=/m/, <v>=/v/.

3.1. Premières étapes de l'apprentissage de la lecture-écriture

Les données de la littérature (cf. les trois premiers articles de ce numéro) indiquent que le décodage grapho-phonémique a un rôle crucial dans les premières étapes de l'apprentissage de la lecture. À ce stade, il faudrait donc apprendre d'abord aux enfants les correspondances les plus fréquentes et, surtout, les plus consistantes : par exemple, pour les voyelles : <a>, <é>, <o>, <eu> et <ou>, plus <i> et <u> (sauf lorsque ces dernières sont suivies par une autre voyelle comme dans *ciel* et *lui*), ainsi que le <e> en fin d'un mot précédé par une consonne (*le, je, te*). Les consonnes à introduire en premier sont les fricatives (<f>, <j>, <ch>, <v>), parce qu'elles se lisent presque toujours de la même façon et que leur reconnaissance en isolat est plus facile que celle des occlusives (qui ne permettent pas de produire un son continu). Les autres consonnes fréquentes ayant une prononciation régulière pouvant aussi être présentées au début de l'apprentissage sont <l> et <r> ainsi que, par la suite, <p>, , <m>, <n>, <d> et <qu> (qui se retrouve dans de nombreux mots fonctions) plus <s> (en début de mot).

Il faudrait toutefois introduire aussi très tôt certains mots fréquents contenant des CGPh peu régulières dans la mesure où, sans eux, il n'est pas possible de construire des énoncés. C'est le cas, par exemple, pour de nombreux mots fonction (*et/cet, pendant*), pour quelques noms (*dix, sept, femme, monsieur*) et pour certains verbes irréguliers (*il est grand, il a eu*).

D'autres principes devraient être respectés pour les débuts de cette progression. Ainsi, il faudrait utiliser en premier des mots courts composés de syllabes simples, avec une consonne (C) à l'initiale suivie par une voyelle (*la, le, jeu, feu, fou*), les items CVC (*bal, sol, jour*) et CCV (*tri, blé, bleu*) étant introduits dans un

5. Ces principes reprennent partiellement ceux présentés dans Dehaene (éd.) (2011), avant la mise en place de la base de données Manulex-MorphO.

second temps 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*).

3.2. Dans une seconde étape : morphologie et règles contextuelles

Dans un second temps, il faudrait aborder les difficultés de l'orthographe du français, certaines s'expliquant par la morphologie, d'autres par des règles contextuelles. Ces dernières, qui n'ont pas été intégrées dans Manulex-MorphO, concernent, entre autres, la lecture de <s>, de <c/g> et de <e>.

Ainsi, 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/ (*arrosé, arrosoir, chinoise*), sauf lorsque la voyelle qui précède est une nasale (et se termine par une consonne graphique : *danse*)⁶. Les graphèmes <c> et <g> se prononcent /k/ et /g/, respectivement, devant une consonne ou devant les voyelles commençant par <a>, <o>, <u> (<a>, <au>, <ai>, <an[m]> ; <o>, <ou>, <on[m]> ; <u>, <un>). Ils se prononcent /s/ et /ʒ/, respectivement, devant les autres voyelles (<e-é-è-ê> et <i-y>). Enfin, le graphème <e> se prononce /E/ devant une double consonne (*belle, cette, presse*), ainsi que dans une syllabe fermée phonologiquement (*sel, sec*) ou orthographiquement (*les, effet, user, nez*).

Pour les lettres muettes de fin de mots, il faudrait mettre rapidement l'accent sur le fait que, lorsqu'elles sont suivies par un mot commençant par une voyelle, la liaison se réalise entre l'article et le nom, ainsi qu'entre le pronom et le verbe (Encrevé 1988). Cette capacité est, en fait, maîtrisée assez tôt par les enfants français qui, dès 3 ans, comprennent la différence entre *il arrive* et *ils arrivent* (Legendre *et al.* 2010).

En revanche, l'apprentissage d'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 consistantes⁷ (<tt>, <mm>, <rr>, <nn>), comparativement à celles qui ne sont géminées que rarement (<bb>, <dd>). Plusieurs études (v. Pacton 2008) ont en effet montré que les enfants sont très tôt sensibles à ces régularités statistiques qui, à la différence du double <s>, sont purement orthographiques.

6. Sauf dans quelques mots pluri-morphémiques, comme *parasol* ou *tournesol*.

7. Ce n'est pas le cas du double <ll>.

Tableau 3 : Fréquence textuelle (FrqU) et consistance (Consist) des relations graphème-phonème : (1) voyelles et semi-voyelles [FrqU : gras >15 000, ital >5 000 – Consist : gras >95 % – Cellules 1 à 4 avec fond gris : graphèmes les plus fréquents et consistants (1^{er} au 11^e : cf. colonne centrale)]

Voyelles orales		Exemples. Règle contextuelle [RC]: Morphologie: flexions noms et verbes [FN], [FV], support flexion / dériv [SFD]; Orthographe rectifiée [OR] <i>Gris italique: CGPh exceptionnelles (fr<500)</i>			Voyelles nasales et semi-voyelles		Exemples. Idem colonne de gauche		
<>	FrqU	Consist		1-6	FrqU	Consist			
<>	=/A/	143.285	>95	ma / mal (+ aout [OR]); cake, lady, football)	<an>	=/ã/	29.147	>95	tante (+ gentleman: /an/)
<>	=/A/	17.642		5 mots: ô, là, ça, voilà, déjà	<am>	=/â/	1.151	100	ampoule (+ tamtam: /am/)
<>	=/A/	1.485		40 mots: âge, âne, château, gâteau, râtelier...	<ant>	=/ã/	3.395	100	en faisant (FV)
<>	=/æ/	266.114	74	file (79%); fine [FN: 12%]; il vole (FV: 6%)		=/ê/	31.908	82	lente
<ent>	=/æ/	9.710	3	amie (FN: 3%); il joue (FV: <1%); lycée		=/ê/	4.098	100	membre
<ent>	=/æ/	3.754	68	ils chantent [FV]	<ent>	=/ê/	3.504	100	rarement (adverbes en 'ment')
<eu>	=/E/	23.315	98	deux / neuf (+ œil)	<en>	=/ɛ̃/	6.890	18	[RC] chien + examen (+ Ben: /En/)
<eu>	=/E/	561	2	il a eu [FV]	<in>	=/ɛ̃/	8.310	100	fin
<eu>	=/E/	84.535	24	9 mots: scieur (+ <ue> de cueille)	<im>	=/ɛ̃/	876	>99	impossible (+ Jim: /im/)
<>	=/E/	84.535	24	[RC]: elle, cette...; sel, reste, verte...; les, mes...	<in>	=/ɛ̃/	624	100	frein
<>	=/E/	38.875	100	été	<ain>	=/ɛ̃/	3.500	100	bain
<>	=/E/	11.773	100	lève	<aim>	=/ɛ̃/	165	3	mots: faim, essaim, daim
<>	=/E/	5.932	100	arrêt / arrête (+ Noël; canoë [OR])	<ym>	=/ɛ̃/	48	3	mots: symbole, thym, olympique
<eb>	=/E/	1.121	100	neige	<on>	=/ô/	33.125	97	point
<ai>	=/E/	22.816	98	aigle (+ nous faisons: <ai> = /ə/)	=/O/ə/	709		3	mots: bonheur, bonhomme; monsieur
<ait>	=/E/	9.909	100	il chantait (fréquence de <ais> =2440)	=/ô/	5.333	>99	ombre (+ Tom: /om/)	
<et>	=/E/	20.290	98	jouet (95%); fouet (SFD: 5%); prêt [FN]	<um>	=/u/	19.074	100	un
<er>	=/E/	16.422	93	3 mots: net, cet, basket [SFD]	<um>	=/u/	46	39	parfum
<ez>	=/E/	2.918	100	nager (FV: 78%); léger (SFD: 11%); cahier (11%)	=/om/	71	61	album, maximum, aquarium...	
<>	=/E/	1.210	7	mer	<u>	=/u/	24.746	20	ciel (+ février: /ij/ = <18%>)
<>	=/E/	97.911	79	vous lavez (FV: 60%); chez (40%)	<y>	=/j/...	2.770	44	yeux; crayon; voyage (+ pays: /E/ = 8%)
<>	=/E/	3.036	48	lys	<il>	=/i/	1.807	98	chandail (+ gentilhomme: /ij/)
<>	=/E/	258	40	[OR]: sans accent (boite, chaîne, entraîner...)	<il>	=/i/	2.841	48	oreille; paille (après voyelle)
<>	=/E/	40	5	5 mots: héroïne, mais, ouïe, inouï, Loïc	<il>	=/ij/	1.959	32	fille (après consonne + million: /ij/ =2%)
<>	=/O/	49.707	>95	mot / mort (+ clown)	<oi>	=/wa/	20.725	100	moi
<>	=/O/	1.658	26	26 mots: aussitôt, bientôt, plutôt, plutôt, côté...	<oy>	=/we/	1.339	100	envoyer (+ water: /w/)
<au>	=/O/	17.003	100	auteur	<oi>	=/w/	2.078	100	loin
<eau>	=/O/	6.102	100	5 mots: alcool, zoo, coordination, foot, igloo	<u>	=/u/	11.318	18	lui (+ /y/)
<u>	=/U/	51.248	82	lune (+ club, puzzle); [OR] <u> sauf ambiguïté	Le signe ↔ signale que le graphème a 2 prononciations (<u> = /y/ - /u/)				
<ou>	=/u/	50.024	97	tour (+ /w/ de 'bouée'); <ou> (un seul mot)					

Tableau 4 : Fréquence textuelle (FrqU) et consistance (Consist) des relations graphème-phonème : (2) consonnes [FrqU : gras>15 000, ital.>5 000 – Consist : gras>95 % – Cellules 1 à 4 avec fond gris : graphèmes les plus fréquents et consistants (1^{er} au 11^e : cf. colonne centrale)]

Graphèmes pour /p/-/t/-/k/; /b/-/d/-/g/; /m/-/n/-/ŋ/ (+ <h>) FrqU Consist	Exemples (règle contextuelle; morpho- logie; Orthographe rectifiée; CGPh exception): cf. Tableau 3	Graphèmes pour /f/-/s/-/ʃ/; /v/-/ʒ/-/ʒ/; FrqU Consist	Exemples (règle contextuelle; morpho- logie; Orthographe rectifiée; CGPh exception): cf. Tableau 3
<p> =/p/ 81.615 95 =/h/ 4.664 5 <pp> =/p/ 2.603 100	pomme coup (SFD: 43%); loup (58%) nappe	<f> =/f/ 27.396 >99 <ff> =/f/ 1.946 100 <pf> =/f/ 2.930 100 <v> =/z/ 713 95 <ch> =/ʃ/ 18.614 >99	four (+ nerf; /#) offrir pharmacie zone (+ riz; /#) chat (+ sandwich; /ʃ/)
<h> =/h/ 136 <1 <t> =/t/ 99.257 71 =/h/ 37.784 27 =s 2.783 2 <t> =/t/ 6.843 100	chorale tour cent (SFD: 52%); lit (FV: 22%); têt: 26% addition attendre (+ <th>; /the)	<s> =/s/ 67.552 32 =/h/ 126.443 61 =/z/ 14.408 7 <ss> =/s/ 11.826 100 <ʃ/ʃc> =/s/ 22.795 34 =/ks/ 2.696 100 =/z/ 239 27	amis (FN: 64%); gris (SFD: 4); jus (33%) basse garçon (+ science; /s/) [RC]: accent
<c> =/k/ 42.805 64 =/h/ 865 >1 <c> =/k/ 647 73 <qu> =/k/ 29.668 >99 <q> =/k/ 371 =/k/ 396 <v> =/z/ 713 95 <h> =/ʃ/ 18.614 >99	[RC] col caoutchouc (+ seconde; /g/) [RC] accord qui (+ /kw/; aquarium) 3 mots: cinq, coq, pique 25 mots: kilo...	<x> =/h/ 10.955 71 =/ks/ 2.657 17 =/gz/ 1.338 9 =/z/ 588	jeux (FV: 49%); deux (SFD: 38%)... Taxi 3 mots: deuxième, dixième, sixième
<h> =/h/ 28.564 98 =/b/ 459 2 <d> =/d/ 100.46 93 =/h/ 7.580 7	bon + <bb> (abbé et mots de la famille) obscur (+ /#; plomb) dur (+ <dd>; addition et sa famille) fond (FV: 12%; SFD: 60%); foulard (24%)	<f> =/f/ 16.507 >99 <ge> =/ʒ/ 329 100 =/ʒ/ 10.780 47 <v> =/z/ 713 95 =/ʃ/ 160.533 >99 =/h/ 456 =/v/ 13.732 88	jeu (+ blue-jean; /dʒ/) [RC] mangéons (FV); bourgeois [RC] gel (+ suggérer; /ʒ/) lac 6 mots: fusil, outil, gentil, persil; fils elle, nulle; ville (+ fille; /ij/: 12%)
<g> =/g/ 11.109 48 =/h/ 1.085 5 <gu> =/g/ 1.563 96	[RC] gare (+ <gg>; toboggan) long (SFD: 46%); étang (54%) guide (+ /gy/ aiguille)	<v> =/z/ 713 95 =/ʃ/ 160.533 >99 =/h/ 456 =/v/ 13.732 88	vol wagon /v/, western /w/
<h> =/h/ 58.937 >99 <mm> =/m/ 6.057 100 <n> =/n/ 45.517 100 <nn> =/n/ 5.284 98 <gn> =/n/ 2.728 100	mur (+ <mm> = /#; automne) pomme nid année (+ <enn> = /ân/, ennui: 2%) gagner (+ <ign>; oignon [OR=agnon])	<v> =/z/ 713 95 =/ʃ/ 160.533 >99 =/h/ 456 =/v/ 13.732 88	rue (mots finissant par <er>; Annexe A) barre
<h> =/h/ 6.898 100	heure; bonheur	<v> =/z/ 713 95 =/ʃ/ 160.533 >99 =/h/ 456 =/v/ 13.732 88	vol wagon /v/, western /w/

4. CONCLUSION

La progression proposée est centrée sur la mise en place du décodage. 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. Le travail sur la compréhension d'énoncés et de textes est, bien entendu, indispensable. Toutefois, il devra s'effectuer principalement à l'oral, tant que les enfants n'ont pas bien automatisé le décodage. Il n'est évidemment pas davantage possible de travailler sur la production d'énoncés ou de textes, étant donné le retard de l'écriture sur la lecture (Fayol, 2013 : 96). Le fait de travailler la compréhension principalement à l'oral au début de l'apprentissage de la lecture n'est pas un problème, le processus de compréhension étant très largement amodal, *i.e.* similaire à l'écrit et à l'oral, comme l'a clairement montré l'étude princeps de M. A. Gernsbacher, K. R. Varner et M. E. Faust (1990).

En conclusion, la base de données Manulex-MorphO peut être utilisée pour construire du matériel pour l'enseignement (entre autres, par les concepteurs de manuels scolaires) ou pour la recherche. Elle peut également être utilisée pour la formation des enseignants. Quel que soit le contexte, il faut souligner que son apport principal est dans la prise en compte de la morphologie, qui a pour conséquence d'augmenter la consistance des CGPh et, surtout, celle des CPhG. Si, de plus, l'on considère que les règles contextuelles n'ont pas été intégrées dans Manulex-MorphO, l'orthographe du français est, contrairement à une idée reçue, plutôt consistante, tout au moins pour la lecture.

Dans le prolongement du travail décrit dans cet article, le projet Morphiste (Peereman & Sprenger-Charolles) cible la morphologie dérivationnelle dans le lexique écrit de l'enfant (au-delà du cas des morpho-phonogrammes). Plusieurs travaux indiquent, en effet, un impact des connaissances morphologiques de l'enfant sur les choix orthographiques qu'il effectue. Par exemple, S. Pacton, M. Fayol et P. Perruchet (2005) observent chez des enfants francophones que l'utilisation de l'orthographe <ette> (et non <aite>, <ête>) augmente lorsque la terminaison /Et/ est identifiée comme un suffixe diminutif dans l'item à orthographier (v. aussi Casalis & Colé 2018 dans ce numéro). L'objectif principal de Morphiste est de caractériser statistiquement les différentes variables morphologiques sur l'ensemble des corpus lexicaux du CP au CM2 : par exemple, diversités et fréquences des préfixes et suffixes, ainsi que celles des bases et racines.

Références bibliographiques

- CASALIS S. & COLÉ L. (2018), « Le morphème, une unité de traitement dans l'acquisition de la littéracie », *Langue française* 199. (ce volume)
- CATACH N. (1973), « Que faut-il entendre par système graphique du français ? », *Langue française* 20, 30-44.
- CATACH N. (2005), *L'orthographe française*, Paris, Armand Colin.

- CLEMENTS G. N. (1990), "The role of the sonority cycle in core syllabation", in J. Kingston & M. E. Beckman (eds.), *Papers in Laboratory Phonology, 1: Between the Grammar and Physics of Speech*, Cambridge, Cambridge University Press, 283-333.
- DEHAENE S. (éd.) (2011), *Apprendre à lire : des sciences cognitives à la salle de classe*, Paris, Odile Jacob.
- DELATRE P. (1965), *Comparing the Phonetic Features of English, French, German and Spanish: an interim report*, Heidelberg, Jumiuss Gross Verlag.
- ENCREVÉ P. (1988), *La Liaison avec et sans enchaînement : phonologie tridimensionnelle et usages du français*, Paris, Éditions du Seuil.
- FAYOL M. (2013), *L'acquisition de l'écrit*, Paris, Presses Universitaires de France.
- GERNSBACHER M. A., VARNER K. R. & FAUST M. E. (1990), "Investigating differences in general comprehension skills", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16 (3), 430-445.
- GINGRAS M. & SÉNÉCHAL M. (2017), "Silex: A database for silent-letter endings in French words", *Behavior Research Methods* 49 (5), 1894-1904.
- LEGENDE G. *et al.* (2010), "Comprehension of infrequent subject-verb agreement forms: Evidence from French-learning children", *Child Development* 81 (6), 1859-1875.
- LÉTÉ B., SPRENGER-CHAROLLES L. & COLÉ P. (2004), "MANULEX: A grade-level lexical database from French elementary-school readers", *Behavior Research Methods, Instruments, and Computers* 36 (1), 156-166.
- PACTON S. (2008), « L'apprentissage de l'orthographe lexicale du français », dans A. Desrochers, F. Martineau & Y. C. Morin (éds), *Orthographe française : évolution et pratique*, Ottawa, Les Éditions David, 331-354.
- PACTON S., FAYOL M. & PERRUCHET P. (2005), "Children's implicit learning of graphotactic and morphological regularities", *Child Development* 76 (2), 324-339.
- PEEREMAN R. & CONTENT A. (1999), "LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic words", *Behavior Research Methods, Instruments, and Computers* 31 (2), 376-379.
- PEEREMAN R., LÉTÉ B. & SPRENGER-CHAROLLES L. (2007), "Manulex-Infra: Distributional characteristics of grapheme-phoneme mappings, and infralexical and lexical units in child-directed written material", *Behavior Research Methods* 39 (3), 579-589.
- 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.
- POTHIER B. & POTHIER P. (2003), *ÉOLE : échelle d'acquisition en orthographe lexicale*, Paris, Retz.
- SPRENGER-CHAROLLES L. *et al.* (2003), "Development of phonological and orthographic processing in reading aloud, in silent reading, and in spelling: A four-year longitudinal study", *Journal of Experimental Child Psychology* 84 (3), 194-217.
- TERS F., MAYER G. & REICHENBACH D. (1977⁵), *L'échelle Dubois-Buyse d'orthographe usuelle française*, Paris, OCDL.
- TREIMAN R. & KESSLER B. (2014), *How Children Learn to Write Words*, Oxford, Oxford University Press.
- VÉRONIS J. (1986), « Étude quantitative sur le système graphique et phono-graphique du français », *Cahiers de Psychologie cognitive* 6, 501-531.
- ZIEGLER J. C., JACOBS A. M. & STONE G. O. (1996), "Statistical analyses of the bidirectional inconsistency of spelling and sound in French", *Behavior Research Methods, Instruments, and Computers* 28 (4), 504-515.