MINISTÈRE
DE L'ÉDUCATION
NATIONALE,
DE L'ENSEIGNEMENT
SUPÉRIEUR
ET DE LA RECHERCHE

Liberté
Égalité
Fraternité

Conseil scientifique
de l'éducation nationale

Note du CSEN —

— Septembre 2025, n°14

Pourquoi et comment faciliter les débuts de l'apprentissage de la lecture et de l'écriture en français ?

Rédigée par Liliane Sprenger-Charolles et Ronald Peereman[a]

Résumé

Quel que soit le système d'écriture, apprendre à lire et à écrire nécessite un enseignement explicite. Dans une écriture alphabétique comme celles de l'anglais, du français ou de l'espagnol, cet enseignement doit s'appuyer sur la consistance et la fréquence des liens graphème-phonème (pour décoder un mot écrit) et phonème-graphème (pour encoder un mot entendu). Pour cela, il faut disposer de statistiques développées à partir du vocabulaire utilisé dans les textes destinés aux enfants. Par exemple, pour le français, *Manulex-Infra* et *Manulex-Morpho* fournissent des informations sur les relations graphème-phonème et phonème-graphème ainsi que, tout au moins pour *Manulex-Morpho*, sur les marques morphophonémiques grammaticales (par exemple, le *s* du pluriel des noms et le *ent* de celui des verbes) et lexicales (par exemple, le *t* de *dent* support des dérivés *dentiste*, *dentition*…). Pour cette raison, après une présentation des principales bases existantes, l'article[b] décrit en détail *Manulex-Morpho* : en premier, les caractéristiques des mots de cette base (catégories, fréquence, longueur…), puis celles des relations graphème-phonème et phonème-graphème (nombre, consistance, fréquence…) et enfin celles des lettres finales des mots en fonction de leur statut (muettes ou non et ayant ou non une fonction morphologique). Ces données permettent de dégager des principes à partir desquels il devient possible de définir une progression pédagogique optimale, susceptible de faciliter le début de l'apprentissage de la lecture ainsi que celui de l'orthographe. Cette progression est confrontée aux résultats d'études qui ont, pour certaines, observé les effets de différentes pratiques pédagogiques sur les débuts de l'apprentissage de la lecture et, pour d'autres, évalué ces effets dans le cadre d'expérimentations utilisant des jeux éducatifs implémentés sur ordinateur.

a **Liliane Sprenger-Charolles**, directrice de recherche, Emérite, CNRS, Centre de Recherche en Psychologie et Neurosciences à Aix Marseille Université et **Ronald Peereman**, chargé de recherche, CNRS, Laboratoire de Psychologie et de Neuro-Cognition à l'université de Grenoble Alpes.

b Nous remercions **Anne Abeillé**, professeure à l'université Paris Cité, Laboratoire de linguistique formelle ; **Stanislas Dehaene**, professeur de sciences cognitives expérimentales au Collège de France ; **Anna Potocki**, maître de conférences en sciences de l'éducation et membre du LaRAC à l'université Grenoble Alpes ; **Johannes Ziegler**, directeur de recherche au CNRS, responsable recherche fondamentale du pôle pilote AMPIRIC à l'université d'Aix-Marseille, pour leurs contributions.

La compréhension écrite dépend à la fois du niveau de compréhension orale de celui qui lit et de sa maitrise de mécanismes spécifiques à la lecture. L'exemple de la lecture d'une partition de musique peut permettre de cerner ce que sont ces mécanismes. En effet, l'incapacité de lire ce type de document n'est en général pas attribuée à des difficultés de compréhension de la musique. Ce qui est supposé faire défaut, ce sont les mécanismes qui permettent au musicien expert d'associer dans sa tête, automatiquement, une suite de notes écrites à un bout de mélodie. Il en va de même pour la lecture : un enfant qui a une bonne compréhension orale ne peut comprendre ce qu'il lit que s'il a automatisé les mécanismes spécifiques à la lecture, ceux qui lui permettent d'identifier les mots écrits. Par conséquent, l'objectif principal de l'enseignement doit être de permettre à l'enfant de parvenir à comprendre ce qu'il lit de la même façon qu'il comprend ce qu'il entend. Pour cela, et tout au moins dans une écriture alphabétique comme celle du français, il faut l'aider à développer des capacités de décodage précises et rapides[c].

Pour construire un apprentissage structuré et systématique des correspondances graphème-phonème (CGPh, utilisées pour décoder les mots en lecture) et phonème-graphème (CPhG, utilisées pour encoder les mots en écriture), il faut disposer de données statistiques fiables. De telles données sont disponibles, par exemple celles issues de travaux de linguistes [8]-[9] et de psycholinguistes [10]-[11]-[12] mais elles présentent des limites. Certaines sont datées [8]-[9]. Dans d'autres, la consistance[d] et la fréquence des CGPh et des CPhG n'ont été calculées que pour des mots monosyllabiques [10]-[11], ou encore proviennent de corpus issus du lexique écrit adressé à l'adulte [12]. Il existe également des bases de données provenant du Canada, mais elles portent sur un lexique relativement différent de celui du français [13]. L'intérêt de *Manulex-Infra* [14] et *Manulex-Morpho* [15] provient du fait que leur corpus est issu de *Manulex* [16], une base de données qui prend en compte le lexique écrit adressé à l'enfant de l'école élémentaire.

Après une introduction signalant les principales bases de données sur l'orthographe du français, la suite est centrée principalement sur *Manulex-Morpho*. Ce choix s'explique par le fait que cette base contient des statistiques sur les relations entre graphèmes et phonèmes et celles entre phonèmes et graphèmes ainsi que sur les marques morphologiques. Ces marques (par exemple le <d> de grand) posent un problème au jeune enfant qui apprend à lire et, surtout, à écrire dans la mesure où elles doivent s'écrire alors qu'elles sont souvent muettes à l'oral en français. La version remaniée en 2023 de cette base (V2)[e] comprend deux documents de format xls :

- la base des **données lexicales** : les mots avec leur catégorie, leur fréquence, leur longueur et leur consistance ;
- celle des **données infra-lexicales**, avec la consistance et la fréquence des CGPh et des CPhG, y compris pour les marques morpho-phonémiques.

Sont présentées après cette introduction les principales caractéristiques des unités lexicales du corpus de *Manulex-Morpho* (les mots) puis celles des unités infra-lexicales en soulignant les spécificités de l'orthographe du français : d'une part, la présence de nombreux digraphes qui transcrivent un seul phonème, certains n'ayant pas d'équivalent plus simples (<ou>, <on>, <ch>, <gn>[f]…) ; d'autre part, le fait que de nombreuses lettres qui doivent s'écrire en fin de mots ne se prononcent pas. Ces données permettent de dégager des principes à partir desquels il devient possible de définir un ordre de présentation des CGPh susceptible de faciliter l'apprentissage de la lecture. C'est ce qui est au cœur de l'avant-dernière section de l'article qui présente les résultats d'études qui ont, pour certaines, observé les effets de différentes pratiques pédagogiques sur les débuts de l'apprentissage de la lecture et, pour d'autres, évalué ces effets dans le cadre d'expérimentations utilisant des jeux éducatifs implémentés sur ordinateur.

Données lexicales et infra-lexicales : l'exemple de *manulex_morpho*

Manulex-Morpho comprend une base lexicale (les mots) et deux bases infra-lexicales (l'une sur les correspondances entre graphèmes et phonèmes pour le décodage, l'autre sur celles entre phonèmes et graphèmes pour l'encodage) qui intègrent des informations sur les marques morpho-phonémiques[g] : entre autres, les flexions nominales et verbales du nombre (<s>, <x>, <ent>) ainsi que les pivots de flexion ou de dérivation (<d> : *rond, ronde, rondeur*). Ces marques, qui doivent s'écrire alors qu'elles sont souvent muettes à l'oral, sont source de difficultés pour l'apprentissage de la lecture et, surtout, pour celui de l'écriture [17]-[18]-[19].

La base des données lexicales : les mots

Il est nécessaire d'utiliser des mots que les élèves sont en mesure de comprendre, et donc des mots qui appartiennent au lexique écrit qui leur est adressé. Or la référence disponible sur le site du Ministère de l'Education

c . Voir les notes du Conseil Scientifique de l'Education Nationale, CSEN [1]-[2]-[3], celles du Ministère de l'Education Nationale, MEN [4]-[5] et celles de l'observatoire National de la Lecture, ONL [6] ; voir aussi [7].

d Le substantif *consistance* est issu du verbe *consistere* (emprunté au latin classique) qui signifie « se tenir ensemble » (https://www.cnrtl.fr/etymologie/consister). Dans la présente note, comme dans la plupart des études sur les correspondances entre graphèmes et phonèmes, il est utilisé à la place de *régularité*.

e La version remaniée en 2023 (V2) de *Manulex-Morpho* est accessible via github (https://inframorph.github.io/morpho/telechargement.html) avec celle de *Manulex-Infra*.

f Les graphèmes sont entre crochets et les phonèmes entre des barres obliques (<ch>-/S/). L'annexe présente les transcriptions de l'Alphabet Phonétique International (API) ainsi que celles des bases informatisées.

g Marques morphologiques correspondant à un phonème. Exemples : <er> des verbes du 1er groupe mais par <ir> de ceux du 2nd groupe pour les flexions ; <d> de *grand* mais pas <eur> dans *grandeur* pour les dérivations.

Nationale est celle de Brunet [20][h] , qui, comme l'indique son titre (*Le vocabulaire français de 1789 à nos jours*), a pour objectif d'évaluer l'évolution du vocabulaire dans le temps (en l'occurrence, de 1789 à 1969). Ce corpus ne peut donc pas ''rend[re] compte de la langue que lisent les élèves francophones'' d'autant plus que les mots retenus ''ont été ramenés à leur base lexicale'' (c'est-à-dire lemmatisés, sans les marques de flexions nominales ou verbales (comme le <s> ou le <ent> pour le pluriel). En conséquence, parmi les 100 mots les plus fréquents de *Manulex-Morpho*, 25% ne sont pas dans Brunet : entre autres, les déterminants et pronoms portant les marques du genre et du nombre (*la, les, une, ils…*) et les verbes conjugués (*est, dit, va…*).

Le corpus de *Manulex_Morpho* [15] est issu de *Manulex* [16]. Il intègre les 10.000 mots les plus fréquents de cette base, soit environ 20% des 50.000 entrées lexicales mais plus de 90% des entrées textuelles, celles qui tiennent compte de la répétition des mots dans le corpus. La base lexicale de *Manulex-Morpho* présente les mots en fonction de leur consistance, de leur fréquence, de leur longueur (nombre de lettres, de graphèmes, de phonèmes et de syllabes), et de leur structure syllabique. Un résultat notable est que les statistiques issues de l'ensemble de ce corpus diffèrent fortement de celles provenant des 100 mots les plus fréquents de cette base (cf. Tableau 1). Ainsi, alors que la plupart des mots du corpus (à l'exception des noms propres) sont des noms (45,71%), des verbes (35,87%) et des adjectifs (14,01%), la proportion des termes grammaticaux est faible (4,41%). Ces derniers, souvent dénommés *mots outils* dans la tradition pédagogique (en l'occurrence les adverbes, les pronoms, les déterminants, les prépositions et les conjonctions) sont en revanche omniprésents parmi les 100 mots les plus fréquents de la base (66) qui ne contient que 14 noms, 8 adjectifs et 12 verbes.

Un constat identique ressort des statistiques sur la longueur des mots. Alors que, dans l'ensemble du corpus, presque la moitié des mots sont bisyllabiques (48,8%), les monosyllabiques sont sur-représentés parmi les 100 mots les plus fréquents (90%). De même les mots de 5 à 8 lettres dominent dans l'ensemble du corpus alors que 82% des 100 mots les plus fréquents ont moins de 5 lettres. Ces données permettent de cerner les items qu'il faudrait utiliser en début de progression compte tenu des caractéristiques des plus consistants et fréquents.

Tableau 1. *Manulex-Morpho* : Mots par catégorie et longueur (au total et les 100 les plus fréquents)

Par catégorie	%	100 les + fréquents	Longueur Lettres	%	100 les + fréquents	Longueur NbSyllabes	%	100 les + fréquents
Noms communs	45,71	11	1	0,03	3	1	22,68	90
Verbes	35,87	14	2	0,62	31	2	48,75	10
Adjectifs	14,01	8	3	2,55	25	3	23,38	
Adverbes	2,70	11	4	7,09	23	4	4,54	
Pronoms	0,80	23	5	13,90	13	> 5	0,65	
Déterminants	0,29	18	6	19,87	4			
Prépositions	0,41	9	7	19,17	1			
Conjonctions	0,21	6	8	15,11				
Mots grammaticaux (adv, pro, dét, prép, conj)	4,41	67	9 >10	10 11,65				

h https://eduscol.education.fr/186/liste-de-frequence-lexicale

La base de données infra-lexicales : relations graphème-phonème et phonème-graphème

En dehors des problèmes posés par la morphologie, une autre difficulté de l'apprentissage de la lecture-écriture provient de ce que le nombre de phonèmes du français dépasse celui des lettres de l'alphabet, surtout pour les voyelles (6 lettres dans l'alphabet pour les 10 à 16 voyelles en français). Ainsi, pour les voyelles orales, en plus du /i/ (de *rire*), du /u/ (de *fou*) et du /y/ (de *pur*), le français a 4 voyelles de cette nature qui peuvent, selon les locuteurs, correspondre à 1 ou 2 phonèmes : /e-ɛ/ de *célèbre*, /a-ɑ/ de *patte* et *pâte*, /ɔ-o/ de *pot* et *port* et /ø-œ/ de *deux* et *neuf*. Également selon les locuteurs, les voyelles nasales sont au nombre de 3 (/ɑ̃/ pour *an*, /ɔ̃/ pour *on* et /ɛ̃/ pour *fin*) ou 4 (pour ceux qui différencient /ɛ̃/ de *brin* de /œ̃/ de *brun*). Toujours selon les locuteurs, la voyelle dite muette (/ə/) peut être ou non prononcée : *gare* est un monosyllabique à Paris, pas à Marseille.

Les 3 semi-voyelles du français (/ɥ/, /w/ et le yod /j/) peuvent aussi s'écrire de plusieurs façons, sauf le /ɥ/ qui s'écrit toujours <u> (*lui*, *nuit*…). Ainsi, /w/+/a/ s'écrit (entre autres) <oi>, <wa>, <oua>, <oê> (*loi*, *water*, *ouate*, *poêle*…). De même le yod s'écrit <i> ou <y> (*iode* et *yod*), ces deux graphèmes pouvant aussi se lire /i/ (*si* et *jury*). Toutefois le yod peut également s'écrire <il> (cf. *ail*), <ll> (*fille*) ou encore avec un <y> qui transcrit deux /i/ (*paye* : /a+i/ qui se lit <é> suivi par un yod). En outre, la voyelle qui précède ces semi-voyelles peut s'entendre plus ou moins (*biais* et *billet*, *pied* et *piller*, *trois* et *troua*…).

Ces spécificités du français expliquent la présence de nombreux digraphes dans cette langue qui ne sont pas pris en compte dans les débuts des progressions classiques. Ces dernières partent des lettres de l'alphabet et n'intègrent que dans une seconde étape les digraphes. Or pour certains d'entre eux (<ou>-/u/, par exemple) la correspondance graphème-phonème est non seulement fréquente mais également très consistante. De plus /u/ est plus facile à produire que /y/ qui a une prononciation pouvant poser des problèmes aux élèves allophones parlant une langue dans laquelle le phonème /y/ n'existe pas.

En outre, il a été observé que la présence de digraphes *fréquents* facilite la précision et la rapidité du décodage [21]. C'est le cas, par exemple, à nombre de lettres équivalent, des mots tels que *poudre, tante, sapin, écharpe, oncle*… par rapport à *pilote, tulipe, samedi, avril, abri*. D'après d'autres études [22] déstructurer l'unicité des digraphes (*champignon* écrit ***c ha mp ig no n*** et non ***ch am p i gn on***) a un effet négatif sur la précision et la rapidité de la réponse qui est plus notable quand les items *et* la tâche sollicitent fortement la phonologie (pseudomots versus mots ; dénomination vs. décision lexicale). Cette étude [22] (voir aussi [23]) signale également que les digraphes sont traités dans une aire spécifique de l'hémisphère gauche du cerveau, proche, mais distincte, de la zone spécialisée pour la lecture (la Visual Word Form Area : VWFA [7], [24]).

Pour construire une progression pédagogique, il faut donc s'appuyer sur des statistiques fiables. L'encadré 1 explique comment ces statistiques ont été calculées pour les CGPh et CPhG dans *Manulex*.

ENCADRÉ 1 : Consistance des relations graphème-phonème et phonème-graphème

La consistance lexicale des correspondances graphème-phonème (CGPh) est le rapport entre la fréquence propre d'une relation spécifique et la fréquence globale du graphème. Par exemple, dans 553 mots du corpus <ch> se prononce /ʃ/ alors que dans 9 il se prononce /k/, la consistance lexicale <ch>-/k/ est donc de 1,60% (9/[553 + 9]) x 100 = 1,60%) et celle de <ch>-/ʃ/ de 98,40%. **La consistance textuelle** est obtenue par le même calcul mais à partir des données tenant compte du nombre de répétition des mots dans le corpus. Si parmi les mots contenant <ch> ceux dans lesquels ce graphème se prononce /ʃ/ sont plus fréquents que ceux dans lesquels il se prononce /k/, la consistance textuelle de l'association <ch>-/ʃ/ va augmenter et celle de l'association <ch>-/k/ va baisser. Sachant que la fréquence totale est de 184 pour les 9 mots dans lesquels <ch> se prononce /k/ et de 22.388 pour les 553 mots dans lesquels ce graphème se prononce /ʃ/, la consistance textuelle de <ch>-/k/ est de 0,82% (184/[22.388 + 184]) x 100 = 0,82%) et celle de <ch>-/ʃ/ de 99,18%.

Exemple	Consistance lexicale (%)	Consistance textuelle (%)
ch-/k/ (chorale)	1,60	0,82
ch-/ʃ/ (chat)	98,40	99,18

La consistance, lexicale et textuelle, des CPhG se calcule de la même façon mais en partant du phonème versus le ou les graphèmes qui correspondent.

Comme l'indique le tableau 2, les correspondances graphème-phonème (CGPh) sont moins nombreuses que celles phonème-graphème (CPhG) : 187 CGPh contre 263 CPhG, aussi bien pour les voyelles (88 CGPh vs. 122 CPhG) que pour les consonnes (76 CGPh vs. 95 CPhG) et les semi-voyelles (22 CGPh vs. 30 CPhG). Toutefois, parmi l'ensemble des CPhG qui codent les phonèmes du français, certaines ne se rencontrent que dans moins de 10 occurrences. Pour les **voyelles**, c'est par exemple le cas du /a/ écrit <e> dans *femme*, ou encore du /u/ écrit <oo> dans *foot*. Pour les **consonnes**, on peut citer celles qui sont doublées, sans incidence sur leur prononciation : <r> et <rr> se lisent toujours /r/ mais pour écrire selon les normes, il faut savoir qu'il n'y a qu'un seul <r> à *mourir* mais deux à *nourrir*.

Ces données indiquent qu'il est plus facile d'apprendre à lire que d'apprendre à écrire en français. Elles suggèrent aussi des pistes de simplification concernant, entre autres, les singletons[i] ou quasi singleton (Cf. pour des explications et des propositions, la note de synthèse du CSEN *Rationaliser l'orthographe pour mieux l'enseigner* [18], voir aussi [19]).

Tableau 2. Nombre de voyelles, de semi-voyelles et de consonnes dans CGPh et les CPhG

	CGPh 187	CPhG 263	Exemples
Voyelles : 10 à 16	**89**	**122**	Le graphème <é> se lit toujours /e/ (sauf le 2nd <é> d'*événement*, corrigé dans les rectifications de 1990) Le phonème /e/ peut s'écrire d'au moins 10 différentes manières : <é> dans *été*, <er> dans *parler*, <et> dans *et*, <ez> dans *nez*, <ai> dans *mai*, <ef> dans *clef* (qui s'écrit aussi clé), <ë> dans *canoë* (corrigé dans les rectifications de 1990), <a> dans *lady*, <ers> dans *volontiers*…
Semi-Voyelles : 3	**22**	**30**	Suivi par une voyelle, le graphème <u> se lit (/ɥ/) et /ɥ/ s'écrit <u> (cf. *lui, nuit, bruit*…) Le <oi> se lit /w+a/ (loi…) mais /w+a/ peut aussi s'écrire <wa> ou <oua> (*water, ouate*…) Les graphèmes <i> et <y> peuvent se lire /i/ ou /j/ (cil – *ciel* ; y – *yeux*) mais il y a plus de 10 façons d'écrire /j/ (<i> dans *ciel*, <y> dans *yeux*, <il> dans *ail*, <ll> dans *fille*…)
Consonnes : 17	**76**	**95**	Les différences proviennent surtout de celles qui sont doublées
Lettre muettes	**0**	**16**	Graphèmes ne correspondant à aucun phonème : <a> (*aout*), <p> (*sept*), <t> (*art*), <s> (*ans*)

La morphologie et les relations entre graphèmes et phonèmes (lecture) versus celles entre phonèmes et graphèmes (écriture)

Il ressort des calculs précédents que les CPhG sont plus nombreuses que les CGPh. Il est donc plus difficile d'apprendre à écrire que d'apprendre à lire en français. Toutefois, comme le signale la figure 1, la prise en compte de la morphologie permet de réduire cette différence : le gain de consistance en fin de mot est de plus 30% pour les CPhG contre moins de 2% pour les CGPh.

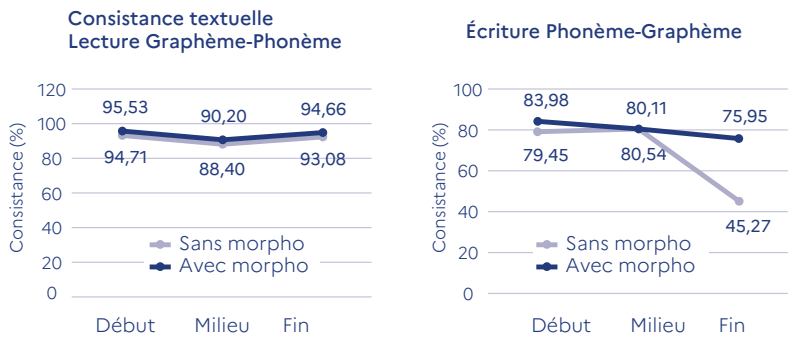

Figure 1. Différence de consistance textuelle des CGPh et CPhG en fonction de la position dans le mot et de la prise en compte des marques morpho-phonémiques

i Ce terme est issu de l'anglais *single* qui lui-même vient du français (cf. *singulier, singularité*).

Statut des lettres finales muettes dans les noms de *Manulex-Morpho*

Les résultats d'une analyse qui a pris en compte les lettres finales muettes (pas les graphèmes) qui étaient les plus fréquentes dans les noms du corpus de *Manulex-Morpho* sont présentées dans la figure 2. Les résultats de cette analyse indiquent que les lettres les plus fréquentes sont <e> et <s> (respectivement 37,4 et 33% des cas). Toutefois <e> n'est que rarement une marque morphologique indiquant le genre (4,8% des occurrences, par exemple, pour ceux désignant un métier : *bergère, chanteuse, couturière, directrice, employée, maitresse, marchande, paysanne, servante*…).

En revanche, <s> signale presque toujours un pluriel, voire une marque de dérivation (95,4% des cas, les exceptions se trouvent dans quelques noms tels que *héros, jus, dessus, sous, dessous, poids, temps, puits, parcours, secours* ou encore *repas, repasser* n'étant pas un dérivé de *repas*).

Il en est de même pour <x>, qui est rare (2,3%) mais, comme <s>, signale le plus souvent une marque de flexion, voire de dérivation (96,3%). Il faut toutefois rappeler que l'utilisation de <x> en tant que marque morphologique résulte d'une erreur [18]-[19][j] et que seuls 7 noms en <ou> ont un pluriel en <x> (les fameux *bijou, caillou, chou, genou, hibou, joujou, pou*). En revanche, le <x> est majoritaire dans les noms en <(e)au> et <eu>, avec des exceptions (*landaus, bleus, pneus*).

Le <n> (10,5% des finales muettes), le plus souvent intégré dans une voyelle nasale (*paysan, voisin, bon, brun*…), est, dans la moitié des cas (52,3%), un support de flexion ou de dérivation (cf. *voisin-voisine-voisinage*).

Le <t> et le <r> (respectivement, 8,3% et 2,6% des lettres muettes) ne se prononcent pas surtout quand ils sont intégrés dans les graphèmes <(i)er> ou <et>. Le premier est un support de flexion ou de dérivation dans 29,6% des occurrences : dans des noms qui sont aussi des verbes (*le diner, le déjeuner, le toucher*…) ainsi que dans certains noms de métier (*le boucher, le teinturier*…) ou d'arbres (*le cerisier, le peuplier, le pommier*…). Dans 61,9% des cas, <t>-<et> sont des morpho-phonogrammes (*art, cent, lot ; sujet, muet*…).

Les noms de ce corpus se terminant par <d> muet sont rares (1,1%) mais ce graphème note presque toujours une marque morphophonémique (87,5% des cas : *bond, bord, froid, grand*), qui est souvent dans une rime en <ard> (cf. *lézard, regard, retard*…).

Enfin, ceux qui se terminent par un <l> muet, qui eux aussi sont rares (0,8%), proviennent dans plus de 59,3% des cas du <il> qui code un yod (cf. *travail, réveil, sommeil*…) ou se transforme en un yod (*fusil - fusillade* ; *outil - outillage*…).

Il faut ajouter <u> à la série des lettres finales muettes : présent dans un nombre non négligeable d'occurrences (2,5% des lettres muettes), muet parce qu'il est le plus souvent intégrée dans un digraphe : <au> ou <eau> (*noyau, peau*…), <eu> (*bleu, feu*…), <ou> (*bijou, fou, genou*…)[k].

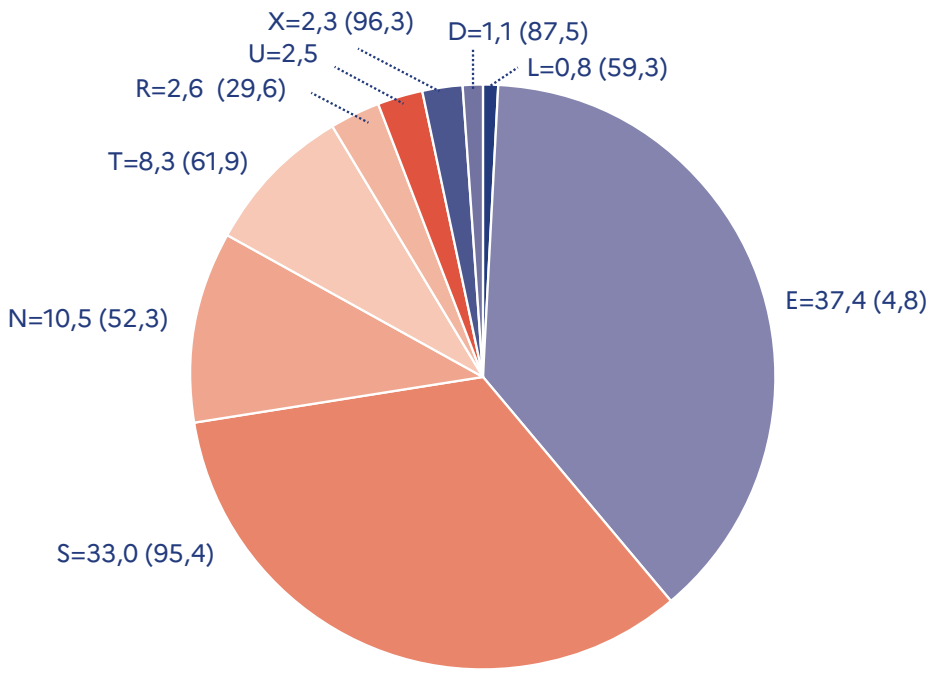

Figure 2. Lettres muettes (% sur total des lettres muettes) et lettres morphophonémiques (% sur le nombre d'occurrence de la lettre cible) : Fin des noms de *Manulex-Morpho* (hors noms propres)

j La lettre <x> était initialement une abréviation pour les terminaisons latines en <us>

k Les autres lettres muettes relevées dans les noms de *Manulex-Morpho* (<m>, <p>, <g>, <z>, <f>, <b>, <k>, <j>, <v> et <w>) ne représentent que 1,4% du corpus. Elles ne sont pas intégrées dans le graphique.

Correspondances de base nécessaires pour apprendre à lire en français

Le tableau 3a présente les correspondances graphème-phonème (CGPh) qu'il faut enseigner en priorité pour apprendre à lire en français : les plus consistantes et les plus fréquentes. Les autres CGPh qui pour la plupart sont régies par des règles contextuelles, sont intégrées dans le tableau 3b.

D'après le tableau 3a, et si on ne tient pas compte de la différence entre son vocalique ouvert et fermé pour <a>, <o> et <eu>, 7 graphèmes fréquents ont une prononciation consistante, les plus nombreux étant <a>, <i>, <o> et <é>. La valeur du <ou> et celle du <u> sont proches. Toutefois <u>, comme <i>, sont peu consistants en milieu de mot parce que, dans ce contexte, <i> peut coder la semi-voyelle de ciel (/j/) et <u> celle de lui (/ɥ/). A cet ensemble s'ajoutent 4 voyelles nasales, 2 étant très consistantes (<an> et <on>). D'après le même tableau, 16 graphèmes consonantiques ont une prononciation consistante, sauf pour 12 d'entre eux en fin de mots. Enfin <s> (qui est fréquent), comme <gn> (qui est rare), ne sont consistants que dans une position.

Tableau 3a. Relations graphème-phonème les plus consistantes (> 85%) et les plus fréquentes (> 300)

Consistance par position									Fréquence	Fréquence par position					
			Début		Milieu		Fin		Total	Début		Milieu		Fin	
		Exemp	Type[l]	Tok	Type	Tok	Type	Tok		Type	Tok	Type	Tok	Type	Tok
Voyelles															
1	<a>-/a/A/	ami	**99,8**	**99,8**	**99,9**	**99,9**	**100**	**100**	**2975**	**623**	**831**	**2331**	**3030**	21	29
2	<o>-/O/	sol	**95,1**	**95,7**	**89,5**	**89,0**	0	0	**1398**	117	152	**1281**	**1665**	0	0
	<o>-/o/	saut	4,9	4,3	10,5	11,0	**100**	**100**	194	6	7	148	203	40	56
3	<é>-/e/	été	**100**	**100**	**99,8**	**99,7**	**100**	**100**	**1301**	**303**	**383**	**884**	**1101**	114	148
4	<ou>-/u/	ours	**87,5**	**87,4**	**93,8**	**94,4**	**100**	**100**	**767**	6	51	**698**	**976**	34	48
5	<u>-/y/	bulle	**100**	**100**	**75,8**	**74,9**	**100**	**100**	**748**	32	45	**689**	**881**	27	36
6	<i>-/i/	ami	**100**	**100**	72,5	72,3	**100**	**100**	**2318**	64	88	**2183**	**2810**	71	121
7	<eu>-/2/	deux	**100**	**100**	34,7	34,3	**100**	**100**	153	1	3	127	174	25	46
	<eu>-/9/	neuf	0	0	65,3	65,7	0	0	239	0	0	239	**334**	0	0
8	<on>-/§/	ronde	**100**	**100**	**99,0**	**98,5**	**100**	**100**	**579**	6	11	**394**	**518**	179	224
9	<an>-/@/	tante	**100**	**100**	**100**	**100**	**92,3**	**92,0**	**473**	23	27	**438**	**592**	12	17
10	<in>-/5/	cinq	**100**	**100**	**100**	**100**	**100**	**100**	240	157	190	64	84	19	25
11	<un>-/1/	un	0	0	**100**	**100**	0	0	3	0	0	3	4	0	0
Consonnes															
1	<l>-/l/	lama	**100**	**100**	**99,9**	**99,9**	**95,2**	**96,6**	**1767**	270	**398**	**1477**	**1911**	20	35
2	<r>-/R/	roche	**100**	**100**	**100**	**100**	**99,3**	**99,0**	**4845**	**779**	**995**	**3771**	**4970**	295	**423**
3	<f>-/f/	feu	**100**	**100**	**100**	**100**	66,7	65,8	**651**	**445**	**620**	198	245	8	12
4	<ch>-/S/	chat	**98,6**	**98,7**	**98,3**	**98,5**	50,0	61,3	**553**	211	306	**341**	**431**	1	2
5	<v>-/v/	ville	**100**	**100**	**100**	**100**	0	0	**1128**	**365**	**535**	**763**	**1008**	0	0
6	<j>-/Z/	jeu	**99,1**	**99,4**	**99,3**	**100**	0	0	166	106	165	60	87	0	0
7a	<s>-/s/	sol	**100**	**100**	52,4	52,3	15,3	12,8	**1213**	**645**	**899**	**551**	**706**	17	27
7b	<s>-/z/	rose	0	0	47,6	47,7	0	0	**500**	0	0	**551**	**706**	0	0
8	<d>-/d/	dur	**100**	**100**	**99,9**	**99,8**	3,4	2,4	**1319**	**600**	**817**	**718**	**913**	1	1
9	<b>-/b/	balle	**100**	**100**	**96,1**	**96,1**	50,0	62,0	**892**	**421**	**568**	**470**	**601**	1	1
10	<p>-/p/	page	**100**	**100**	**97,8**	**97,4**	28,6	22,3	**1810**	**1064**	**1481**	**744**	**964**	2	3
11	<t>-/t/	terre	**100**	**100**	**91,0**	**91,5**	5,1	4,6	**2489**	**489**	**677**	**1993**	**2589**	7	9
12	<qu>-/k/	qui	**100**	**100**	**99,3**	**99,3**	0	0	**331**	50	90	281	**359**	0	0
13	<n>-/n/	note	**100**	**100**	**100**	**100**	**100**	**100**[m]	**877**	162	235	**711**	**908**	4	5
14	<m>-/m/	main	**100**	**100**	**99,8**	**99,8**	0	0	**1458**	**577**	**805**	**881**	**1140**	0	0
15	<gn>-/N/	bagne	0	0	**100**	**100**	0	0	123	0	0	123	160	0	0
16	<z>-/z/	zéro	**100**	**100**	**100**	**100**	0	0	33	6	8	27	37	0	0

l Fréquence par **type** (fréquence lexicale) et par **token** (fréquence textuelle, tenant compte du nombre de répétition du mot dans le corpus).
m Pour 4 mots issus de l'anglais : *clown* au singulier et au pluriel, *western* et *jean*

Le tableau 3b intègre les autres CGPh nécessaires pour apprendre à lire en français, celles qui sont fréquentes sans être consistantes et qui, en conséquence, ne sont pas dans le tableau 3a. C'est le cas de la voyelle la plus fréquente du français, <e>. En effet, cette voyelle se prononce au mieux comme le <eu> de *feu* quand elle est dans un mot monosyllabique (*ce, de, je*…) ou quand elle fait partie d'une syllabe dans des mots plurisyllabiques tels que *table* ou *arbre*. En outre, dans cette position et selon l'origine du locuteur, elle peut aussi être muette dans des mots tels que *gare* et elle l'est toujours dans des mots comme *lycée*. Dans d'autres positions, le <e> peut se prononcer /e-ɛ/. Deux des trois graphèmes qui codent les semi-voyelles font partie des autres correspondances nécessaires pour apprendre à lire en français : <oi> pour /wa/ et <i> pour le yod (tout au moins celui de *ciel*). Enfin, 2 consonnes fréquentes présentes dans l'alphabet sont dans ce tableau : <c> et <g>. Comme pour <s>, la prononciation de ces consonnes est régie par des règles contextuelles.

Tableau 3b. Autres relations graphème-phonème nécessaires pour apprendre à lire en français

Consistance par position> 85%									Fréquence frq> 300)	Fréquence par position					
			Début		Milieu		Fin			Début		Milieu		Fin	
		Exemp	Type	Tok	Typ	Tok	Typ	Tok	Total	Type	Tok	Typ	Tok	Typ	Tok
Voyelles et semi-voyelles															
17	<e>-/%/	table	0	0	37,2	37,8	14,4	15,4	**956**	0	0	**593**	**787**	**363**	**546**
	<e>-/°/	gare	0	0	60,8	60,3	**99,7**	**99,3**	**2350**	0	0	**357**	**446**	**1993**	**2751**
	<e>-/#/	lycée	0	0	1,2	1,2	6,2	6,1	176	0	0	19	24	157	216
18	<i>-/j/	ciel	0	0	25,8	26,1	0	0	**777**	0	0	**777**	**1015**	0	0
19	<oi>-/wa/	oiseau	71,4	79,8	**100**	**100**	**100**	**100**	283	5	10	259	**383**	19	32
Consonnes															
20a	<c>-/k/	car	**88,9**	**88,1**	56,6	56,3	**90,0**	**92,5**	**1392**	**822**	**1100**	**561**	**722**	9	16
20b	<c>-/s/	cil	11,1	11,9	41,7	42,0	0	0	**516**	103	149	**413**	**538**	0	0
21a	<g>-/g/	gare	77,7	78,2	33,2	32,3	25,0	19,5	**412**	220	287	191	241	1	1
22b	<g>-/Z/	gel	21,9	21,4	66,0	66,4	0	0	**442**	62	79	**380**	**494**	0	0

Progressions respectant au mieux la consistance et la fréquence des CGPh

Dans un premier temps, devraient être présentées les correspondances graphème-phonème (CGPh) les plus consistantes et les plus fréquentes et, dans un second temps, celles qui dépendent de règles contextuelles ou de régularités statistiques non motivées phonologiquement. Cette progression devrait s'appuyer sur des mots présentant les caractéristiques des plus fréquents.

Quelles sont les CGPh à enseigner en priorité au début de l'apprentissage du décodage et dans quel type de mots ?

- Les voyelles nasales :
  - → <é>, <ou>, <a>, <o> et <eu>, sans différencier, pour les trois dernières, les voyelles ouvertes de celles qui sont fermées ;
  - → <i>, sauf lorsque cette voyelle, suivie par une autre voyelle, est un yod (*miel*) ;
  - → <u>, sans faire la différence entre le <u> de lu (/y/) et celui de lui (/ɥ/) ;
  - → <e>, seulement quand il est à la fin d'un monosyllabique (cf. *le, je, te*…).
- Les voyelles nasales :
  - → <on> et <an>, mais pas <en> qui est un allographe de <an> (*en*) et de <in> (*bien*) ;
  - → <in> et <un>, ce dernier étant fréquent dans le vocabulaire scolaire (*un, 1, lundi*).
- Les consonnes :
  - → Une priorité devrait être accordée aux fricatives <f>, <ch>, <j> et <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. Il en est de même pour le <s> mais uniquement en début de mot.

→ En raison de leur fréquence, il faut également intégrer parmi les consonnes prioritaires <l> et <r>.
→ Suivent les consonnes occlusives <p>, <b>, <d>, <qu> ainsi que <t> (sauf quand il se prononce /s/ comme dans *solution*…) et les nasales <m>, <n> et <gn>.

- Ces CGPh doivent être insérées dans des mots respectant les caractéristiques présentées dans la section précédente, à savoir des mots ayant des syllabes simples comprenant une consonne à l'initiale suivie par une voyelle (CV : *la, le, jeu, feu, chou*), les CVC étant introduits dans un second temps (*bal, sol, jour*), puis les CCV (*tri, blé, bleu*) et CCVC (*fleur, pleur*). En outre, pour focaliser l'attention des enfants sur l'importance de l'ordre des graphèmes, il est nécessaire de travailler sur les couples CV-VC (tu-ut, *fou-ouf*) et CVC-CCV (*tir-tri, dur-dru*), mais en tenant compte du fait que les monosyllabiques VC sont rares en français.

Quelles sont les CGPh à enseigner dans un second temps de l'apprentissage du décodage ?

Il faudrait aborder les difficultés de l'orthographe du français, entre autres celles qui s'expliquent soit par des règles contextuelles, soit par des régularités statistiques non motivées par la phonologie.

- CGPh dépendant de règles contextuelles

Parmi les plus connues, on relève celles qui régissent la prononciation du <c> et du <g> qui est dite dure (c'est-à-dire celle d'une occlusive) ou douce (c'est-à-dire celle d'une fricative) en fonction du graphème qui suit. Ainsi devant une autre consonne et devant les voyelles <a>, <o>, <u> et tous les graphèmes qui commencent par ces voyelles, <c> et <g> se prononcent comme des occlusives (*clou, cri, car, col, culture, coup, conte*… ; glace, grave, *gare, golfe, guerre, gant*). Dans les autres cas ils se prononcent comme des fricatives (cf. *cerf, cire, cygne, centre, cinq*… ; *gel, génie, gilet, genre*…).

D'autres règles contextuelles régissent la prononciation du <s> et du <e>. Par exemple la consonne <s> se lit /s/ en début de mot et quand elle est précédée ou suivie par une consonne (*liste, bourse*). Entre deux voyelles ou semi-voyelles, elle se lit /z/ (*rose, bise*), sauf lorsque la voyelle qui précède est une nasale et se termine par une consonne graphique (*danse*). En fin de mot, le <s> est le plus souvent une marque morphologique qui s'écrit mais ne se prononce pas, tout comme le <e> qui, dans la même position, se prononce au mieux comme le <eu> du mot feu (*ce, de, je, que*…). Toutefois <e> se prononce /E/ dans une syllabe fermée phonologiquement (*sel, sec*) ou orthographiquement (*les, effet, user, nez*) ainsi que devant une double consonne, sauf cas exceptionnels. En particulier <e> se prononce /a/ dans le mot femme et les adverbes en <emment>. On peut citer en tant qu'autres anomalies le cas des mots *ennui* (versus *ennemi*) et dessus (versus *paresse*).

- Régularités statistiques non motivées phonologiquement

L'apprentissage d'autres spécificités du français nécessite la mémorisation de configurations orthographiques 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> ainsi que <nn> et <mm>[n], mais pas <ll>[o]), comparativement à celles qui ne sont géminées que rarement (<bb>, <dd>). Plusieurs études 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 [25].

- Lettres muettes pouvant ou non coder des marques morphologiques

Comme pour les données infra-lexicales et lexicales, les informations issues des lettres muettes de fin de mot, et de leur statut, permettent d'orienter les actions pédagogiques. On peut suggérer que, durant la première période de l'apprentissage de la lecture, les graphèmes muets soient écrits dans une couleur spécifique. Dès le milieu du CP, les séquences pédagogiques spécifiques doivent être mises en place sur le marquage de la morphologie grammaticale et lexicale à l'écrit. Il faudrait aussi travailler dès la maternelle sur les liaisons. De fait, par exemple lorsqu'une lettre muette est suivie 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[p]. Cette capacité est maitrisée par les enfants français qui, dès 3 ans, comprennent la différence entre *il arrive* et *ils arrivent* [26].

n Les doubles <n> et <m> proviennent d'une dénasalisation de la voyelle qui précède, un processus progressif qui a débuté au 13ème siècle. Ainsi, sauf dans certaines régions du sud-ouest de la France continentale, les mots *sonner* et *grammaire* ne se prononcent plus depuis fort longtemps comme *son* et *grand-mère*. Cela explique pourquoi certains mots récents (comme *sonore* et *sonate*) ne prennent qu'un seul <n>.

o Le double <l> a deux prononciations, /l/ comme dans *ville* ou /j/ comme dans *fille*, la seconde étant de loin la plus fréquente (cf. *bille, taille, abeille, feuille, brouillon*…).

p Pour une étude plus exhaustive sur la liaison en français, voir Encrevé P. (1988) : La Liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français. Editions du Seuil.

Évaluations : des observations de pratiques aux expérimentions

Que nous apprennent les résultats des études qui ont examiné les pratiques pédagogiques utilisées par les enseignants français au début de l'apprentissage de la lecture [1], [27] ou qui ont expérimenté de nouvelles pratiques utilisant la consistance et la fréquence des correspondances entre graphèmes et phonèmes [28]-[29]-[30] ?

Que nous apprennent les résultats de deux études observationnelles récentes ?

Deux études à plus ou moins grande échelle ont examiné les effets sur le devenir en lecture de pratiques pédagogiques centrées ou non sur le décodage en début de CP [27], [1]. La première, publiée en 2016 [27], a pris en compte 2507 élèves et 131 enseignants expérimentés de CP et examiné les progressions pédagogiques qu'ils utilisent pour enseigner les relations graphème-phonème en partant de celles de la base *Manulex* et en ajoutant celles habituellement étudiées au CP d'après les tables des matières d'une dizaine de manuels.

L'autre étude, publiée en 2024 [1], a porté sur 9340 enseignants de classes de CP (près de 20% de l'ensemble des enseignants de ce niveau) qui ont répondu à un questionnaire en ligne sur leurs pratiques d'enseignement durant les premiers mois de l'année scolaire. L'objectif de cette étude était également de vérifier, d'une part, quelles méthodes d'enseignement de la lecture sont mises en œuvre en début d'année au CP en France et, d'autre part, quels sont les effets de ces méthodes sur les résultats des élèves. Pour répondre à la première question, les 34 manuels de lecture utilisés par les enseignants de cette cohorte ont été classés en 4 catégories : plus ou moins centrés sur le décodage (c'est-à-dire sur les correspondances graphème-phonème) et plus ou moins mixtes. Et pour répondre à la seconde question, les résultats aux évaluations nationales CP-CE1 des 139.288 élèves de ces enseignants ont été examinés.

Selon les auteurs de la première étude (2016, [27]) l'examen des correspondances enseignées durant les 9 premières semaines de classe a permis de constater que l'élévation de ce nombre ''influence significativement et positivement les apprentissages jusqu'à 15 correspondances grapho-phonémiques enseignées pendant les neuf premières semaines de classe''(p.55, seconde colonne). Toutefois même si ces enseignants expérimentés étudient principalement des correspondances fréquentes, ce ne sont jamais toutes ''ni exclusivement celles qui font partie des 15 plus fréquentes en français'' (p.59, seconde colonne).

Ces auteurs soulignent également que ''certains professeurs prennent garde à ce que les textes proposés aux élèves soient suffisamment déchiffrables et ils semblent avoir raison'' (p.59, seconde colonne) comme l'indiquent leurs résultats. De même, en utilisant également la plateforme *Anagraph* pour évaluer la déchiffrabilité des items[q] , les auteurs de l'autre étude [1] ont observé qu'un enseignement avec une méthode phonémique stricte lors des premières semaines du CP est le plus efficace, aussi bien pour la fluence que pour la compréhension écrite. Les différences d'efficacité sont notables dès le mois de janvier de CP et restent présentes à l'entrée du CE1. En outre, l'impact des méthodes d'enseignement est plus fort chez les élèves fragiles, en particulier ceux scolarisés dans les écoles à recrutement social populaire.

Que nous apprennent les résultats des expérimentations avec des jeux éducatifs implémentés sur ordinateur ?

L'intérêt des méthodes implémentées sur ordinateur est qu'elles permettent de contrôler les progressions pédagogiques. C'est le cas de 3 études françaises récentes qui ont évalué l'impact d'une méthode d'apprentissage de la lecture utilisant une progression pédagogique fondée sur la consistance et la fréquence des CGPh, ainsi que sur la fréquence des mots. Deux de ces études ont été effectuées en CP [28]-[29] et une en grande section de maternelle [30].

Expérimentation en CP avec le logiciel *GraphoGame*

Cette étude [28] s'est étalée sur deux ans. Elle a utilisé une progression basée sur *Manulex-Morpho* et a impliqué 921 élèves de CP scolarisés dans des zones d'éducation prioritaire, une moitié ayant bénéficié la première année d'un entrainement au décodage en lecture (avec *GraphoGame*, cf. l'encadré 2), l'autre d'un entrainement aux maths (avec le logiciel *Fiete Math*), la seconde année, toujours avec des élèves de CP. La progression *GraphoGame* utilise les données de *Manulex-Morpho* (cf. pour l'ordre d'introduction des correspondances, voir l'annexe de l'article [28], pp. 467-468). Sont introduites en premier les correspondances graphème-phonème les plus consistantes et les plus fréquentes dans des items ayant une structure syllabique simple. Dans un second temps, sont abordées les correspondances dépendant du contexte puis les exceptions. En fin de parcours, plusieurs séquences portent sur la morphologie, principalement grammaticale : genre et nombre.

q Riou, 2017 ; http://anagraph.ens-lyon.fr

ENCADRÉ 2 : Quelques exemples d'entrainements avec le logiciel *GraphoGame*

Dans le premier jeu, il faut choisir, parmi 3 graphèmes présentés à l'écran, celui qui correspond au phonème entendu.

Dans le second et le troisième jeu, il faut choisir, parmi 3 syllabes (ou 3 mots) présentés à l'écran, la syllabe (ou le mot) qui correspond à la syllabe (ou au mot) entendu.

Dans le quatrième jeu, il faut placer les graphèmes présentés à l'écran dans l'ordre afin de former un mot. Un jeu similaire (non présenté) dans lequel il faut placer des mots présentés à l'écran dans l'ordre afin de former une phrase ayant un sens a aussi été développé.

Les pré-tests, effectués en novembre, ont été suivis par les entrainements qui ont duré environ 4 mois, de janvier à mai, puis par les post-tests réalisés en juin. Les effets de *GraphoGame* ont été évalués sur 4 tâches : conscience phonémique, choix orthographique et fluence (une en lecture de mots et une en lecture de pseudomots). Les effets des entrainements lecture-décodage sont faibles mais significatifs pour 3 des 4 post-tests : ceux évaluant la conscience phonémique, les capacités orthographiques et la fluence en lecture de mots (taille des effets : respectivement 0,23, 0,27 et 0,18). La relative faiblesse des effets obtenus pour l'ensemble des élèves peut provenir du caractère tardif des entrainements qui n'ont débuté qu'en janvier, et non dès le début du CP. En effet, les élèves devaient tous suivre la même progression qui commence par les correspondances graphème-phonème les plus consistantes et les plus fréquentes dans des items ayant une structure syllabique simple. Ce stade est, en janvier de CP, normalement dépassé par les élèves lecteurs typiques. En revanche ce n'est pas le cas pour les 50 % d'élèves les plus faibles au départ pour lesquels la taille des effets est plus que doublée.

Pour cette raison, une nouvelle version du jeu a été créée : *GraphoGame* Classroom. Elle propose une progression optimale sur toute l'année du CP, toujours en lien avec la consistance et la fréquence des correspondances entre graphèmes et phonèmes, en introduisant de 2-3 graphèmes par semaine. Elle dispose d'une interface permettant à l'enseignant d'assigner des exercices, de suivre les progrès des élèves et de créer des groupes de besoin. Il s'agit d'un véritable compagnon numérique pour les enseignants souhaitant mettre en œuvre une méthode phonémique rigoureuse. Cette version est actuellement en cours d'expérimentation en Nouvelle Calédonie et dans l'Académie Aix-Marseille.

Expérimentation en CP avec le logiciel ELAN

L'explication des effets modestes de *GraphoGame* sur l'ensemble de la cohorte par le caractère tardif des interventions est également étayée par les résultats d'une autre étude [29], avec le logiciel *ELAN*. Cette étude a pris en compte deux groupes témoins : un entrainé aux mathématiques avec le logiciel *Fiete* (comme dans l'étude [28]), et un sans entrainement spécifique (business as usual). Les entrainements ont eu lieu durant la même année scolaire, un groupe ayant eu en premier l'entrainement lecture-décodage, l'autre l'entrainement en mathématiques. Des post-tests ont été effectués après chacun des entrainements (les premiers entre le début et le milieu de février ; les seconds de mi-juin à fin juin).

Dans ce logiciel, l'ordre d'entrée des correspondances graphème-phonème pour les entrainements au décodage, déterminé par la consistance et la fréquence de ces correspondances dans *Manulex*, diffère toutefois légèrement de celui de *GraphoGame* : les voyelles les plus fréquentes sont présentées en premier, avant les consonnes fricatives qui, comme dans *GraphoGame*, sont introduites avant les occlusives[r].

Les résultats de 113 élèves du groupe lecture-maths et de 103 du groupe maths-lecture ont pu être examinés[s]. Par rapport au élèves entrainés aux maths, ceux qui ont bénéficié de l'entrainement décodage ont amélioré leurs capacités de fluence en lecture de mots et de pseudomots (les résultats sont présentés pour les deux tâches ensemble, sans les différencier) ainsi que leur compréhension. Toutefois l'impact de l'intervention lecture-décodage n'est significatif que pour les élèves qui ont bénéficié de cette intervention au cours du premier semestre. En outre, ces élèves ont continué à progresser entre le premier et le second post-test (avantage de 0,34 écart-type en fluence et de 0,31 en compréhension lors du second post-test). En revanche, l'utilisation d'*Elan* au cours de la seconde moitié de l'année n'a pas permis d'améliorer significativement les performances.

Pourquoi les élèves qui ont utilisé le logiciel *ELAN* au cours du premier semestre ont-ils progressé plus que ceux qui l'ont eu au cours du second semestre ? Une réponse possible vient du moment de passation. En effet, comme pour *GraphoGame*, cette version du jeu exigeait que l'élève termine chaque leçon avant de passer à la suivante. Ainsi, ceux qui ont utilisé le jeu au cours de la première partie de l'année ont bénéficié d'exercices plus appropriés à leur niveau, en synchronie avec l'enseignement prodigué en classe, alors que ceux qui l'ont eu dans la deuxième session ont reçu des exercices probablement trop faciles, qui n'étaient pour la plupart d'entre eux qu'une simple révision de ce qu'ils avaient déjà appris en classe depuis plusieurs mois.

Expérimentation en grande section de maternelle avec le logiciel Kalulu

Encouragés par les résultats de l'étude précédente, la même équipe a évalué les effets d'une version simplifiée du logiciel *ELAN* (rebaptisé *KALULU*) dans une population plus jeune : des élèves de grande section de maternelle [30]. La principale modification a été la suppression des tâches de compréhension en lecture de mots et de phrases. Quarante classes (1092 élèves) ont été affectées de façon aléatoire à un entrainement lecture-décodage puis maths pendant la première moitié de l'année, et l'inverse pour deux autres groupes pendant la seconde moitié de l'année. Le suivi longitudinal de cette cohorte a pris en compte les résultats des évaluations nationales de CP.

Tous les participants aux interventions lecture-décodage ont conservé un avantage dans les épreuves évaluant la conscience phonémique et la maitrise du principe alphabétique au début du CP, mais cet avantage ne s'est pas traduit par des gains en matière d'alphabétisation en milieu d'année (janvier de CP).

Un Bilan des évaluations présentées

Les résultats des études menées en CP indiquent l'existence d'une fenêtre temporelle très spécifique, et précoce, durant laquelle un enseignement basé sur les correspondances graphème-phonème les plus consistantes et les plus fréquentes est le plus avantageux, surtout pour les élèves fragiles. L'amélioration de la compréhension en lecture grâce à ce type d'enseignement précoce, intense et explicite permet de soutenir l'hypothèse selon laquelle l'automatisation des compétences de décodage libère des ressources cognitives pour la compréhension. Toutefois cette précocité ne semble pas fonctionner en deçà du CP, et ce à la différence de ce qui est relevé en anglais [31].

Pour conclure, et comme l'ont souligné les auteurs de l'étude [27] (p.59) : ''si le pouvoir *déchiffrer* offert aux élèves est retenu comme facteur de réussite en lecture, des recommandations peuvent être faites sur la nature des correspondances à étudier en priorité, en particulier pour favoriser les apprentissages des élèves les plus dépendants de l'intervention pédagogique. Peut-être serait-il bon d'alerter les enseignants sur ce point, surtout si l'on fait l'hypothèse qu'ils n'en sont pas tous conscients''.

Conclusion

Les résultats de ces études sont repris dans les objectifs d'apprentissage présentés dans les dernières instructions officielles dédiées au français en CP [4]-[5]. Ces documents, comme ceux plus anciens du Centre national d'Etude des Systèmes Scolaires (CNESCO) [32] ou de l'Observatoire National de la Lecture (ONL) [6] insistent sur l'importance de l'apprentissage du code alphabétique et que, pour cela, il faut s'accorder sur une progression et l'organiser sur l'année du CP.

Les résultats des études observationnelles et expérimentales présentées dans cet article, s'appuient sur des statistiques issues de bases de données qui rendent compte de la consistance et de la fréquence des correspondances pour les voyelles, les semi-voyelles et les consonnes. Ces résultats permettent de définir un (ou des) ordre(s) de présentation des correspondances susceptible(s) de faciliter le début de l'apprentissage de la lecture[t]. En conséquence, il serait bien d'offrir aux formateurs les moyens d'interroger facilement les données issues de ces différentes bases et de rendre ces statistiques accessibles à tous les enseignants en charge de cet apprentissage.

Enfin, les données chiffrées concernant par exemple les exceptions (tels que les deux <rr> de *nourrir, fourrure, courrier* alors qu'il n'y en a qu'un seul à courir) ou encore la part relative des pluriels en <s> par rapport à ceux en <x>, pourraient servir d'appui pour poursuivre la simplification de l'orthographe du français.

r Voir la présentation de l'ensemble de la progression dans l'annexe de l'article ([29], p. 208)

s Les pertes sont dues au fait que certains élèves étaient absents à une des trois sessions de test et que, pour d'autres élèves, les enregistrements des sessions d'entrainement n'étaient pas fiables.

t En complément, pour l'orthographe, voir la note de synthèse du CSEN (cf. référence [18])

Pour aller plus loin

- L'efficacité des méthodes d'enseignement de la lecture. Les apports de l'enquête Formalect. Note du CSEN n°11, Juin 2024.
- Rationnaliser l'orthographe du français pour mieux l'enseigner. Synthèse et recommandations du CSEN (2024).
- Pédagogie et manuels pour l'apprentissage de la lecture : comment choisir ? Synthèse de la recherche et recommandations du CSEN (2020)
- Apprendre à lire : du décodage à la compréhension. Synthèse de la recherche et recommandations du CSEN (2022)
- De nouveaux signaux d'alerte sur l'enseignement de la lecture en CP. Note d'alerte n° 1 du CSEN (2022)

Bibliographie

1. Deauvieau J. & Gioia P. (2024). L'efficacité des méthodes d'enseignement de la lecture. Les apports de l'enquête Formalect. Note 11 du Conseil Scientifique de l'éducation nationale (CSEN). https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conseil_scientifique_education_nationale/notes_csen/Note_CSEN_2024_11.pdf

2. Dehaene S. dir de (2020). Pédagogies et manuels pour l'apprentissage de la lecture : comment choisir ? https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conseil_scientifique_education_nationale/MANUELS_CSEN_VDEF.pdf

3. Sprenger-Charolles L. & Ziegler J. (2024). Apprendre à lire : du décodage à la compréhension. Synthèse de la recherche et recommandations, Conseil Scientifique de l'éducation nationale (CSEN). https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conseil_scientifique_education_nationale/CSEN_Synthese_Apprendre-a-lire.pdf

4. Ministère de l'Education Nationale de la jeunesse et des sports (2021). Pour enseigner la lecture et l'écriture au CP : un guide fondé sur l'état de la recherche. Seconde édition https://eduscol.education.fr/document/1508/download?attachment

5. Ministère de l'Education Nationale de l'enseignement supérieur et de la recherche (2024). Instructions Officielles : Programme de français pour le cycle 2 https://www.education.gouv.fr/sites/default/files/ensel135_annexe3.pdf

6. Morais J. & Robillart G. (1998). Apprendre à lire au cycle des apprentissages fondamentaux (GS, CP, CE1) Analyses, réflexions, propositions (Observatoire National de la lecture). CNDP et Editions Odile Jacob

7. Dehaene S. et al. (2011). Apprendre à lire : des sciences cognitives à la salle de classe. Paris, Odile Jacob.

8. Catach N. (1984). Les listes orthographiques de base (LOB). Nathan

9. Véronis J. (1988). From sound to spelling in French: simulation on a computer. *Cahiers de Psychologie Cognitive*, 8, 315-334.

10. 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, & Computers*, 28, 504-515.

11. Peereman R. & Content A. (1999). Lexop. A Lexical database with Orthography-Phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, 31, 376-379.

12. Gimenes M., Perret C. & New B. (2020). Lexique-Infra: Grapheme-phoneme, phoneme-grapheme regularity, consistency, and other sublexical statistics for 137,717 polysyllabic French words. *Behavior Research Methods*, 52, p. 2480-2488.

13. Stanké B. et al. (2019). ÉQOL : Une nouvelle base de données québécoise du lexique scolaire du primaire comportant une échelle d'acquisition de l'orthographe lexicale. *Corpus* [En ligne], 19 | 2019 ; https://doi.org/10.4000/corpus.3818

14. Peereman R., Lété B. & Sprenger-Charolles L. (2007) : Distributional characteristics of grapheme-phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods*, 39, 579-589. doi.org/10.3758/BF03193029. https:3inframorph.github.io/index.html

15. 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, *Année Psychologique*, 113, p. 3-33. doi.org/10.4074/S0003503313001012. https://inframorph.github.io/index.html

16. Lété B., Sprenger-Charolles L. & Colé P. (2004). Manulex: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, 36, 156-166.

17. Fayol M. (2020). L'acquisition de l'écrit. Paris : PUF (Que sais-je ?)

18. Sprenger-Charolles L., Abeillé A., Cerquiglini B. (2024). *Rationaliser l'orthographe du français pour mieux l'enseigner.* Synthèse de la recherche et recommandations, Conseil Scientifique de l'éducation nationale (CSEN). https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conseil_scientifique_education_nationale/syntheses_et_recommandations/17_rationnaliser_orthographe/CSEN_Synthese_17_orthographe_WEB__002_.pdf.

19. Abeillé A. (2024). L'accord et l'orthographe dans la Grande Grammaire du français, *Le Français moderne*, 92-2, 254-268.

20. Brunet E. (1981). Le vocabulaire français de 1789 à nos jours. Genève : Slatkine ; Paris : Champion, 1981. 3 tomes (Travaux de linguistique quantitative). ISBN 2-05-100361-0.

21. Deacon H., Desrochers A. & Levesque K. (2017), Learning to Read French. The crosslinguistic study of reading and spelling acquisition: The case of French, in L. Verhoeven & C. Perfetti (eds.), Learning to Read across Languages and Writing Systems, Cambridge, Cambridge University Press, 243-269.

22. Bouhali F., Bézagu Z., Dehaene S., & Cohen L. (2019). A mesial-to-lateral dissociation for orthographic processing in the visual cortex. *Proc Natl Acad Sci U S A. 2019 Oct 22;116(43):21936-21946. doi: 10.1073/pnas.1904184116.* PMID: 31591198; PMCID: PMC6815114.

23. Bouhali F., Dubois J., Hoeft F & Weiner K.S. (2024). Unique longitudinal contributions of sulcal interruptions to reading acquisition in children. *bioRxiv* [Preprint]. 2024 Jul 30:2024.07.30.605574. *doi: 10.1101/2024.07.30.605574.* PMID: 39131390; PMCID: PMC11312548.

24. Dehaene-Lambertz G., Monzalvo K. & Dehaene S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biol.* 16, e2004103 (2018).

25. Pacton S. & Commissaire E. (2024). Statistical learning and spelling : The case of graphotactic regularities. *Année Psychologique*, 124, 317-345

26. Nazzi T., Barrière I., Goyet L., Kresh S., Culbertson J. & Legendre G. (2011). Tracking irregular morpho-phonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition*, 120, 119-135.

27. Riou J. & Fontanieu V. (2016). Influence de la planification de l'étude du code alphabétique sur les performances des élèves en décodage au cours préparatoire. *Revue Française de Pédagogie*, 196,49-66, https://journals.openedition.org/rfp/5073

28. Lassault J., Sprenger-Charolles L., Albrand J. P., Alavoine E., Richardson U., Lyytinen H., & Ziegler J. C. (2022). Testing the Effects of GraphoGame Against a Computer-Assisted Math Intervention in Primary School. *Scientific Studies of Reading*, 26(6), 449–468.

29. Potier Watkins C., Caporal J., Merville C., Kouider S. & Dehaene S. (2020). Accelerating reading acquisition and boosting comprehension with a cognitive science-based tablet training. *J. Comput. Educ.* 7, 183–212 (2020).

30. Potier Watkins C. & Dehaene S. (2023). Can a Tablet Game That Boosts Kindergarten Phonics Advance 1st Grade Reading? *The Journal of Experimental Education*, 92(1), 32–55.

31. Shanahan C. & Lonigan J.L. (2010). The National Early Literacy Panel: A Summary of the Process and the Report. *Educational Researcher*, 39(4), 279-285 DOI:10.3102/0013189X10369172

32. CNESCO (2016). *Conférence de consensus. Lire, comprendre, apprendre. Comment soutenir le développement des compétences en lecture ?* Synthèse des recommandations du jury

Retrouvez l'intégralité des textes du CSEN sur le lien suivant :

reseau-canope.fr/conseil-scientifique-de-leducation-nationale

Ce qu'il faut retenir

- **Un enfant qui a une bonne compréhension orale ne peut comprendre ce qu'il lit que s'il a automatisé les mécanismes qui lui permettent d'identifier les mots écrits. Ce sont ces mécanismes qui sont spécifiques à la lecture, les processus de compréhension étant largement amodaux, c'est-à-dire similaires à l'écrit et à l'oral.**
  - → L'objectif principal de l'enseignement doit donc être de permettre à l'enfant de parvenir à comprendre ce qu'il lit de la même façon qu'il comprend ce qu'il entend.
  - → Pour cela, et tout au moins dans une écriture alphabétique comme celle du français, il faut l'aider à développer des capacités de décodage précises et rapides, ce qui nécessite un enseignement des correspondances graphème-phonème qui doit être systématique, intensif et structuré.
- **Le français écrit se caractérise, d'une part, par la présence de nombreux graphèmes ayant plus d'une lettre (<ch> et <ou> dans *chou* : /S/+/u/). Cela provient du fait que le nombre de phonèmes de cette langue dépasse celui des lettres de l'alphabet, surtout pour les voyelles (6 lettres dans l'alphabet, 10 à 16 voyelles en français).** Les progressions classiques partent des lettres de l'alphabet sans prendre en compte, dans les premières étapes de l'apprentissage, les phonèmes transcrits par un digraphe ce qui est une erreur. En effet, la présence de digraphes fréquents facilite la précision et la rapidité du décodage. En outre, certains digraphes sont plus faciles à prononcer que des phonèmes codés par une seule lettre, par exemple, <ou> comparé à <u>.
- **Le français écrit se caractérise également par la présence d'un nombre conséquent de graphèmes qui, en fin de mots, sont des marques morphologiques grammaticales** (le <s> du pluriel des noms et le <ent> de celui des verbes) **ou lexicales** (le <t> de *dent* support des dérivés *dentiste, dentition*…) **qui doivent s'écrire alors qu'elles sont muettes à l'oral**.
  - → Cette différence entre français écrit et oral doit être enseignée.
  - → Elle pose plus de difficultés pour l'écriture que pour la lecture.
- **Les relations entre phonèmes et graphèmes sont, en français, plus nombreuses que celles entre graphèmes et phonèmes.** Un exemple pour les voyelles : <é> se lit toujours /e/ (sauf le 2nd <é> d'*événement*, corrigé dans les rectifications de 1990) alors que le phonème /e/ peut s'écrire d'au moins 10 différentes manières : <é> dans *été*, <er> dans *voler*, <et> dans *et*, <ez> dans *nez*, <ai> dans *mai*, <ef> dans *clef* (qui s'écrit aussi *clé*). **Apprendre à écrire est donc plus difficile qu'apprendre à lire.**
- **Il existe une fenêtre temporelle précoce (les premières semaines du CP), durant laquelle un enseignement basé sur les correspondances graphème-phonème les plus consistantes et les plus fréquentes est le plus avantageux pour les élèves, surtout les plus fragiles.**

Un exemple de progression optimale pour les débuts de l'apprentissage de la lecture

→ **Voyelles orales pouvant être introduites en premier :**
<é>, <ou>, <a>, <o> et <eu>, sans différencier les deux <a>, <o>, et <eu> ;
<i>, sauf lorsque cette voyelle, suivie par une autre voyelle, est un yod (*ciel*) ;
<u>, sans faire la différence entre le <u> de *lu* (/y/) et celui de *lui* (/ɥ/) ;
<e>, seulement quand il est à la fin d'un monosyllabique (cf. *le, je, te*…).

→ **Voyelles nasales pouvant être introduites par la suite :**
<on> et <an>, mais pas <en> qui est un allographe de <an> (*pente*) et de <in> (*chien*) ;
<in> et <un>, ce dernier est fréquent dans le vocabulaire scolaire (*un, 1, lundi*).

→ **Consonnes pouvant être introduites successivement :**
D'abord les fricatives (<f>, <ch>, <j>, <v> plus <s> en début de mot) 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) ainsi que les consonnes dites liquides (<l> et <r>). Dans un second temps, il faudrait introduire les occlusives <p>, <b>, <d>, <qu> ainsi que <t> (sauf quand il se prononce /s/ comme dans *solution*…) et les nasales <m>, <n> et <gn>.

→ **Ces graphèmes doivent être présentés dans des mots respectant les spécificités des mots français les plus fréquents :** d'abord ceux composés d'une consonne suivie par une voyelle (CV : *le, la de, du, jeu, fou, chou*…) puis des CVC (*bal, peur, pour*…), des CCV (*tri, trou*…) et des CCVC (*fleur*). En outre, pour focaliser l'attention des enfants sur l'importance de l'ordre des graphèmes, des couples CV-VC (*tu-ut, fou-ouf*) et CVC-CCV (*tir-tri, dur-dru*) doivent aussi être utilisés, mais en tenant compte du fait que les monosyllabiques VC sont rares en français.

Un exemple de progression pour les étapes ultérieures de l'apprentissage de la lecture qui permet d'aborder les difficultés majeures de l'orthographe du français

→ **Correspondances dépendant de règles contextuelles**
Celles qui régissent la prononciation du <c> et du <g> qui est dite dure (c'est-à-dire celle d'une occlusive) ou douce (c'est-à-dire celle d'une fricative) en fonction du graphème qui suit (*grave* et *gare* vs *gel* et *givre* ; *crise* et *car* vs *ceci*).
D'autres régissent la prononciation du <s> qui se lit /s/ en début de mot et quand il est précédé ou suivi par une consonne (*liste, bourse*). Entre deux voyelles ou semi-voyelles, il se lit /z/ (*rose, bise*), sauf lorsque la voyelle qui précède est une nasale et se termine par une consonne graphique (*danse*). En fin de mot, <s> est le plus souvent une marque morphologique qui s'écrit mais ne se prononce pas.
C'est aussi le cas du <e> qui, en fin de mot, se prononce au mieux comme le <eu> de *feu* (cf. *ce, de, je, que*…). Toutefois <e> se prononce /E/ dans une syllabe fermée phonologiquement (*sel, sec*) ou orthographiquement (*les, effet, user, nez*) ainsi que devant une double consonne, sauf cas exceptionnels (*ennui, femme*…).

- **L'apprentissage d'autres spécificités du français est facilité par des régularités statistiques :** par exemple, les lettres géminées fréquentes et consistantes (<tt>), comparativement à celles qui ne sont géminées que rarement (<bb>, <dd>) ou qui ont plusieurs prononciation (*fille* vs *ville*). Les enfants sont sensibles à ces régularités statistiques qui, à la différence du double <s>, sont purement orthographiques.
- **Lettres muettes pouvant ou non coder des marques morphologiques**
  Au cours de la première période de l'apprentissage de la lecture, les graphèmes muets devraient être écrits dans une couleur spécifique. Par la suite (dès le milieu du CP) un travail systématique devrait être mis en place sur le marquage de la morphologie grammaticale et lexicale à l'écrit. Il faudrait aussi travailler dès la maternelle sur les liaisons. De fait, lorsqu'une lettre muette est suivie 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. Cette capacité est maitrisée par les enfants français qui, dès 3 ans, comprennent – à l'oral – la différence entre *il arrive* et *ils arrivent*.

Les données concernant la fréquence des correspondances graphème-phonème et phonème-graphème devraient aussi être utilisées à des fins de simplification de l'orthographe du français.

ANNEXE. CODAGE DES PHONÈMES : Alphabet phonétique international (API) et *Manulex* (codage pour données informatisées)

API Voyelles	Manulex	Exemples	API Consonnes	Manulex	Exemples
Voyelles Orales			**Occlusives**		
I	I	iris, vie, lys	p	p	pin, loupe, appel
U	U	ours, loup	t	t	terre, vite, attaque
Y	Y	une, lune	k	k	col, accord, qui, coq, kilo
E	E	égal, nez	b	b	brosse, cube
ɛ	E	lève, aile, jouet	d	d	danse, aide, addition
A ɑ	A A	adulte, table âne, âge	g	g	gare, guide, bague, toboggan
			Fricatives		
ɔ O	O O	sol, bol hôte, eau, jaune	f	f	foule, affaire, phare
			s	s	sol, ce, tasse, science, garçon
Ø Œ	2 9	deux, œufs neuf, œuf, sœur	ʃ	S	chat, vache
			v	v	vent, rêve
ə	% ° #	table, ceci gare, gobelet amie, lycée, gaiement	z	z	zéro, rose
			ʒ	Z	jeudi, gel, bourgeon
Voyelles Nasales			**Nasales**		
ɔ̃	§	on, ombre	m	m	main, femme
ɑ̃	@	an, en, septembre	n	n	nage, laine, panne
ɛ̃	5	cinq, bain, chien, plein	ɲ	N	ligne, peigne, ognon
œ̃	1	un, lundi, parfum	ŋ	G	parking, ring
Semi-Voyelles			**Autres**		
J	J	lieu, paye, paille, fille	l	l	lune, aller, pull
W	W	oie, fois, loin, western	ʀ	R	rue, air, arrière
ɥ	8	huit, lui, nuit			