

Proposition de hiérarchisation des 45 graphèmes de base de l'orthographe du français

Pérez, Manuel

ESPE Académie de Toulouse – Laboratoire CLLE-ERSS – UMR 5263 – Université de Toulouse 2 Le Mirail
manuel.perez@univ-tlse2.fr

Introduction

L'orthographe du français, considérée comme l'une des plus difficiles au monde (Fayol, 2003), appartient au domaine alphabétique dont la spécificité est d'utiliser des caractères pour noter les sons de la langue. Il ne s'agit pas d'un système simple mais plutôt d'un plurisystème régi par deux principes (phonographique et sémiographique) selon lesquels le graphème, unité de base et de description, représente respectivement soit du son soit du sens (Catach, Gruaz & Duprez, 1995 ; Cogis, 2005 ; Fayol & Jaffré, 2008). A ces deux principes correspondent trois fonctions (phonographique, morphographique et distinctive) qui permettent de classer les graphèmes en trois grandes catégories : les phonogrammes, les morphogrammes et les logogrammes. D'autres graphèmes, non fonctionnels, viennent compléter le système : il s'agit de graphèmes étymologiques et historiques dont la suppression n'engagerait pas les trois fonctions citées (Cogis, 2005).

La représentation du son est la fonction phonographique qu'assurent les phonogrammes : ces graphèmes symbolisent graphiquement des phonèmes (e.g. le graphème *in* qui sert à transcrire le phonème [ɛ̃] du mot [sɛ̃ʒ] → *singe*).

La représentation du sens relève du principe sémiographique et est dévolue aux morphogrammes et aux logogrammes. Les morphogrammes sont les graphèmes chargés de la fonction morphographique. Leur forte présence dans l'orthographe du français en fait une de ses principales spécificités (Jaffré, 2005). Ce sont les représentations graphiques des morphèmes. Ces plus petites unités de sens peuvent être distinguées en morphèmes grammaticaux (objet d'étude de la morphologie flexionnelle) et en morphèmes lexicaux (objet d'étude de la morphologie dérivationnelle). A l'image des morphèmes qu'ils représentent, les morphogrammes se subdivisent en morphogrammes grammaticaux et en morphogrammes lexicaux. Par exemple, les morphogrammes grammaticaux ont trait à toutes les flexions verbales qui permettent de différencier les personnes, les temps et les modes : [dɔ̃n] (vs [dɔ̃nɔ] ou [dɔ̃ne]) auquel correspondent trois formes graphiques différentes *donne*, *donnes*, ou *donnent*. En ce qui concerne les morphogrammes lexicaux, Catach *et al.* (1995) distinguent les marques lexicales finales et directes (e.g. le *t* final et muet de *compartiment* qui réapparaît dans *compartimenter*) ou indirectes (e.g. *-x/s* : *heureux/heureuse* ; *-t/tt* : *sot/sotte*), ainsi que les préfixes et les suffixes qui tous deux peuvent poser le problème du doublement des consonnes. Les morphogrammes, qu'ils soient grammaticaux ou lexicaux, sont de deux types : soit ils n'ont aucun lien avec le son et n'assurent que la représentation du sens (e.g. le *t* de *geignait*), auquel cas ce sont des morphogrammes au sens strict ; soit ils représentent tant du sens que du son, ce qui signifie que les fonctions morphographique et phonographique n'y sont pas entièrement détachées. Ce dernier type de morphogrammes est désigné par le mot-valise *morphonogramme*.

La fonction distinctive des graphèmes relève elle aussi du principe sémiographique dans lequel le sens est l'objet de la représentation. Elle est assurée par les logogrammes, unités dont la taille dépasse celle des graphèmes. Ce sont, en effet, « des “ graphies globales de lexèmes ”. Leur fonction est de donner une image spécifique à certains mots, afin d'aider à la reconnaissance rapide de leur sens » Catach *et al.* (1995, p. 262). Les logogrammes servent à différencier graphiquement des homophones hétérographes et hétérosèmes, c'est-à-dire des mots de forme phonique identique mais d'orthographe et de significations différentes (e.g. *cent*, *sans*, *sent*). En général, il s'agit de mots très courts, monosyllabiques, qu'il est préconisé d'enseigner globalement. Comme pour les morphogrammes, on distingue les logogrammes

grammaticaux ou mots-outils (e.g. *a/à, ces/ses*) des logogrammes lexicaux (e.g. *sain/saint/cinq/seing/sein/ceint*).

Les trois fonctions qui viennent d'être présentées sont, dans la réalité, liées les unes aux autres. C'est donc uniquement pour les besoins de l'analyse qu'elles ont été désolidarisées.

	Principes	Fonctions	Unités	Exemples
Zone fonctionnelle	Phonographique	Phonographique	Phonogrammes	<i>avril – poisson</i>
	Sémiographique	Morphographique	Morphonogrammes	Lexicaux : <i>français</i> Grammaticaux : <i>geignait</i>
			Morphogrammes	Lexicaux : <i>grand</i> Grammaticaux : <i>geignait – grands</i>
	Distinctive	Logogrammes	Lexicaux : <i>tante/tente</i> Grammaticaux : <i>sans/sent</i>	
Zone non fonctionnelle	Etymologique		Lettres étymologiques	<i>perdrix</i>
	Historique		Lettres historiques	<i>huile</i>

Tableau 1 : Principes et fonctions de l'orthographe du français (d'après Catach *et al.*, 1995 et Cogis, 2005).

La description de l'orthographe du français (Cf. Tableau 1) serait incomplète si n'étaient évoqués les caractères qui n'ont de fonction ni phonographique, ni morphographique, ni distinctive. Il s'agit des graphèmes étymologiques ou historiques qui ne jouent plus vraiment de rôle dans le système orthographique actuel et dont Catach *et al.* (1995) donnent quelques exemples : appartiennent à la première catégorie les lettres latines (e.g. l'*h* de *homme* ; des consonnes doubles comme dans *appeler* ; des graphies plus ou moins logogrammiques comme *doigt, vingt*), les graphèmes d'origine grecque formés d'une lettre prononcée et d'un *h* muet (e.g. *rhume, théâtre*), les lettres d'autres origines (e.g. l'*ow* de *clown*) ; les lettres historiques sont celles que l'étymologie ne peut justifier, utiles à un temps précis de l'histoire de l'orthographe mais ayant perdu leur valeur informative (e.g. le *h* initial de *huit, huis, huile*, qui n'a rien de latin mais qui fut introduit pour signifier que le *u* initial était bien une voyelle et non pas une consonne : [ɥil] vs [vil]). Comme le souligne Cogis (2005, p. 51), les lettres étymologiques et historiques appartiennent à « une zone floue, instable, voire anarchique. On n'en finirait pas d'énumérer les exceptions, les anomalies, concernant des graphies incorporées pour des raisons aujourd'hui sans intérêt, et qui ne subsistent que comme traces des choix graphiques, plus ou moins heureux, ou comme erreurs et incohérences malheureuses d'une époque antérieure ».

1 Des systèmes orthographiques plus ou moins profonds

La complexité de l'orthographe doit être considérée selon deux angles différents, que l'on se situe dans le domaine de la lecture en réception, ou bien dans celui de l'écriture en production. Dans les deux cas, toutefois, sont concernées les relations que les phonèmes entretiennent avec les graphèmes et *vice versa*. En lecture, on parle de correspondances orthophonologiques (ou correspondances graphie-phonie : C.G.P.), alors que l'écriture renvoie à des correspondances phono-orthographiques (ou correspondances phonographémiques, ou encore correspondances phonie-graphie : C.P.G.) (Bonin, Collay & Fayol, 2008).

Une orthographe alphabétique idéale serait celle qui ne contiendrait que des relations biunivoques quel que soit le sens considéré, C.G.P. ou C.P.G. En d'autres termes, il s'agirait d'un système dont la fonction serait phonographique et dont la sémiographie serait absente. Cet idéal ne semble pas avoir été atteint par

quelque système que ce soit, même si certains s'en approchent comme le finnois, le turc, le serbo-croate ou le tchèque pour les langues alphabétiques (Caravolas, 2004), ou encore l'italien et l'espagnol pour ce qui est des langues romanes. Ainsi, le système espagnol compte 29 graphèmes pour 25 phonèmes (Fayol & Jaffré, 2008) tandis que le français d'aujourd'hui, selon Catach *et al.* (1995) totalise environ 130 graphèmes (175 selon d'autres auteurs, ainsi que le soulignent Fayol et Jaffré, 2008) pour les 36 phonèmes que compte la langue, et que l'anglais, pour sa part, offre 561 graphèmes pour l'écriture de ses 41 phonèmes (Fayol & Jaffré., 2008).

En conséquence, le binarisme qui consisterait à classer les orthographes en systèmes exclusivement phonographiques ou sémiographiques n'est pas de mise. Les auteurs (*e.g.* Borgwaldt, Hellwig & De Groot, 2005 ; Ferrand, 2007 ; Jaffré, 1992, 2005) préfèrent parler de continuum. Il y aurait ainsi des orthographes plus ou moins transparentes ou régulières et des orthographes plus ou moins opaques ou irrégulières, ce qui renvoie à la notion de profondeur orthographique en lecture (Katz & Frost, 1992).

En ce qui concerne les langues européennes et pour rester dans le domaine de la lecture, 13 d'entre elles ont été étudiées par Seymour, Aro et Erskine (2003). Ces auteurs en ont proposé une classification en fonction de leur profondeur orthographique (*Cf.* Tableau 2). Ils ont également pris en considération la structure syllabique des langues concernées. Sont distinguées d'une part certaines langues romanes (italien, espagnol, portugais, français), le grec et le finnois caractérisés par une prédominance de syllabes ouvertes de structure consonne-voyelle avec peu de groupements de consonnes à l'initiale ou en finale (structure syllabique simple), et d'autre part des langues germaniques (allemand, norvégien, islandais, néerlandais, suédois, danois, anglais) qui présentent une structure syllabique complexe (nombreuses syllabes fermées de structure consonne-voyelle-consonne, accompagnées tant à l'initiale qu'en finale de groupements de consonnes).

Dans le continuum de la profondeur orthographique, le français, langue de structure syllabique simple, apparaît donc comme possédant une orthographe plutôt opaque, conséquence des trois différentes fonctions qu'assurent les graphèmes servant à l'écriture.

		Profondeur orthographique				
		Transparence			Opacité	
Structure syllabique	Simple	Finnois	Grec Italien Espagnol	Portugais	Français	
	Complexe		Allemand Norvégien Islandais	Néerlandais Suédois	Danois	Anglais

Tableau 2 : Classification de 13 langues européennes en fonction de leur structure syllabique (simple, complexe) et de leur profondeur orthographique (transparence, opacité) (Seymour *et al.*, 2003, p. 146).

2 La complexité de l'orthographe française

Le plurisystème orthographique du français repose donc sur de solides fondations phonographiques mais est également constitué d'un nombre important de particularités sémiographiques qui en font toute la complexité, notamment lorsqu'il s'agit de production. A titre d'illustration, Catach (2011) cite les statistiques obtenues à partir de l'*Echelle Dubois-Buyse* (Ters, Mayer & Reichenbach, 1982) : sur les 3724 mots composant cette base de données, 3,27% sont des logogrammes et, parmi les graphèmes représentés, 83,50% sont des phonogrammes, 3,67% des morphogrammes et 12,83% des lettres muettes. Par ailleurs, la linguistique statistique nous apprend que 96% des C.G.P. sont régulières, ce qui n'est le cas que pour 71% des C.P.G. (Fayol & Jaffré, 2008). C'est ainsi par exemple que les graphèmes *o*, *au* ou *eau* d'une part ou les graphèmes *f* ou *ph* d'autre part, ne correspondent qu'à un seul et même phonème, respectivement [o] et [f], alors que plusieurs possibilités existent pour écrire chacun de ces phonèmes

(Sprenger-Charolles, Siegel & Bonnet, 1998). Par conséquent, notre orthographe, complexe dans le sens de la lecture, l'est davantage dans celui de l'écriture.

Afin d'en organiser l'enseignement/apprentissage, Catach *et al.* (1995) ont proposé une division de son système graphémique dont ils ont analysé linguistiquement la structure et le fonctionnement. Les auteurs se sont appuyés sur les mesures statistiques de deux variables sous-lexicales : la fréquence des C.P.G. et celle des graphèmes. Ces deux variables ne prennent pas en compte l'utilisation effective du système par les scripteurs. Toutefois, la délimitation de trois niveaux de complexité dénote une prise en considération des utilisateurs en situation d'apprentissage. Chacun des trois niveaux est représenté dans un tableau différent.

Le système des graphèmes dans son intégralité contenant environ 130 graphèmes (Niveau 3) est reproduit en annexe (Cf. Tableau A1). Pour les auteurs, un archigraphème est un « *graphème fondamental, représentant un ensemble de graphèmes, qui sont par rapport aux autres ensembles dans un rapport exclusif, correspondant au même phonème ou au même archiphonème. Ex. : O pour désigner o, ô, eau, etc. Cet ensemble peut ne contenir qu'un élément (ex. GN représentant gn). S'ajoutent en français trois "idiotismes graphiques" : X, OI, OIN, correspondant chacun à deux phonèmes. L'archigraphème est noté, comme l'archiphonème, par la majuscule (O)* » (p. 17). De l'archiphonème, il est dit qu'il est un « *représentant de l'ensemble des traits phoniques pertinents communs à deux ou plusieurs phonèmes, qui sont par rapport aux autres dans un rapport exclusif. Ex. : maison, [e] ou [ɛ], noté [E]* » (p. 16). Sont notés en gras les 45 graphèmes considérés comme les graphèmes de base que les élèves en fin de C.E.1 doivent aisément manier (Niveau 1). Les parenthèses qui entourent certains graphèmes indiquent que ces derniers sont répétés, c'est-à-dire qu'ils correspondent à plusieurs phonèmes et à plusieurs secteurs d'archigraphèmes. Enfin, « *les graphèmes exceptionnels (moins de 5 à 6 unités lexicales et leurs dérivés), d'ailleurs non exhaustifs* » (p. 14), sont signalés d'un astérisque. Quel que soit le phonème considéré, il existe plusieurs possibilités de transcription graphémique. Ainsi, même la possibilité de transcription de la voyelle [y] par le graphème *u*, approximativement considérée par les auteurs comme la plus proche des 100%, peut être écrite avec les 4 autres graphèmes que sont *û, hu, ü, ou eu*.

Le niveau 2 du système, représenté en annexe (Cf. Tableau A2) correspond à un niveau intermédiaire constitué de 70 graphèmes dont le maniement devrait être aisé à la fin de l'école élémentaire (fin de cycle des approfondissements ou Cycle 3). Les graphèmes de base y sont également notés en caractères gras.

Enfin, le niveau 1 (Cf. Tableau 3, ci-après) regroupe les 45 graphèmes de base que les auteurs recommandent comme devant être maîtrisés à la fin du cycle des apprentissages fondamentaux (ou Cycle 2).

3 Les graphèmes de base du français : une description non hiérarchisée

Les 45 graphèmes de base du niveau 1 de la classification proposée par Catach *et al.* (1995) peuvent être considérés comme « *le code minimal de transcription du français* » en-deçà duquel il n'est pas possible d'aller. Le tableau qui les classe selon les phonèmes qu'ils permettent de transcrire. S'y côtoient des graphèmes n'ayant aucun concurrent dans cette transcription (e.g. *d* → [d]) et des graphèmes entretenant avec leurs phonèmes correspondants une relation qui n'est ni unique ni systématique (e.g. *o, au, eau* → [o]). Une telle représentation laisse ainsi émerger certaines zones potentielles de difficulté. Toutefois, elles semblent conférer aux diverses unités graphémiques un statut d'égalité, ce qui ne permet pas d'en dégager toutes les spécificités. Or, si les objectifs demeurent l'apprentissage et le maniement aisé de l'orthographe des mots, il est nécessaire d'identifier ces spécificités qui permettront de hiérarchiser les graphèmes en fonction de leur degré de complexité. Peut-être alors qu'une classification en fonction des phonèmes n'est plus pertinente et que d'autres critères doivent être établis. En ce sens, une collaboration entre linguistique et psycholinguistique peut s'avérer fructueuse. En effet, les chercheurs dans ce dernier domaine disciplinaire sont parvenus à isoler différentes variables sous-lexicales susceptibles d'influer sur la production orthographique

Archigraphèmes	Phonèmes	Graphèmes de base	% approximatif d'utilisation	Exemples
A	[a]	a	92%	papa
E	[e]	e + é	99%	mes, pré
	[ɛ]	(e) + è ai	67,9% 30%	bec, règle chair
I	[i]	i	99%	il
O	[o]	o	75%	zéro, sol
		au	21%	cheveu(x)
		eau	3%	oiseau
U	[y]	u	100%	tu
EU	[œ]	eu	93%	peu, peur
	[ə]	(e)		ch(e)val
OU	[u]	ou	98%	fou
AN	[ɑ̃]	an	44%	(un) an
		en	47%	enlever
IN	[ɛ̃]	in	45%	fin
		(en)	23%	chien
ON	[ɔ̃]	on	92,8%	son
UN	[œ̃]	un	97%	un
Y, IL(L)	[j]	(i)	86%	piéd
		ï		faïence
		y	30%	payer
		il(l)	10%	rail, raille
OI	[wA]	oi	100%	loi
OIN	[wɛ̃]	oin	100%	moins
P	[p]	p	100%	pon(t)
B	[b]	b	100%	bon
T	[t]	t	99%	ton
D	[d]	d	100%	don
C	[k]	c + qu	98%	coque
G	[g]	g + gu	100%	goguette
F	[f]	f	95%	fou
V	[v]	v	100%	vou(s)
S	[s]	s + ss	69%	danse, rosse
		(c) + ç	26%	ci, ça
Z	[z]	(s) intervocalique	90%	rose
		z	10%	zéro
X	[k s] ou [g z]	x	84%	axe, examen
CH	[ʃ]	ch	100%	chou
J	[ʒ]	j	49%	jeu
		g + (ge)	51%	(nous) mangeons
L	[l]	l	100%	la
R	[ʀ]	r	100%	ré
M	[m]	m	100%	mon
N	[n]	n	100%	non
GN	[ɲ]	gn	100%	règne
N.B. : Les consonnes doubles et les lettres muettes ne sont pas décomptées ici.				

Tableau 3 : Tableau des graphèmes de base du français - Niveau 1 (Catach *et al.*, 1995, pp. 10-11).

4 Variables sous-lexicales influant sur la production orthographique

Les 45 graphèmes de base du français tels qu'isolés par Catach *et al.* (1995) ne relèvent pas tous du même type. Le tableau les contenant les présente uniquement selon les phonèmes qu'ils permettent de transcrire. Bien qu'il laisse apparaître des zones potentiellement plus problématiques que d'autres, il ne semble tenir compte ni de leurs différents degrés de complexité orthographique, ni des difficultés empiriques que les scripteurs pourraient rencontrer dans leur acquisition ou leur maniement. Parce qu'elle rend possible l'établissement d'une progression de l'enseignement/apprentissage de l'orthographe des mots, la prise en considération de ces différences nous paraît pourtant essentielle. Elle s'appuie sur l'identification de critères classificateurs qui correspondent à des variables sous-lexicales susceptibles d'avoir un impact sur la production orthographique.

Dans leur étude longitudinale conduite auprès d'enfants de 2^o et 4^o années d'élémentaire, Mousty et Leybaert (1999) ont proposé une catégorisation de graphies particulières. La préférence accordée au terme *graphie* au détriment du terme *graphème* tient à la présence dans une des catégories de groupes consonantiques tels *gr* ou *pr*. Quatre types de graphies sont distingués, que l'enfant doit produire dans des mots destinés à compléter des phrases lacunaires : les graphies consistantes acontextuelles, les graphies consistantes contextuelles, les graphies inconsistantes contextuelles et les graphies dérivables par la morphologie. Les résultats de l'étude indiquent que les graphies consistantes acontextuelles sont les premières à être maîtrisées, d'abord les simples puis les complexes. Les critères définis par cette étude sont donc le contexte d'utilisation, la systématique des C.P.G. et la dérivation morphologique. Ce dernier critère ne concerne pas des phonogrammes mais des morphogrammes muets : il ne peut donc être retenu pour une hiérarchisation des 45 graphèmes du niveau 1 en fonction de leur difficulté empirique. En revanche, nous conserverons les deux suivants : au même titre que la complexité graphémique, le contexte d'utilisation et la systématique des C.P.G. sont des facteurs liés à des unités sous-lexicales reconnus pour leur capacité à exercer une influence sur la production orthographique.

4.1 La complexité graphémique

La complexité graphémique concerne le nombre de lettres qui composent un graphème, (de 1 à 3 pour les phonogrammes du français). Cette variable a été identifiée comme exerçant un effet sur la production orthographique pour laquelle elle est une source de difficulté, quel que soit le degré d'expertise des sujets. Au cours d'une étude longitudinale conduite auprès d'enfants de Cours Préparatoire, Sprenger-Charolles *et al.* (1998) ont analysé l'apprentissage de la lecture et de l'orthographe en français. Les élèves étaient soumis en janvier et en juin à des tâches de lecture et d'orthographe (*i.e.* dictée) de mots et de pseudomots. Plusieurs facteurs étaient manipulés, parmi lesquels la complexité graphémique. Celle-ci concernait des mots complexes qui contenaient les graphèmes complexes *ou* ou *ch* (*e.g.* *tour*, *poche*) et des mots simples dans lesquels les graphèmes étaient tous des lettres uniques (*e.g.* *abri*). Son influence sur la production orthographique était examinée à partir des productions erronées. Les résultats montrent que si, en janvier, les mots complexes étaient moins bien orthographiés que les mots simples, la différence disparaissait en juin. Il semblerait donc qu'au-delà des tout premiers apprentissages, la complexité graphémique ne constitue plus une source de difficulté si l'on s'en tient aux erreurs commises. Pourtant, l'analyse de variables temporelles révèle que les graphèmes complexes sont traités plus lentement que les graphèmes simples, y compris par les experts, lorsque la production orthographique est automatisée. C'est ce qu'indiquent les résultats de l'étude conduite auprès d'adultes par Kandel et Spinelli (2010). Les participants devaient copier des séries de trois mots contenant soit des graphèmes simples (*A* ou *E*), soit des graphèmes complexes de deux ou trois lettres (*AI* / *AIN* ou *EI* / *EIN*) partageant une même lettre à la même position (*e.g.* le graphème simple *A* dans *CLAVIER*, le digramme *AI* dans *PRAIRIE* et le trigramme *AIN* dans *PLAINTE*, la lettre *A* se situant en troisième position dans chacun des trois mots). Étaient mesurées la durée de production de la lettre précédant le graphème cible, celle de sa première lettre (*A* ou *E*) et pour les digrammes et les trigrammes, celle du *I* suivant le *A* ou le *E*. L'analyse des durées d'écriture des lettres indique que le temps de traitement du mouvement dépend de la longueur des graphèmes, et donc de leur complexité. Ainsi, la durée d'un *A* graphème simple est plus courte que celle du *A* dans le digramme *AI*, cette dernière étant à son tour plus courte que celle du *A* dans le trigramme

AIN. Le traitement des graphèmes simples a lieu en parallèle avec leur exécution motrice (*i.e.* en même temps que leur écriture effective), ce qui se traduit par des durées d'écriture plus longues que celles affectant les lettres qui les précèdent. De leur côté, les graphèmes complexes et leur degré de complexité seraient traités avant que leur écriture ne commence, sur la lettre qui les précède : d'une part, plus les graphèmes cibles sont longs et plus les durées des lettres qui les précèdent augmentent et d'autre part, la différence de durée d'écriture des *I* dans ces graphèmes complexes n'est pas significative entre les digrammes et les trigrammes. Du point de vue du codage en mémoire à long terme des représentations orthographiques, les résultats de cette étude vont à l'encontre du modèle de Van Galen (2001) pour lequel les informations qu'elles contiennent ne relèvent que de l'identité et de l'ordre des lettres. Aux informations fournies par la structure syllabique et la structure morphologique des mots (Kandel, Álvarez & Vallée, 2006 ; 2008), viennent ainsi s'ajouter celles concernant la structure des graphèmes et leur complexité. Les résultats de l'étude toute récente conduite par Shen, Damian et Stadthagen-Gonzalez (2013) sur l'anglais à partir d'un paradigme d'amorçage implicite vont dans le même sens.

4.2 Le contexte d'utilisation

Le contexte d'utilisation d'un graphème a également sa part d'importance dans la réalisation orthographique. En effet, les graphies qui conservent la systématisme des C.P.G. répondent à des règles sur lesquelles le contexte peut influencer. Dans leur étude, Mousty et Leybaert (1999) distinguent les règles consistantes acontextuelles (*e.g.* le graphème *a* permet de représenter graphiquement le phonème [a]) que les enfants parviennent à maîtriser rapidement (89,9% de réponses correctes en 2^o année d'élémentaire et jusqu'à 96,4% en 4^o année) des règles consistantes contextuelles (*e.g.* le phonème [ɔ̃] s'écrit avec la lettre *m* devant *p* et *b*) dont la maîtrise est loin d'être généralisée en 4^o année (seuls 10% des enfants y parviennent). Une analyse comparative des productions de graphies consistantes contextuelles montre que le nombre de réponses correctes est plus important lorsqu'elles sont contenues dans des mots fréquents que quand elles appartiennent à des mots rares. Pour les auteurs, ces résultats indiquent que l'application de règles explicitement apprises n'est pas suffisante. Leur maîtrise est également soumise à la mémorisation par les enfants de représentations orthographiques contenant ces graphies contextuelles, mémorisation qui est le résultat de ce que les auteurs nomment « *l'expérience linguistique* » (p. 9). En effet, il semblerait, ainsi que suggéré par plusieurs études, que la prise en compte du contexte de C.P.G. non systématiques se développe au fur et à mesure que s'accroît l'expérience avec l'écrit (pour une synthèse, voir Lété, 2008). En d'autres termes, plus l'enfant lit et écrit des mots contenant de telles graphies, plus ces dernières sont mémorisées, ce qui fait de l'apprentissage implicite rendu possible par la fréquence de rencontre des mots, un facteur déterminant dans leur maîtrise.

La catégorie des graphies contextuelles qui conservent la systématisme des C.P.G. ne peut être conservée pour une classification des 45 graphèmes du niveau 1 dans la mesure où aucun de ces derniers, qu'il soit simple ou complexe, ne semble pouvoir y être affecté si l'on considère que leur maîtrise doit résulter d'un enseignement explicite. En revanche, nous n'écarterons pas le facteur constitué par le contexte d'utilisation puisqu'il peut être un facteur discriminant dans la hiérarchisation des graphèmes dont la relation phonographémique n'est plus systématique.

4.3 La consistance phonographémique

Ainsi que le laisse apparaître le Tableau 3, tous les phonogrammes de notre orthographe n'entretiennent pas avec les unités sonores qu'ils représentent des relations systématiques, que le contexte d'utilisation entre ou non en ligne de compte. Mousty et Leybaert (1999) utilisent à leur endroit la notion d'inconsistance qui implique l'absence de règle unique et systématique dans les C.P.G. et, en conséquence, l'impossibilité d'un enseignement explicite visant leur maîtrise. La notion de consistance/inconsistance orthographique a, dans la littérature psycholinguistique, des significations fluctuantes et ambiguës, à l'image des notions de transparence/opacité et de régularité/irrégularité avec lesquelles il lui arrive de se confondre (Kreiner, 1996). Les raisons de cette confusion sont à chercher dans le fait que les trois notions partagent une seule et même origine, à savoir les relations entre les unités

sonores et les unités graphiques de la langue. L'historique de leur évolution tel que présenté dans leurs synthèses respectives par Ferrand (2007) pour la lecture et Bonin *et al.* (2008) pour la production écrite permet néanmoins d'en dégager les spécificités. Ainsi que nous l'avons vu précédemment, la notion de transparence/opacité s'applique aux systèmes orthographiques qu'elle sert à différencier. Il ne s'agit pas d'une opposition binaire mais plutôt d'un continuum sur lequel le curseur peut se déplacer en fonction de la stabilité des relations entretenues entre unités sonores et unités orthographiques de la langue. Comme le rappellent Bonin *et al.* (2008), la distinction entre les notions de régularité et de consistance est pertinente dans le domaine de la lecture à haute voix. Ferrand (2007) cite les travaux de Venezky (1970) pour qui la régularité est étroitement liée à la fréquence des C.G.P. : dans un mot régulier, les C.G.P. sont les plus fréquentes (*e.g. rame* → [R a m]) alors qu'un mot irrégulier en contient au moins une de rare (*e.g. femme* → [f a m] au lieu de [f ε m]). Il convoque par la suite les critiques de Parkin (1982) à l'encontre de cette définition, critiques selon lesquelles la régularité dépend non seulement de la fréquence des C.G.P. mais aussi de la fréquence d'usage des mots qui les contiennent. En ce qui concerne la consistance, l'auteur rappelle que cette notion a été proposée par Glushko (1979) qui la définit à partir des relations entretenues uniquement par une rime orthographique et sa prononciation : deux mots qui partagent la même rime orthographique sont dits inconsistants si leur prononciation diffère (*e.g. jars, mars et gars*). Les deux notions sont donc de type binaire, la consistance étant exclusive de l'inconsistance, comme la régularité l'est de l'irrégularité. Plus tard, l'évaluation de la consistance dépasse le cadre de la rime pour s'appliquer soit à des unités lexicales, les mots, soit aux unités sous-lexicales qui composent un mot. Dans le premier cas, il s'agit : des homophones hétérographes hétérosèmes (*e.g. chœur / cœur ; paire / père ; temps / taon*, Ferrand, 1999), d'homophones homographes hétérosèmes (*e.g. louer un appartement / louer un dieu*, Ferrand, 1999), et d'hétérophones homographes hétérosèmes (*e.g. un fils / [fis] – des fils / [fil]*, Ferrand, 1999). Dans le second cas, lorsque la consistance concerne toute unité sous-lexicale quelle qu'elle soit (*e.g. l'attaque, la voyelle, la rime*), elle est définie « *comme la proportion des mots qui possèdent des lettres communes ayant la même prononciation* » (Bonin *et al.*, 2008, p. 521). Elle cesse alors d'être une mesure binaire pour devenir une variable continue fondée sur les relations statistiques entre unités orthographiques et phonologiques. Dans ce contexte, le degré de consistance d'une C.G.P. se calcule généralement par le rapport entre le nombre de mots contenant une unité orthographique de prononciation spécifique et le nombre total de mots contenant cette même unité orthographique sans distinction de prononciation (Peereman & Content, 1999). Il s'agit ici de la consistance orthographique prise dans le sens de la lecture, c'est-à-dire de celle qui a pour point de départ le code orthographique et pour point d'arrivée le code phonologique, plus précisément appelée consistance ortho-phonologique. La notion symétrique de consistance phono-orthographique (ou phonographémique) concerne pour sa part la production orthographique. Dans ce domaine et en ce qui concerne le français, la distinction entre les variables de consistance et de régularité n'est plus opérationnelle. Bonin, Collay, Fayol et Méot (2005) ont en effet établi qu'une forte corrélation existe entre elles, un mot consistant dans ses C.P.G. étant régulier, et un mot inconsistant irrégulier. C'est dans cette corrélation que pourrait se trouver l'explication de l'utilisation indifférenciée de l'un ou l'autre des deux termes dans la littérature concernant ce domaine. Néanmoins, d'autres chercheurs essaient de les éviter, à l'exemple de Kreiner (1996) qui leur préfère le terme de *polygraphie*, sujet à moins d'ambiguïté. Il définit la polygraphie comme le degré de transparence contenue dans les C.P.G., ce qui en fait une mesure statistique et donc objective, à l'instar de la consistance telle qu'elle a pu être présentée dans diverses études ou bases de données consacrées au français (*e.g. Lété, 2008 ; Lété, Sprenger-Charolles & Colé, 2004 ; Peereman & Content, 1998, 1999 ; Peereman, Lété & Sprenger-Charolles, 2007 ; Ziegler, Jacobs & Stone, 1996*). Dans le cadre de cet article, malgré son ambiguïté soulignée par Kreiner (1996) et en raison de son utilisation majoritaire dans la littérature (Bonin *et al.*, 2008), nous utilisons le terme *consistance* au détriment de *polygraphie*. En règle générale, les effets de la consistance phonographémique ont été plutôt délaissés par la recherche au bénéfice de ceux exercés par la consistance ortho-phonologique en lecture et en reconnaissance de mots (Bonin *et al.*, 2008 ; Lété, 2008). Pour le français, plusieurs travaux fondés sur des tâches différentes (*i.e. dénomination écrite d'images, dictée, copie*) ont montré l'influence de la consistance dans la production orthographique tant chez des adultes que chez des enfants, influence relevée dans l'examen des productions en tant que telles (*i.e. réponses correctes vs réponses erronées*), mais aussi dans l'analyse de variables aussi diverses que les durées de production, les latences d'initialisation ou les fixations oculaires

(e.g. Alegria & Mousty, 1996 ; Bonin, Peereman & Fayol, 2001 ; Delattre, Bonin & Barry, 2006 ; Kandel, Valdois & Orliaguet, 2003 ; Kandel & Valdois, 2005 ; Lambert, Alamargot, Larocque & Caporossi, 2011 ; Lété, Peereman & Fayol, 2008).

Dès lors, la consistance phonographémique, une des contraintes les plus importantes dans l'apprentissage de l'orthographe des mots, peut être retenue comme l'un des facteurs linguistiques qui nous permettront de hiérarchiser les 45 graphèmes de base.

5 Les graphèmes de base du français : proposition de hiérarchisation

Afin de concevoir leur test *ORTHO3* d'évaluation de l'orthographe, Mousty et Leybaert (1999) avaient catégorisé un choix de graphies au moyen de variables telles que la complexité graphémique, le contexte d'utilisation et la consistance des C.P.G. Leur application aux graphèmes du niveau 1 permet de distinguer des graphèmes simples et complexes, acontextuels et contextuels, consistants et inconsistants. Il s'agit donc soit de graphèmes composés d'une seule lettre accompagnée ou non d'un signe diacritique, soit de graphèmes composés de plusieurs lettres, soit de graphèmes servant à transcrire un seul et même phonème quel que soit leur contexte environnant, soit de graphèmes dont le phonème transcrit change en fonction du contexte, soit de graphèmes inclus dans une C.P.G. consistante en étant l'unique possibilité de transcription d'un phonème, soit de graphèmes relevant d'une C.P.G. inconsistante qui les place en concurrence avec d'autres graphèmes pouvant eux aussi transcrire le même phonème.

Le croisement de ces caractéristiques nous permet d'aboutir aux catégories suivantes, classées par ordre croissant de complexité orthographique empirique :

- 1 - graphèmes simples, acontextuels, consistants,
- 2 - graphèmes complexes, acontextuels, consistants,
- 3 - graphèmes simples, contextuels, consistants,
- 4 - graphèmes complexes, contextuels, consistants,
- 5 - graphèmes simples, acontextuels, inconsistants,
- 6 - graphèmes complexes, acontextuels, inconsistants,
- 7 - graphèmes simples, contextuels, inconsistants,
- 8 - graphèmes complexes, contextuels, inconsistants.

Aucun des graphèmes du niveau 1 n'est concerné par les deux catégories 3 et 4 de graphèmes contextuels consistants (e.g. *am – em – im – om*) qui peuvent dès lors être exclues. Il reste 6 catégories présentées dans le Tableau 4, dans lesquelles chaque graphème n'apparaît qu'une seule fois.

Difficulté	Caractéristiques	Graphèmes
1	Graphèmes simples acontextuels consistants	a – i – u – p – b – t – d – f – v – l – r – m – n
2	Graphèmes complexes acontextuels consistants	eu – ou – on – un – oi – oin – ch – gn
3	Graphèmes simples acontextuels inconsistants	é – è – o – ï – y
4	Graphèmes complexes acontextuels inconsistants	ai – au – eau – an – en – in – il(l) – qu
5	Graphèmes simples contextuels inconsistants	e – g – s – c – ç – z – j – x
6	Graphèmes complexes contextuels inconsistants	gu – ss – ge

Tableau 4 : Classification des 45 graphèmes de base du français selon Catach *et al.* (1995) en fonction de leur degré de complexité orthographique.

6 Conclusion

L'association d'une approche psycholinguistique et de l'utilisation de variables linguistiques nous a permis de déterminer 6 degrés de complexité orthographique pour le système graphémique de base du français tel que proposé par Catach *et al.* (1995). La complexité graphémique, le contexte d'utilisation et la consistance des C.P.G. sont des variables sous-lexicales identifiées par la psycholinguistique comme pouvant exercer un effet sur la production orthographique. A ce dernier, il convient toutefois d'ajouter des effets de fréquence qui peuvent être associés, eux aussi, à des unités inférieures au mot. Catach *et al.* (1995) ont souligné l'importance de la fréquence. Ils ont utilisé celle des graphèmes pour les répartir selon trois niveaux. Ils ont également eu recours à la fréquence des C.P.G. en présentant les pourcentages d'utilisation des graphèmes dans la transcription d'un phonème donné (*e.g.* *a* : 92% des transcriptions de [a]; *in* : 45% des transcriptions de [ɛ̃]; *z* : 10% des transcriptions de [z]). Ces deux fréquences d'éléments sous-lexicaux et de leurs relations (*i.e.* fréquence des graphèmes et des C.P.G.) s'avèrent aussi importantes que les trois variables utilisées pour la caractérisation et la hiérarchisation des graphèmes de base (*i.e.* la complexité graphémique, le contexte d'utilisation et la consistance des C.P.G.). Ceci vaut également pour la fréquence d'apparition des phonèmes. En effet, dans le cadre d'un enseignement explicite et formel des principes phonographiques de notre orthographe en lien avec celui de la lecture, cette fréquence phonémique vient peser sur les choix à opérer à l'heure d'établir la progression de cet enseignement/apprentissage. Les C.P.G. correspondant aux phonèmes les plus fréquents gagneront ainsi à être introduites plus rapidement que celles des phonèmes plus rares, ce qui vient pondérer les degrés de complexité orthographique définis dans le Tableau 4. Par exemple, les graphèmes *s*, *c*, *ç* et *ss* qui permettent de transcrire le phonème [s] apparaissent dans ce tableau comme des unités de degrés de complexité élevés (degrés 5 et 6), tandis que le graphème *ch* correspondant au phonème [ʃ] est moins complexe puisqu'il relève du degré 2. Cette différence de degrés pourrait laisser entendre que la correspondance [ʃ] → *ch* doit être enseignée avant la correspondance [s] → *s*, *c*, *ç* et *ss*. Or, l'échelle de fréquence de Haton et Lamotte (1971), cités par Catach *et al.* (1995, p. 49), indique un pourcentage d'occurrences dans les textes plus élevé pour le phonème [s] que pour le phonème [ʃ], avec respectivement 5,90% contre 0,56%. Aussi, l'introduction de la C.P.G. [s] → *s*, *c*, *ç* et *ss* primera sur celle de la C.P.G. [ʃ] → *ch* malgré la plus grande simplicité de ce dernier graphème.

Pour conclure, les variables sous-lexicales telles que la complexité graphémique, le contexte d'utilisation et la consistance phonographique sont reconnues pour peser d'un poids particulièrement lourd sur la production orthographique et l'écriture manuscrite, notamment à leur début, au cours de la phase d'apprentissage (Peereman *et al.*, 2007). D'autres variables ont également été identifiées par la psycholinguistique comme susceptibles de contraindre la production orthographique et l'écriture manuscrite, indépendamment du degré d'expertise des scripteurs. Il s'agit de variables lexicales (*i.e.* concernant le mot) telles que la fréquence, le voisinage, la durée acoustique ou la longueur orthographique. Bien que dépassant le cadre de cet article, leur prise en compte est tout aussi nécessaire dans un objectif d'enseignement/apprentissage de l'orthographe des mots.

Références bibliographiques

- Alegria, J. & Mousty, P. (1996). The Development of Spelling Procedures in French-speaking, Normal and Reading-Disabled Children: Effects of Frequency and Lexicality. *Journal of Experimental Child Psychology*, 63, 312-338.
- Bonin, P., Collay, S. & Fayol, M. (2008). La consistance orthographique en production verbale écrite : une brève synthèse. *L'Année Psychologique*, 108, 517-546.
- Bonin, P., Collay, S., Fayol, M. & Méot, A. (2005). Attentional strategic control over nonlexical and lexical processing in written spelling to dictation in adults. *Memory and Cognition*, 33 (1), 59-75.
- Bonin, P., Peereman, R. & Fayol, M. (2001c). Do Phonological Codes Constrain the Selection of Orthographic Codes in Written Picture Naming? *Journal of Memory and Language*, 45, 688-720.
- Borgwaldt, S. R., Hellwig, F. M. & De Groot, A. M. B. (2005). Onset entropy matters. Letter-to-phoneme mappings in seven languages. *Reading and Writing: An Interdisciplinary Journal*, 18, 211-229.

- Caravolas, M. (2004). Spelling Development in Alphabetic Writing Systems: A Cross-Linguistic Perspective. *European Psychologist*, 9 (1), 3-14.
- Catach, N. (2011). *L'orthographe*. (10^e éd. révisée). Paris : Presses Universitaires de France.
- Catach, N., Gruaz, C. & Duprez, D. (1995). *L'orthographe française*. (3^e éd. révisée). Paris : Nathan Université.
- Cogis, D. (2005). *Pour enseigner et apprendre l'orthographe*. Paris : Delagrave.
- Delattre, M., Bonin, P. & Barry, C. (2006). Written Spelling to Dictation: Sound-To-Spelling Regularity Affects Both Writing Latencies and Durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32 (6), 1330-1340.
- Fayol, M. (2003). L'orthographe française est une des plus difficiles du monde. Comment les enfants en déjouent les pièges ? *Cerveau et psycho*, 3, 2-5.
- Fayol, M. & Jaffré, J.-P. (2008). *Orthographier*. Paris : Presses Universitaires de France.
- Ferrand L. (1999). 640 homophones et leurs caractéristiques. *L'Année Psychologique*, 99 (4), 687-708.
- Ferrand, L. (2007). *Psychologie cognitive de la lecture : Reconnaissance des mots écrits chez l'adulte*. Bruxelles : De Boeck Université.
- Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 361-379.
- Haton, J.-P. & Lamotte, M. (1971). Etude statistique des phonèmes et diphonèmes dans le français parlé. *Revue d'Acoustique*, 16, 258-262.
- Jaffré, J.-P. (1992). Le traitement élémentaire de l'orthographe : les procédures graphiques. *Langue Française*, 95 (1), 27-48.
- Jaffré, J.-P. (2005). Introduction: The orthography of French. *Educational Studies in Language and Literature*, 5 (3), 353-364.
- Kandel, S., Álvarez, C. J. & Vallée, N. (2006a). Syllables as Processing Units in Handwriting Production. *Journal of Experimental Psychology: Human Perception and Performance*, 32 (1), 18-31.
- Kandel, S., Álvarez C. J. & Vallée, N. (2008). Morphemes also serve as processing units in handwriting production. In M. Baciú (Ed.), *Neuropsychology and Cognition of language Behavioral, Neuropsychological and Neuroimaging Studies of Spoken and Written Language* (pp. 87-100). Kerala, India: Research Signpost.
- Kandel, S. & Spinelli, E. (2010). Processing complex graphemes in handwriting production. *Memory and Cognition*, 38 (6), 762-770.
- Kandel, S. & Valdois, S. (2005). The effect of orthographic regularity on children's handwriting production. *Current Psychology Letters* [Online], 17, 3. URL: <http://cpl.revues.org/index463.html>
- Kandel, S., Valdois, S. & Orliaguet, J.-P. (2003). Etude de la production écrite en copie: Une approche visuo-orthographique et graphomotrice. *Le Langage et l'Homme*, 38 (2), 5-24.
- Katz, L. & Frost, R. (1992). The Reading Process is Different for Different Orthographies. The Orthographic Depth Hypothesis. In R. Frost & L. Katz (Eds.), *Orthography Phonology, Morphology, and Meaning*. (pp. 67-84). Oxford, England; North Holland: Elsevier Science Publishers.
- Kreiner, D. S. (1996). Effects of word familiarity and phoneme-grapheme polygraphy on oral spelling time and accuracy. *Psychological Record*, 46 (1), 49-70.
- Lambert, E., Alamargot, D., Larocque, D. & Caporossi, G. (2011). Dynamics of the Spelling Process During a Copy Task: Effects of Regularity and Frequency. *Canadian Journal of Experimental Psychology*, 65 (3), 141-150.
- Lété, B. (2008). La consistance phonographique : une mesure statistique de la complexité orthographique. In C. Brissaud, J.-P. Jaffré & J.-C. Pellat, *Nouvelles recherches en orthographe* (pp. 85-99). Limoges : Lambert-Lucas.
- Lété, B., Peereman, R. & Fayol, M. (2008). Consistency and word-frequency effects on spelling among first-to fifth-grade French children: A regression-based study. *Journal of Memory and Language*, 58, 952-977.
- 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, 156-166.

- Mousty, P. & Leybaert, J. (1999). Evaluation des habiletés de lecture et d'orthographe au moyen de BELEC. Données longitudinales auprès d'enfants francophones testés en 2^o et 4^o années. *Revue Européenne de Psychologie Appliquée*, 49 (4), 325-342.
- Parkin, A. J. (1982) Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depends upon how regularity is defined. *Memory and Cognition*, 12, 287-292.
- Peereman, R. & Content, A. (1998). Quantitative analyses of orthography to phonology mapping in English and French. <http://homepages.ulb.ac.be/~acontent/OPMapping.html>
- Peereman, R., Lété, B. & Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme-phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods*, 39, 593-603.
- Seymour, P. H. K., Aro, M. & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143-174.
- Shen, X. R., Damian, M. F. & Stadthagen-Gonzalez, H. (2013). Abstract graphemic representations support preparation of handwritten responses. *Journal of Memory and Language*, 68, 69-84.
- Sprenger-Charolles, L., Siegel, L. S. & Bonnet, P. (1998). Reading and Spelling Acquisition in French: The Role of Phonological Mediation and Orthographic Factors. *Journal of Experimental Child Psychology*, 68, 134-165.
- Ters, F., Mayer, G. & Reichenbach, D. (1982). *L'Echelle Dubois-Buyse d'orthographe usuelle française*. (6^o éd. critique). Neuchâtel : Mousseillier.
- Van Galen, G.P. (1991). Handwriting: Issues for a Psychomotor Theory. *Human Movement Science*, 10, 165-191.
- Venezky, R. L. (1970). *The structure of English orthography*. The Hague, Netherlands: Mouton.
- Ziegler, J. C., Jacobs, A. M. & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, and Computers*, 28 (4), 504-515.

Annexes

Archigraphèmes	Graphèmes	Sous-Graphèmes
Voyelles		
A	a à â	(em) (prudemment) *(en) solennel *(on) (paonne)
E [e] [ɛ]	e + é (e) + è ai ê ei ë	œ (fœtus) æ (et cætera) aî (paître) *eî (reître) *ea (break)
I	i (y) ï î	hi (trahi) *ee (meeting)
O	o au eau ô (u) (m)	ho (cahot) *oo (alcool) *aô (Saône) *ü(m) (capharnaüm)
U	u û	hu (cahute) *ü (Saül) *eu (j'ai eu)
EU	eu oeu	*ue (cueillir) *(ai) (faisons) *(u) (club) *(on) (monsieur) *eû (jeûne)
	(e)	
OU	ou où oû	*aou (saoul) *aoû (août) *(oo) (football)
AN	an am en em	*aon (faon) *aen (Caen)
IN	in im (en) ain ein	*în (vînmes) *aim (faim) yn (synthèse) ym (symphonie)
ON	on om	*(un) (punch) *(um) (rumsteack)
UN	un	*um (parfum) eun (à jeun)
Semi-Voyelles		
Y, IL(L)	(i) (ï) y il(l), ll	hi (hier) *hy (hyène)

OI	oi (ou + voy.)	ôî (croît) *w (water) *(u) (jaguar) *(œ)(mœlle) *oê (poêle)
OIN	oin ouin	ooing (shampooing)
Consonnes		
P B	p b	pp (nappe) *bb (abbé)
T D	t d	tt (botte) th (théâtre) dd (addition)
C G	c + qu k q g + gu	(cc) (accorder) cqu (acquitter) ck (ticket) ch (technique) *cch (saccharine) *gg (aggraver) *(c) (second) *gh (ghetto)
F V	f ph v	ff (affaire) *w (wagon)
S Z	s + ss (c) + ç (t) (+ i) (s) (intervoc.) z	sc (science) *sth (asthme) *(x) (soixante) *(x) (deuxième) *zz (mezzanine)
X	x cc (+ e, i) xc (+ e, i)	
CH J	ch j (g) + ge	*sch (schéma)
L R	l r	ll (ville) rr (guerre) rh (rhume)
M N	m n	mm (gomme) nn (bonne)
GN	gn	*ign (oignon)
NG	ng	(parking)
N.B. : Il faudrait y ajouter la lettre <i>h</i> , équivalent du tréma à l'intérieur des mots, mais aussi « couverture » de la voyelle initiale dans certains cas : <i>les/héros</i> .		

Tableau A1 : Tableau des graphèmes du français - Niveau 3 (Catach, Gruaz & Duprez, 1995, pp. 14-15).

Archigraphèmes	Graphèmes	% approximatif d'utilisation	Exemples
Voyelles			
A	a	92%	papa
	à	7%	à (la ville)
	â	1%	pâte, pâté
E	e + é	99%	mes, pré
	(e) + è	67,9%	bec, règle
	ai	30%	chair, j'aimais
	ê	2,1%	bête, bêler
	ei ë		pleine Noël
I	i	99%	il
	(y)	1%	type
	ï		héroïsme
	î		gîte
O	o	75%	sot, sotté, cote
	au	21%	cheveu(x), saute
	eau	3%	oiseau
	ô		côte
	(u) (m)		minimum
U	u	100%	tu, pluie
	û		mûr
EU	eu	93%	feu, fleur
	oeu		oeuf
	(e)	non décompté	ch(e)val, autrement
OU	ou	98%	fou
	où		où (vas-tu ?)
	où		goût
AN	an	44%	(un) an
	am		lampe
	en	47%	enlever
	em		embellir
IN	in	45%	fin
	im		impossible
	(en)	23%	chien
	ain	21%	sain
	ein		plein
ON	on	92,8%	son
	om		ombre
UN	un	97%	un
Semi-Voyelles			
Y, IL(L)	(i)	86%	pied
	(ï)		aïeul
	y	3%	payer
	il(l), ll	10%	rail, raille, fille
OI	oi	100%	loi
	(ou + voy.)		loua, troua
OIN	oin	100%	moins
	ouin		chafouin, ine

Consonnes			
Les consonnes doubles (environ 2%) et les lettres muettes (environ 10%) ne sont pas décomptées.			
P	p	100%	pont
B	b	100%	bon
T	t	99%	ton
D	d	100%	don
C	c + qu k	98%	coque coke coq
G	g + gu	100%	goulette
F	f ph	95%	fou phare
V	v	100%	avoir
S	s + ss (c) + ç (t) (+ i)	69% 26% 3,3%	sac, rosse ci, ça nation
Z	(s) (intervoc.) z	90% 10%	rose zéro
X	x cc (+ e, i) xc (+ e, i)	84%	axe, examen accéder, occident excès
CH	ch	100%	cher
J	j (g) + ge	49% 51%	jeu mangea
L	l	100%	la
R	r	100%	ré
M	m	100%	mon
N	n	100%	non
GN	gn	100%	règne
NG	ng	100%	parking

Tableau A2 : Tableau des graphèmes du français - Niveau 2 (Catach, Gruaz & Duprez, 1995, pp. 12-13).