

LE GRAPHÈME

Nina CATACH

Depuis Saussure, on présente le signe linguistique comme ayant deux faces : Signifiant (St) et Signifié (Sé). Depuis Martinet, on accepte un modèle théorique du langage comme comportant deux articulations : première articulation (dont l'unité de base, le monème ou *morphème*, présente bien les deux faces du signe : St/Sé) ; deuxième articulation, dont l'unité de base, le *phonème*, est certes un élément essentiel, constitutif du morphème, mais n'est en lui-même que *distinctif*, et ne présente par conséquent qu'une seule face du signe, la face signifiante.

Bien que cet ensemble théorique ait été critiqué et continué à l'être, nous le prendrons provisoirement pour base, pour tenter le plus simplement possible de situer par rapport à lui l'unité fondamentale de l'écriture (comprise en tant que système de signes graphiques), le *graphème*.

« La langue est indépendante de l'écriture », disait Saussure, qui ne parlera donc que de « graphies ». Il faut attendre une époque assez récente (Stetson, 1937, à notre connaissance) pour qu'apparaisse le mot de *graphème*, terme calqué sur *phonème*, *morphème*, etc. Mais l'apparition du mot ne signifie nullement qu'on ait, du moins en France, tenté d'intégrer la communication scripturale dans l'analyse globale du fonctionnement du langage.

Matériellement, le graphème continue trop souvent d'être confondu avec la lettre :

« Graphème. *En linguistique, unité d'écriture, caractère distinctif du code écrit : Les lettres de l'alphabet sont des graphèmes* ». (Grand Larousse de la langue française, t. 3, 1973).

On retrouve, à des degrés divers, la même confusion entre lettre et graphème chez des gens aussi différents que Thimonnier (*Pour une pédagogie rénovée*, 1974, pp. 31 et 88), Benveniste-Chervel (1969, p. 119 et tableau p. 134), Genouvrier (*Bref*, janv. 1973, n° 8, p. 18) (1), etc., sans compter, naturellement, la plupart des manuels en cours.

(1) — Thimonnier p. 31 : « Ceux-ci [les 97 graphèmes] se répartissent en 26 graphèmes simples (les 26 lettres de l'alphabet) et en 71 graphèmes composés... » ; p. 88 : « Graphème : lettre (graphème simple) ou groupe de lettres (graphème composé) servant à la transcription d'un phonème... ».

— Benveniste-Chervel : « On est donc fondé de réserver le terme de graphème aux vingt-six lettres de l'alphabet... ». Pourtant, les auteurs donnent bien par ailleurs la différence entre graphème et lettre, comme entre phonème et son. Mais il semble que pour eux la lettre constitue l'aspect matériel du graphème, sa « réalisation particulière », italique, majuscule, etc., et non, comme pour nous, quelque chose de fondamentalement différent.

— Genouvrier : « On doit donc s'en tenir, sauf choix totalement arbitraire, à considérer que les graphèmes du français sont les 26 lettres de l'alphabet auxquels sont adjointes cinq signes graphiques non autonomes : les diacritiques ».

Nous n'insisterons pas ici sur cet aspect, pourtant grave en pédagogie, dû à l'absence totale d'analyse dont le graphème a souffert jusqu'ici (2).

Du point de vue de ses rapports avec la langue, le mot écrit (puisque c'est là que Saussure s'arrête, et cela aussi est significatif) n'est pas considéré comme un *signe* linguistique : il est le *reflet* du mot parlé, sa « *photographie* », bien qu'il « *se mêle si intimement au mot parlé dont il est l'image, qu'il finit par usurper le rôle principal* » (C.L.G., éd. crit., p. 45).

Même analyse actuelle, nous l'avons vu dans la définition donnée plus haut, pour ce qu'on appelle le graphème : comme les phonèmes, dont ils sont la « *translation* » (Haas, 1970), les graphèmes ne sont considérés que comme des « *entités oppositives, relatives et négatives* » (Saussure, à propos des phonèmes), c'est-à-dire distinctives seulement. Les graphèmes comme les phonèmes sont des « *figures* », non des signes. Il y a les mêmes rapports entre eux qu'entre les fameuses pièces du jeu d'échecs citées par Saussure, qui peuvent être en bois ou en ivoire sans changer aucunement les règles du jeu.

C'est pourquoi Hjelmslev les appelle des « *cénèmes* » (de *kevos*, « vide »), alors que les morphèmes sont des « *plérèmes* » (du grec « *plérès* », « plein », éléments chargés de sens). A la suite de Haas (1976), nous retiendrons ces deux mots, pour parler des éléments graphiques, afin de ne pas créer nous aussi la confusion en parlant, comme à l'oral, de signifiants et de signifiés (3).

I. Les graphèmes sont-ils des éléments « vides » ?

1) Caractère secondaire de l'écriture

Il nous faut d'abord parler d'un aspect essentiel de la discussion, qui est le caractère *secondaire* du « code » graphique.

La priorité absolue, historique et génétique, du langage oral, le fait que des civilisations entières n'aient pas connu l'écriture (l'homme remonte à 3 millions d'années, l'écriture à quelque 5000 ans) ; l'infinité possible des codes de substitution (dont l'écriture n'est que le modèle privilégié), opposé aux lois générales et spécifiques du langage articulé en tant que système social de communication, la possibilité, abondamment prouvée par le grec, le latin, l'espagnol, l'italien, le turc et de multiples langues dans le monde, de limiter assez strictement le rôle du graphème à une fonction secondaire et bien délimitée de transcription du phonème, tout cela est indéniable.

Cependant, cela ne change rien à la situation qui existe en synchronie, et qui est le « *lien plus ou moins direct entre la chaîne écrite et la réalité extralinguistique* [et j'ajouterai : linguistique] » (Vachek). Par conséquent, le caractère secondaire de l'écriture est aussi peu pertinent pour l'étude du système (je n'en dirai pas autant pour la pédagogie) que l'est l'origine du langage pour les théories linguistiques.

(2) — Voir à ce sujet N.C. « Le graphème et la lettre », *Journal des Instituteurs*, oct. 1977, n° 1. Rappelons que pour nous un graphème peut être constitué d'une ou plusieurs lettres, lettres accentuées ou pourvues de signes diacritiques, présentant certaines caractéristiques à l'intérieur d'un système de signes graphiques d'une part, par rapport au système de langue d'autre part. Le recensement des graphèmes du français a peu de choses à voir avec l'alphabet. Voir plus loin notre définition du graphème.

(3) — Les vues de Hjelmslev sur l'orthographe et son vocabulaire sont bien expliquées dans Benveniste-Chervel, 1969, pp. 40-43.

2) Le phonème est-il le signifié du graphème ?

Nous ne suivrons pas non plus V. G. Gak lorsqu'il dit : « *L'union de la graphie (signifiant) et du phonème qu'elle transcrit (signifié) forme le signe graphique, le graphème, unité minimale constitutive du système graphique* » (1976, p. 24).

Malgré le lien essentiel qui les unit effectivement, il n'y a pas, nous semble-t-il, entre le phonème et le graphème (dans une écriture phonologique ou une notation phonétique s'entend) le même type de rapports qu'entre un signifié et un signifiant.

Lorsque nous disons, par exemple, *man* « signifie » *homme*, *hand* « signifie » *main*, nous produisons une tautologie, qui ne nous apprend rien sur l'homme ou sur la main : un dictionnaire anglais-français établit des rapports entre *signifiants* (type I), à moins qu'il n'ajoute : « *Homme (man)* peut « signifier » soit l'humanité toute entière, soit le contraire de la femme » (type II) auquel cas il touche au signifié de *homme* et de *man*.

Les rapports entre *homme*/[ɔm], *man*/[man], sont fondamentalement de type I, et non de type II. Comme le dit Hjelmslev, ce sont des rapports entre formes de l'expression ayant des substances de l'expression différentes :

Plan de l'expression

	substance	forme	
oral (ex. [ɔm])	substance phonique (phones)	formes phoniques (phonèmes)	Plan du contenu
écrit (ex. <i>homme</i>)	substance graphique (graphes)	formes graphiques (graphèmes)	

Notons cependant, en passant, que nous avons utilisé dans *homme*/[ɔm] deux formes grapho-visuelles ayant bien le même correspondant articulatoire-acoustique, mais l'une en italique, l'autre entre crochets, les caractères étant différents : ces trois traits, italique, crochets, caractères différents, ajoutent des informations sémiologiques évidentes sur lesquelles nous reviendrons tout à l'heure.

3) La phonation est-elle indispensable au langage ?

Même si, du point de vue phylogénique et ontogénique, l'usage de la langue précède toujours la prise de conscience du système de la langue, leur association constitue, selon Saussure, un « dépôt », qui devient premier à son tour : nous intériorisons la parole sous forme de modèle et, à ce stade « *La phonation, c'est-à-dire l'exécution des images acoustiques, n'affecte en rien le système lui-même* » (p. 36). Répondant à Whitney, qui affirmait que « *notre choix est tombé par hasard sur les organes vocaux* », Saussure trouve qu'il va trop loin, mais il reconnaît :

« *La langue est une convention, et la nature du signe dont on est convenu est indifférente. La question de l'appareil vocal est donc secondaire dans le problème du langage* ». (p. 96).

Le fait que « *ce n'est pas le son, c'est la langue qui fait l'unité du langage* » apparaît de façon évidente, si l'on songe aux sourds-muets, au Braille, au morse, à tous les langages de remplacement, par gestes ou par graphismes divers. D'où la possibilité (acquise) pour l'homme cultivé de sauter plus ou moins complètement l'étape phonique, en passant directement du St graphique au Sé, selon le modèle :

<p>St graphique → modèle linguistique → Sé (pattern)</p>
--

4) Le graphème représente-t-il toujours le phonème ?

Centrés que nous sommes sur nos écritures alphabétiques, nous avons tendance à ignorer que même si l'on considère le graphème comme un substitut de langage (ce qui en fait n'épuise pas toutes ses possibilités), ce n'est pas forcément un substitut de phonème. Je ne referai pas ici toute l'histoire des écritures, des *signes-situation* (scènes de chasse, actes religieux, signes-discours) aux *signes-choses* ou *signes-symboles* (pictogrammes), aux *signes-notions* ou *signes-mots* (idéogrammes et logogrammes), aux *signes-morphèmes* (morphogrammes) ou *signes-syllabes* (syllabogrammes), les seconds souvent issus des premiers par le système du rébus. Tous ces types de notations, supplantés il est vrai de plus en plus par les *signes-phonèmes*, infiniment plus économiques (phonogrammes), constituent tous des unités de discours de valeur et de grandeur différentes, qui co-existent encore aujourd'hui dans la plupart des langues, et doivent être considérés au même titre comme des graphèmes (4).

En réalité, chacun des types de graphèmes mentionnés ci-dessus représente une *classe* de signes, qui peuvent être d'une infinie variété suivant les langues. Sapir cite le nootka, langue pour qui la suite « i-n-i-a-l-u-d-am » (« je viens pour le lui donner ») est un « mot », de même que la suite que l'on pourrait traduire « lorsque-comme-on-le-dit-il-eut-été-absent-quatre-jours ». Va-t-on appeler *morphogrammes* tous les signes de catégories, comme les déterminatifs (« classifers » en anglais) de l'ancien égyptien, ou les 214 « clés » du chinois, qui peuvent être soit des indicateurs de séries morphologiques, soit des indicateurs sémantiques, des « sémiogrammes » ? (5). Ce n'est

(4) — Cf. Pulgram (1976) : « In reducing a language to writing, that is, in making visible marks that evoke or recall linguistic performance, it would seem that each mark must represent a syntagme or a lexeme or a morpheme or a phoneme or whatever other kind of unit the inventor of the system may choose as his basis ». Et Pulgram définit le graphème de la façon suivante : « Any minimal functional distinctive unit of writing on whatever level of linguistics analysis » (toute unité distinctive fonctionnelle minimale d'écriture sur quelque niveau d'analyse linguistique que ce soit).

(5) — Ex. de déterminatifs (je dirais plutôt « classificateurs ») morphogrammiques pour l'ancien égyptien : trois petits traits à côté du mot indiquant la pluralité ; ex. de déterminatif sémantique : le même signe « aller » pour tous les mots qui expriment le mouvement ; pour le hittite : le même signe marquant la « déité » placé en tête d'une série de 18 façons différentes d'écrire le nom de la déesse Kupapa (dans Haas, 1976, p. 48) ; ex. de « clé » sémantique pour le japonais : l'image ou cheval suggérant l'ensemble de la catégorie des bêtes (M. Cohen, 1958, planche 19), etc. En ce qui concerne les logogrammes, la différence entre « mot » (?) et morphème est difficile à faire pour les langues à tendance monosyllabique, et elle varie en fonction des langues. En réalité, une classification stricte des écritures est aussi difficile que la classification universelle des langues, et nos connaissances sont très incomplètes. On en est jusqu'ici à classer, très grossièrement, les principales tendances *prédominantes* des écritures, ce qui ne les empêche pas de présenter, pour la plupart, plusieurs sous-types ou intersections. La terminologie et son contenu varieront en conséquence.

pas le lieu d'en discuter ici. Gelb (1973, p. 112) énumère au moins six sortes de *logogrammes* ou signes-mots, mais on discute encore pour savoir si les caractères actuels du chinois, qui notent un ou plusieurs morphèmes, sont bien des signes-mots, etc.

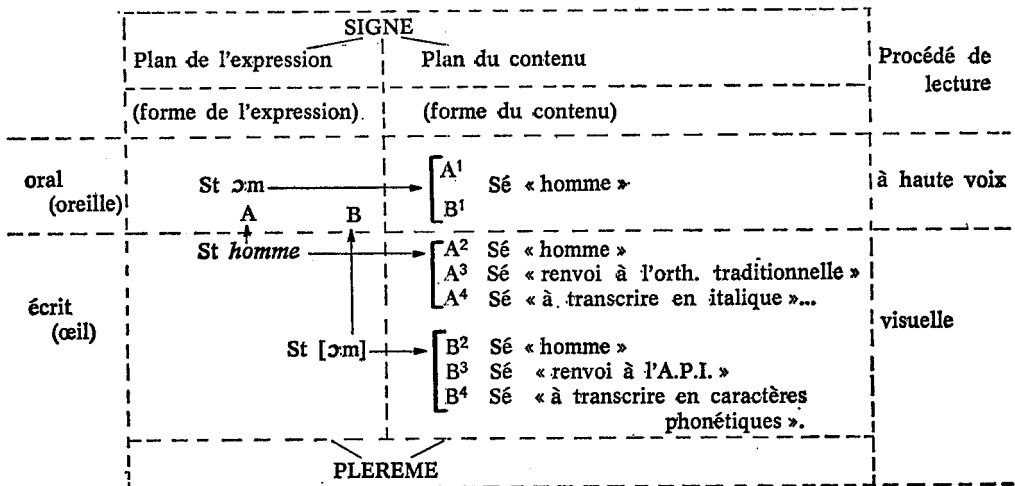
5) Motivation et information des graphèmes

Ce qui est important pour nous, c'est de savoir qu'il devient impossible, dans ces conditions, de parler du graphème comme étant uniquement et toujours un signe « vide » (cénème). Les pictogrammes (comme les deux enfants se tenant par la main et indiquant la proximité d'une école) sont des signes directement ou indirectement « motivés » (*motivated* selon Haas) par rapport à la réalité extérieure. Les idéogrammes et les logogrammes non alphabétiques qui atteignent un haut degré d'abstraction deviennent en général de moins en moins « motivés », de plus en plus seulement « *informed* », c'est-à-dire fonctionnels, par rapport aux autres éléments de la langue.

Ainsi, la forme de nos signes de parenthèses ou de crochets ouverts et fermés évoque plus ou moins l'ouverture et la fermeture, mais surtout ils ont pour fonction d'indiquer une rupture du discours, ou, comme pour les crochets de [ɔm] (homme) tout à l'heure, accompagnés de caractères particuliers, que l'on réfère à l'Alphabet phonétique international, ce qui est une *information*, mais non une *motivation* (6).

Il est difficile de rendre dans un cas comme celui-ci par un schéma toute la complexité des informations apportées, qui (grâce au caractère pluri-dimensionnel de la page écrite) atteignent l'œil et le cerveau en faisceau, mais qui sont en réalité analysées et analysables, cette analyse étant si rapide qu'elle apparaît instantanée.

Essayons cependant :



(6) — Que le mot entier, ou l'image visuelle, prenne accessoirement une motivation d'ordre esthétique, psychologique, etc., ne nous intéresse pas ici, et est valable quel que soit le type d'écriture.

Commentaire : Les graphèmes permettent à la fois le renvoi au St phonique en cas de lecture à haute voix, et le renvoi direct au Sé en cas de lecture visuelle. Les Sé indiqués ici ne sont pas limitatifs, et dépendent des individus et des situations. A noter que nous ne considérons pas les Sé A¹ et A² (homme : genre humain), B¹ et B² (homme : antonyme de femme), etc. comme absolument identiques. Car ils réagissent l'un sur l'autre, tout comme les signifiants entre eux. La « motivation » des crochets ouverts et fermés, évoquant des bras qui entourent, n'est pas mentionnée ici.

6) La double articulation du graphème

Comme on le voit sur ce schéma, le rôle du graphème est double : il est un signifiant (forme écrite) renvoyant à un signifiant forme orale, (c'est un signifiant de signifiant, un cénème), ce qui est son rôle de base dans une écriture alphabétique. Il peut être *en même temps* ou séparément un signifiant de signifié, un plérème. Sa relative indépendance par rapport au phonème fait qu'on peut le prononcer de façon différente, parfois totalement différente dans le cas du logogramme du chinois. Savoir dans quelle mesure, dans ce cas, il est « dérivé » (*derived*, selon Haas) de la langue ou « original », c'est-à-dire autonome, si cette autonomie est immédiate ou acquise, totale ou relative, est une autre question, qui ne sera pas débattue ici. Chose curieuse, Saussure lui-même, pour le chinois, semble considérer leurs caractères comme autonomes :

« Pour le Chinois, dit-il, l'idéogramme et le mot parlé sont au même titre des signes de l'idée : pour lui, l'écriture est une seconde langue... les mots chinois de différents dialectes qui correspondent à une même idée s'incorporent également bien au même signe graphique » (p. 48).

Le cas des signes de ponctuation est particulièrement intéressant. Comme les symboles scientifiques et techniques, les chiffres, les blancs, dans une certaine mesure les sigles, les abréviations et les majuscules, tous signes qui existent pratiquement dans toutes les langues, les signes de ponctuation sont des plérèmes, bien que certains d'entre eux (parfois les mêmes) puissent correspondre à des unités prosodiques de l'oral, des prosodèmes. L'information apportée par nos morphogrammes et nos logogrammes est double elle aussi : une partie des éléments (*inférieurs* au mot, ce qui est important) renvoie au signifiant oral, une autre (parfois la même) rattache le mot à une ou plusieurs catégories sémiques ou morphologiques, et ces éléments peuvent être prononcés ou muets, ce qui rend leur accessibilité beaucoup plus difficile, ex. :

Plan cénémique	Plan plérémiq
[so] } ← sot ← saut ← seau ← sceau	→ sottise... → sauter, saltimbanque... → ? → ?, sceller, scellement...

Le réseau d'oppositions paradigmatiques entre ces quatre mots est triple :

- s initial / sc
- o / au / eau
- t final / marque zéro.

Seul le *s* est ici entièrement « cénémique », tous les autres éléments sont « marqués », les marques des deux derniers mots (leurs « informations ») n'étant apparemment pas motivés, sinon par opposition aux deux premiers, et, pour le dernier, par le lien secondaire établi avec *sceller*, *scellement*, etc.

II. Les graphèmes du français

L'étude générale (trop hâtive) faite précédemment était nécessaire pour mieux situer ce que l'on appelle « l'orthographe » du français, qui est en réalité une *écriture*, d'une haute sophistication il est vrai, mais avec ses défauts et ses qualités. Rappelons que ses unités matérielles ont été empruntées au latin, emprunt qui prouve qu'une écriture, comme le langage d'ailleurs, mais de façon plus nette encore, n'est pas uniquement une abstraction linguistique ; le fait s'est produit pour beaucoup de langues (le sumérien a été emprunté par l'akkadien, le chinois par le japonais, l'alphabet latin par le turc, etc.), avec des succès divers. L'influence alphabétique de l'écriture latine se fait sentir (positivement) jusqu'aujourd'hui, par le fait qu'il est plus facile d'écrire *respectabilité* que *respect*, *ostréiculteur* qu'*huître*, etc... *honorabilité* qu'*honneur*. Rappelons également qu'il s'agit d'une écriture « vieillie » (M. Cohen), et qui aurait sérieusement besoin d'être rajeunie. Celà dit, elle continue à *fonctionner*, dans le sens donné à ces mots tout à l'heure, et ce fait mérite qu'on s'en occupe.

Tout d'abord, et compte tenu du double caractère de cénème et de plérème que peut prendre le graphème, nous en proposons la définition suivante : « *La plus petite unité de la chaîne écrite ayant un correspondant phonique et/ou sémique susceptible d'une analyse linguistique* », ce qui rejoint à peu près la définition de Pulgram.

1) Les phonogrammes

Tout graphème susceptible d'avoir un correspondant phonique est appelé *phonogramme*, même s'il présente d'autres caractéristiques par ailleurs. L'écriture du français est et reste fondamentalement alphabétique, même si elle a un complément plérémique. Il faut le dire et le redire, à l'heure où, confondant perception syncrétique et perception synthétique, psychologie et linguistique, beaucoup parlent à son sujet d'idéographie.

a) *Concentration des règles* : La concentration extraordinaire des « règles phonologiques » de l'écriture du français peut être prouvée de la façon suivante :

— Sur 33 zones d'archigraphèmes (voir lexique p. 20), sept seulement (E, O, AN, IN, S, X, J) comprennent des graphèmes de base d'un rendement inférieur, en discours, à 86 — 90 %, sans l'aide d'aucun contexte, ou avec un contexte très simple (du type *c* (*g, qu*) + *e, i* ; *s* intervocalique ; *i* + voy. ; consonne + *e* à la finale, etc.), tout à fait comparable à ce qui existe dans les autres langues romanes.

— La seule simplification des consonnes doubles, quand elles ne posent pas de problèmes phoniques ou sémiques (4 % des occurrences dans les textes), et des lettres grecques (1 %), amènerait nos phonogrammes consonantiques de base à un rendement approchant 100 %, avec les réserves faites ci-dessus (contexte très simple).

— Si 45 graphèmes de base ont un rendement graphie-phonie d'environ 85 % à 90 %, le passage à 70, puis à 130 et plus, ne permet pas d'augmenter sensiblement le rendement des règles. Un programme de phonétisation automatique (c'est-à-dire de conversion automatique des graphèmes ordinaires en notation phonétique) comprenant environ 1000 règles, n'en utilise en fait que 39 %, les autres cas ne se rencontrant pratiquement jamais dans les textes (programme de Proust, utilisé au LIMSI pour l'Icophone VI).

— Sur 500.000 mots que compte actuellement le vocabulaire large du français, des graphèmes comme *æ, ae, ei, eû, ou, aou, aoû, aon, aim, eun, oi, oê, yn, ym, k, q, sth, ign, sch, etc.*, touchent moins de quelques dizaines d'unités lexicales, et ne risquent plus d'apparaître dans les créations (7).

— Sur quelque 600 mots et formes fléchies parmi les plus fréquentes du français, 37 % sont entièrement phonogrammiques, ce qui ne signifie pas que toute difficulté en soit absente (*Listes orthographiques de base*, à paraître chez Nathan).

b) *Double articulation des phonogrammes* : La caractéristique de nos phonogrammes, comme de ceux de l'anglais, est de présenter ce que l'on pourrait appeler aussi deux articulations, mais différentes de celles de l'oral, où elles ne peuvent apparaître séparément. La face signifiante est très forte, mais non totale, pour les graphèmes représentés par les archigraphèmes : or, c'est la première acquise par les enfants (cf. à ce sujet l'article de M. Chaumont, dans *Repères* 54). Ainsi, ils n'ont aucune difficulté à mémoriser la plupart des consonnes, *p* = [p], *b* = [b], *f* = [f], *v* = [v], etc.

La face plérémique pose des problèmes, non de décodage, mais surtout d'encodage, et nécessite un entraînement global et progressif de l'ensemble de la compétence linguistique. Cependant, là encore, et la preuve en est amplement faite, l'extrême concentration des règles, syntagmatiques et paradigmatiques, ne les place pas hors de portée d'un cursus pédagogique normal, à condition de ne pas mettre la charrue avant les bœufs, et de passer par étapes successives des règles *cénémiques* et *très fréquentes* aux règles *plérémiques* et *peu fréquentes*. Ainsi, *s* et *ss* sont moins « marqués » que *c* et *ç*, qui en général « rappellent », comme on dit, l'étymologie et la filiation des sens ; *an* est moins « marqué » que *en*, pour les mêmes raisons, etc.

Les principales valeurs plérémiques (morphogrammiques) des phonogrammes vocaliques internes ne sont pas très nombreuses :

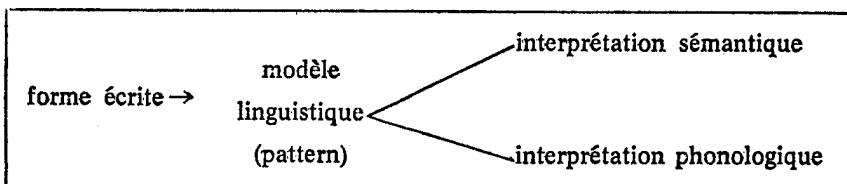
— *ai* : 30 % d'occurrences dans les textes, concentration dans les finales, information grammaticale et lexicale (alternance avec *a*, nasale *ain*), ex. finales verbales, adj. en *ain, aine*, alternance *sain/saine/santé/sanitaire*, etc. ;

— *au* : 21 % d'occurrences dans les textes, information phonogrammique (note [o] en syllabe fermée, *saule, épaule, aube*, etc.), grammaticale (pl. en *aux*) et lexicale (alternance avec *al, autrui/altruisme, autre, alterner*).

— *eau* : 3 % d'occurrences seulement, concentration quasi-exclusive dans les finales, information suffixale (diminutif) et lexicale (alternance avec *el*), ex. : *agneau/agnelle/agneler*, etc.

C'est ici que l'on peut, de la façon la plus nette, montrer la double articulation décrite par Haas (p. 88) :

(7) — Selon nous, tout graphème se retrouvant dans moins de 5 radicaux ou familles lexicales est à considérer comme logogrammique. Même remarque pour les noms propres, mots étrangers, etc.



Pour les consonnes, si l'on excepte là encore les finales, c'est sans doute le redoublement à la limite préfixe-radical qui semble la principale valeur morphogrammique : *cc + e, i* (*accès* de *accéder*, et non *axé* de *axe*), *xc* (*excès*, de *excéder*), *nn* (*innerver*, mais *inouï*), *mm* (*immaculé*, mais *imaginer*), etc. Elle peut avoir, comme on le voit, des répercussions sur la prononciation. Mais elle semble beaucoup moins accessible à la conscience linguistique spontanée, surtout chez les enfants.

2) Les morphogrammes

Ce n'est donc que pour l'analyse que nous séparons nettement phonogrammes et morphogrammes, comme d'ailleurs morphogrammes et logogrammes. Mais cette séparation prend, pour le français, toute sa valeur à la finale. Il vaudrait mieux parler, comme le propose C. Gruaz, de *valeur* ou *fonction* phonogrammique, morphogrammique, logogrammique du graphème, et on y est obligé dans certains cas.

Le statut général du morphogramme a été mal étudié jusqu'ici. Parmi les « classiers » mentionnés par Gelb (pp. 115-126) on trouve autant des signes de ponctuation (marques non prononcées séparant les mots dans les hiéroglyphes), que des indicateurs sémantiques, phonétiques ou prosodiques. Beaucoup, comme en français, avaient un équivalent phonique qui a cessé, en partie ou complètement, d'être prononcé. Selon les cas, ces marques graphiques permettent d'opposer des différences de nature, de fonction (genre, nombre, mode, temps, personne, etc.) ou de familles lexicales (ainsi les catégories « divin », « animaux », « mouvement », etc.). Le gros problème, nous l'avons dit, se pose pour les langues entièrement ou partiellement monosyllabiques, lorsque le morphème se confond avec le « mot », comme, par exemple en grec et en français, pour les mots-outils (*ce/se, ces/ses*, par ex., sont-ils l'un ou l'autre ?). D'ailleurs l'absence de définition satisfaisante pour le « mot », pourtant réalité linguistique universelle bien vivante, fait que la plupart des analyses confondent dans une même catégorie lexème et morphème, ce qui n'a guère de conséquence à l'oral, mais en a de sérieuses à l'écrit. La tendance actuelle est de parler pour le chinois d'une écriture non pas idéographique, ni logogrammique, mais, pourrait-on dire, « morphémogrammique », la notation du morphème étant dans ce cas globale, non analysable en unités plus petites. Des indicateurs sémantiques (« radicals » en anglais) accompagnent aujourd'hui la majorité des caractères (bisyllabiques) chinois, et on y trouve toujours autant de graphèmes que de morphèmes. Mais, et c'est là que nous retrouvons, de façon démesurément amplifiée, les problèmes du français, 80 % des syllabes du chinois sont homophones-hétérographes, situation qui constitue la justification de base du logogramme français. Ainsi, [nán], écrit avec trois caractères différents, signifiera « difficile », « sud » ou « mâle », etc. L'écrit associe donc plusieurs sens à une seule forme phonique, celle-ci étant notée également, très souvent, par un indicateur supplémentaire.

Ce que nous appelons morphogrammes, pour le français (5 à 6 % des graphèmes en discours), ce sont, essentiellement, nos marques finales muettes (*s, t, e* en particulier), qui peuvent être prononcées dans certaines conditions (en liaison, et aussi lors des modifications paradigmatiques du même mot dans les alternances du genre, de nombre, de flexions verbales, le mot écrit constituant une sorte de matrice permettant en principe de planifier la réalisation de surface). Un morphogramme qui n'a plus aucun lien avec l'oral peut être, sauf exception, considéré comme se détachant du système, à moins qu'il ne présente par ailleurs une valeur logogrammique, ce qui sera discuté plus loin.

Par exemple, dans la mesure où la liaison se fait de plus en plus rare dans les formes non inversées des verbes, comme dans *je prends une tarte* (verbes en /r/, première pers. sing.), et où il n'y a pas de forme inversée, contrairement à la 2^e et 3^e pers. (*prends-en, prenez-en, prend-il, prennent-ils*, et à la première personne du pl. (*prenons-en*), cet *s* se détache du système. De même, dans la mesure où l'on ne dit plus jamais *un gras enfant, un saoul homme*, etc., alors qu'on dit toujours *un petit enfant, un grand homme*, les deux premiers morphogrammes sont deux fois moins fonctionnels que les deux derniers, parce qu'ils ont perdu l'information potentielle d'ordre syntagmatique pour ne plus garder que leur information paradigmatique (lien masc./fém.).

Un tel sous-système (complémentaire et non essentiel, notons-le) ne vaut que par l'extrême concentration des règles, ou risque de ne plus jouer, ce qui est souvent le cas, son rôle dans la conscience linguistique. C'est pourquoi nous faisons d'extrêmes réserves, contrairement à ce qui concerne ces marques grammaticales universelles (notion de genre, de nombre, de personne, de temps, etc.), sur ce que nous appelons les morphogrammes lexicaux, reliant un radical à quelques dérivés, chaque « famille » formant à la limite un micro-système. Le lexique, surtout en français, contrairement à la morphosyntaxe, est un système ouvert, en perpétuel renouvellement et d'un grand éparpillement. Une écriture alphabétique ne peut pas être vraiment une écriture de mots, ni de familles de mots. D'où les contradictions insolubles dans lesquelles nous nous trouvons, dans le cas de dérivation irrégulière (*cru, croire, crédit, croyance...*), multiple (*tabac/tabagie/tabatière*), graphiquement contradictoire (*bras/bracelet, nourrice/nourrisson*) ou en l'absence, fréquente, de lettre muette (*chapeau/chapeauter*, ou le récent *bureau/bureau-tique*, etc.).

Le morphogramme français semble rejoindre (rarement, et jamais avec le même degré de globalité) le « morphémogramme » du chinois, lorsqu'il dépasse le niveau du graphème pour atteindre celui de la syllabe. Ce sont les suffixes *-tion* ou *-ment*, les préfixes *en* et *in*, *amphi-*, *poly-*, etc.) dans leur ensemble que nous conservons en français, et sens et forme sont difficiles à séparer ici. De proche en proche, les indices de catégorie peuvent pénétrer le mot : *prolétaire/prolétariat, livraison/conjugaison*, etc. On arrive ainsi insensiblement au logogramme, mais le morphogramme est en principe toujours inférieur au mot.

3) Les logogrammes

Il ne s'agit en fait que d'une adjonction marginale au système (5 à 6 % des « mots » graphiques en discours), et d'ailleurs se comptant par unités, rarement par séries. Ces unités, en général anciennes, réduites le plus sou-

vent à un ou deux phonèmes, fragilisées par le temps et les aléas syntaxiques et prosodiques du français, sont devenues homophones. Certains graphèmes anciens, ou des lettres muettes, leur ont conféré une sorte de physionomie globale, qu'elles peuvent perdre si elles sont soumises à une analyse linguistique.

Ex. *champ/champêtre, chant/chanter, vain/vanité, vin/vigne, etc.*

Leur utilité est très contestée, surtout lorsque les lettres logogrammiques ne présentent pas ou plus d'information autre qu'étymologique ou historique.

A notre avis, chaque cas devrait être reconsidéré en fonction d'un certain nombre de critères : nécessité d'une certaine fréquence, étymologie différente (où irions-nous, si tous les mots de même étymologie, mais polysémiques devaient s'écrire différemment, comme *dessein* et *dessin, différent* et *différend* ?) mots de même nature et de même catégorie grammaticale, mots pouvant se trouver dans le même contexte, etc. Ainsi, on peut estimer inutile de distinguer par la graphie un *tour* et une *tour*, un *voile* et une *voile*, déjà opposés en genre ; de même, qui va confondre un *mythe* et une *mite*, un *cygne* et un *signe*, qui ne se trouveront jamais dans le même contexte ? Malgré tout, pour les mots-outils en particulier, on ne peut nier, semble-t-il, l'information immédiate apportée à la *lecture* en cas de défaillance du système. En pédagogie, les logogrammes ne peuvent être, pour les plus courants d'entre eux, que mémorisés et écrits durant un certain temps avec l'aide du maître, en attendant une analyse lorsqu'elle est possible. Cependant, leur petit nombre et leur spécificité (distinction d'homophones) ne permet en aucune façon d'extrapoler et d'affirmer, là encore, qu'il s'agit du fondement de notre système d'écriture.

4) Les lettres étymologiques et historiques

Elles ne répondent pas, selon nous, aux critères de définition du graphème (8). L'exemple type est celui des consonnes doubles, ou de l'accent circonflexe. En dehors des fonctions annexes que de telles lettres peuvent jouer dans certains cas, elles constituent purement et simplement dans nos rapports phonie-graphie une entropie, un « bruit », comme disent les spécialistes de l'information. En attendant une réforme qui s'impose, l'attitude du maître devrait donc être bien différente à leur égard, et tout le comportement futur de l'enfant vis-à-vis de l'orthographe, de l'écrit, voire de ses études peut en dépendre.

On voit donc que le graphème du français peut fondamentalement prendre deux formes : signe-phonème ou signe-morphème, les deux fonctions étant en général nettement distinctes. Dans les deux cas, l'analyse est pratiquement (pour l'adulte cultivé, non toujours pour l'enfant) possible *au-delà* de l'unité-mot, et devrait être poussée aussi loin que possible. Si l'on compare notre système à celui de l'anglais par exemple, on s'aperçoit que chez nous les rapports phonogrammiques sont meilleurs et plus concentrés (ce qui est dû... aux réformes académiques du XVIII^e siècle !), que les phonogrammes de l'anglais présentent presque tous la double articulation, la plupart d'entre eux cumulant la fonction phonogrammique avec une ou plu-

(8) — Pour plus de détails sur les différents critères qui opposent les graphèmes aux sous-graphèmes et aux lettres hors-système, ainsi que sur le plurisystème graphique du français, voir la bibliographie, publications CNRS-HESO.

sieurs valeurs de dérivation lexicale ou de distinction logogrammique, ce qui explique leur éparpillement et rend leur description difficile. En revanche, l'anglais a conservé phoniquement ses marques grammaticales, se contentant de leur attribuer une physionomie graphique stable malgré leur réalisation phonétique de surface (ainsi pl. *es* → [s], [z], [iz], marque du passé *-ed* → [d], [t], [ɪd], etc.). En grec, de même, les morphogrammes sont tous prononcés, mais ils sont écrits différemment, et l'importance de la conservation des graphies historiques permet un grand développement du système logogrammique. Les applications d'une grille générale sont en cours pour de nombreuses langues dont les systèmes sont mixtes comme le nôtre, et donnent de bons résultats.

On voit qu'une bonne analyse du graphème peut permettre non seulement d'améliorer la recherche des indices de lecture et les performances d'écriture, mais d'aborder dans les meilleures conditions les langues étrangères, en éclairant leur système graphique comme leur système de langue. C'est d'autant plus important qu'ici l'équation est double : rapports phonie/graphie de la langue de départ, rapports phonie/graphie de la langue d'arrivée. Plus encore pour l'enfant d'immigrés, pour qui, au mieux, les inconnues sont au nombre de trois (graphie de la langue maternelle ; phonie et graphie de la langue à acquérir). L'enseignement phonique et phonogrammique est ici surtout prioritaire. Dans tous les cas, comme pour tous les systèmes, il faudrait aller des tendances *prédominantes* de l'écriture concernée vers les zones marginales, même s'il est souvent nécessaire d'aborder très tôt celles-ci dans certains cas (logogrammes fréquents).

Je ne voudrais pas finir cette étude sans déboucher, ce qui est nécessaire, sur les implications pédagogiques et sociales d'une réflexion théorique. L'écriture reflète non pas tant la langue elle-même que la conscience que l'on a de sa langue à un moment donné. Plus la conscience linguistique s'améliore, sur les deux plans, plus les choses s'éclairent, se décantent et peuvent se modifier. Les enseignants ont un rôle important à jouer à cet égard, en tournant le dos à une application mécanique des règles, pour une meilleure appropriation générale de l'écriture.