

Version française du texte paru sous le titre: « Unidades de lengua y unidades de escritura. Evolucion y modalidades de la segmentación gráfica », in Emilia Feirreró, compiladora., *Relaciones de (in)dependencia entre oralidad y escritura*, Barcelone, Gedisa, Colección LeA, 2002, 51-71 (traduction portugaise. "Relações de (in)dependência entre oralidade et escrita", Porto Alegre, éd. Artmed, 2003).

UNITES DE LANGUE ET UNITES D'ECRITURE

EVOLUTION ET MODALITES DE LA SEGMENTATION GRAPHIQUE

Marie-José Béguelin
Université de Neuchâtel

Introduction

Le présent chapitre est consacré aux principes qui régissent l'élaboration des unités d'écriture. Il y sera question de la segmentation graphique ainsi que des divers types de pertinence qui la sous-tendent. Y seront évoqués, sous quelques-uns de leurs aspects, les systèmes d'écriture qui ont servi à consigner les langues de l'Antiquité (grec, hittite, latin...), auxquels seront comparées, le cas échéant, certaines pratiques d'écriture contemporaines.

L'objectif de cette étude n'est pas, au premier chef, de contribuer à l'histoire des systèmes d'écriture. Si le témoignage des langues anciennes y est sollicité, c'est d'abord et surtout en vue de dégager les principes pragmatiques fondamentaux selon lesquels les sujets écrivains (re)créent et gèrent, depuis des siècles, les unités d'écriture.

Avant d'entrer dans le vif du sujet, on rappellera brièvement les conceptions parfois conflictuelles auxquelles a donné lieu, tout au long du XXe siècle, le statut de l'écriture. Dans la foulée du *Cours de linguistique générale* de Saussure, la vieille controverse sur la relation écrit - oral a été périodiquement ravivée: s'agit-il bien, comme on le pense communément, d'une relation de représentation, et si oui, de quelle nature? Par ailleurs, la linguistique des débuts a-t-elle eu raison de conférer une primauté à l'oral pour décrire scientifiquement le fonctionnement de la langue et des langues? Après avoir tiré les leçons d'un débat qui n'a perdu ni de son acuité, ni de son actualité, on verra, en deuxième partie, comment la question de la représentation graphique peut être repensée à la lumière de ce que nous avons ailleurs appelé une *pragmatique des opérations de segmentation* (Berrendonner & Reichler-Béguelin, 1989). À l'appui de ce propos, nous examinerons le fonctionnement d'unités graphiques comme le syllabogramme ou la lettre de l'alphabet, ou encore, à un échelon supérieur de la structure, des démarcateurs de "mots". Enfin, en troisième et dernière partie, nous tirerons quelques conclusions à portée plus générale sur la construction des unités graphiques et sur les propriétés

du savoir épilinguistique¹ véhiculé par les systèmes d'écriture. Ces propriétés s'éloignent assez notablement, on le verra, de la rationalité qui fonde les entités de la linguistique scientifique, élaborées avec un souci de consistance logique et dans une visée démonstrative.

I. Le statut linguistique de l'écrit

En simplifiant beaucoup, et en polarisant le débat, on peut dire que les philosophes, les grammairiens, les linguistes qui se sont penchés sur le problème de l'écrit se sont partagés en deux positions antagonistes².

Les premiers ont conçu l'écriture comme un *code substitutif*, voué purement et simplement à la transcription du langage parlé, un code au second degré dont les unités "signifient" des sons ou des suites de sons. On peut suivre d'Aristote au premier structuralisme, en passant par Rousseau³, cette ligne de pensée qualifiée de *phonocentrique*, pour qui les signes vocaux sont le seul véhicule direct, si ce n'est légitime, des significations, et pour qui l'écrit n'a pas d'autre fonction que d'être le médium — souvent trompeur — de l'oral. Dans ce courant, l'apparition à la fois tardive et partielle de l'écriture dans l'histoire des collectivités et des individus permet de traiter la graphie comme un phénomène secondaire, ne méritant qu'une place accessoire dans la description du langage⁴. D'un point de vue proprement scientifique, la position phonocentrique tend à vider de tout intérêt la question de savoir quelles fonctions sémiotiques assument les graphèmes et les marques de ponctuation, dans la mesure où ces fonctions, visiblement, ne se réduisent pas au principe phonographique; radicalisée, elle accrédite l'idée simpliste, mais relativement répandue dans le public cultivé, selon laquelle les systèmes d'écriture à base phonologique seraient les plus évolués.

En réaction à ce courant de pensée, plusieurs auteurs, philosophes ou anthropologues, ont cependant revalorisé l'écriture comme médium linguistique à part entière, invoquant les effets décisifs que son apparition exerce sur les processus cognitifs et sur l'évolution des institutions sociales⁵. De leur côté, les linguistes se trouvaient face au problème que la quasi totalité des systèmes "naturels" de notation du langage fonctionnent tôt ou tard, et fort avantageusement,

¹ Ce terme a été introduit par A. Culioli, et repris par S. Auroux, pour qualifier la "métalinguistique spontanée" qui s'exprime dans l'activité langagière, à l'écart de toute préoccupation scientifique au sens strict.

² Nous synthétisons ici les développements présentés dans Reichler-Béguelin, 1992.

³ Cf: "L'analyse de la pensée se fait par la parole, et l'analyse de la parole par l'écriture; la parole représente la pensée par des signes conventionnels, et l'écriture représente de même la parole." (Rousseau)

⁴ "Certaines applications de la linguistique, comme les recherches relatives à la traduction mécanique, par l'accent qu'elles mettent sur la forme écrite du langage, pourraient faire croire à l'importance foncière des divisions du texte écrit et faire oublier que c'est de l'énoncé oral qu'il faut toujours partir pour comprendre la nature réelle du langage humain", écrit Martinet, 1966: 53. En rapport avec cette citation, il faut souligner que si la primauté de l'oral a été mise en avant avec autant de détermination par le premier structuralisme, c'est qu'il avait à rompre avec le "graphocentrisme" propre à la tradition littéraire et philologique héritée du XIXe siècle.

⁵ Cf. entre autres Derrida, 1967, Goody, 1979 et 1986, Detienne (éd.), 1988.

avec une composante idéographique⁶. Au même titre que certains acquis de la théorie de la lecture, ce recours fréquent à l'idéographie, sur laquelle nous reviendrons plus bas, montre bien que la phonation n'est pas un truchement obligé, et que, dans une théorie de la compétence langagière, la préséance de l'oral — essentiellement historique et acquisitionnelle — ne doit pas être surévaluée.

Sur la base de telles réflexions, certains auteurs ont été conduits à caractériser l'écrit et l'oral comme des *codes parallèles*, ne différant entre eux que par la substance des signifiants, elle-même conditionnée par ces facteurs externes que sont les avantages respectifs des sons et des graphies selon les conditions de production et de réception des énoncés⁷. A l'opposé du courant phonocentrique, l'école dite *autonomiste*, non contente de réclamer un statut linguistique pour l'écrit, est allée jusqu'à préconiser une approche du signifiant graphique en abstraction du langage oral, rompant le plus possible avec la conception "représentationnaliste" de l'écriture⁸. D'autres théoriciens comme Blanche-Benveniste & Chervel, 1969: 40, se sont bornés à refuser d'"établir une hiérarchie" entre les manifestations écrite et orale, de "classer l'une comme primitive et l'autre comme secondaire".

Si donc, en accord avec ces propos, on admet qu'en synchronie, l'écrit, autant que l'oral, "réalise" la langue, la question de la segmentation graphique et des unités d'écriture retrouve la pertinence théorique que lui déniait le point de vue phonocentrique; il redevient légitime, en particulier, de s'intéresser aux procédures d'analyse reflétées par les unités d'écriture et aux fonctions qu'elles assument. _En figurant spatialement la chaîne linguistique, en la pérennisant, l'écriture porte en effet la trace d'une activité d'analyse qui, à la longue, accrédite certains découpages et certains classements. Pour un sujet lettré ou alphabétisé, la connaissance concrète des structures linguistiques est grandement déterminée par la modélisation rudimentaire, par la *grammaire*, pour faire une figure étymologique, qu'en fournit l'écriture⁹. Afin de mieux comprendre le point de vue non savant sur la langue, il importe donc de mettre en lumière les propriétés des modélisations véhiculées par les systèmes graphiques, au regard des analyses conduites selon les principes de la linguistique scientifique.

2. Élaboration et fonction des unités graphiques

2.1. Le graphème

⁶ Ainsi, corrélée à la séquence phonique [ilpaRl], la séquence graphique du fr. *ils parlent* contient les phonogrammes *i, l, p, a, r, l*, mais aussi un morphogramme discontinu à valeur de pluriel *s.....nt*, dépourvu de réalisation phonique correspondante - sans compter l'intermot, "topogramme" à fonction démarcative.

⁷ Cf. Prieto, 1975: 85-92 et 1983; voir déjà Vachek, 1939.

⁸ Voir à ce sujet le débat entre l'autonomisme radical d'Anis, 1983 et 1988, et l'autonomisme modéré de Catach, 1988: 23.

⁹ "Si donc la théorie de la langue qui se matérialise sous forme d'écriture est bien à rejeter comme théorie linguistique *per se*, elle doit par contre y être incorporée en tant que support (entre autres) de cette composante régulatrice qui fait partie intégrante de toute activité langagière." (P. Achard, in Catach (éd.) 1988: 74)

Au niveau inférieur de la structure linguistique, l'articulation "naturelle" de la chaîne assurée par les systèmes d'écriture présente des caractéristiques qui prennent tout leur sens une fois qu'on les rapporte aux autres niveaux d'analyse (celui du "mot", mais aussi celui de la "phrase" qui, pour des raisons de place ne sera pas évoqué ici¹⁰).

Dans les traditions graphiques des langues indo-européennes, l'adoption d'une écriture étrangère d'une part, le traditionnalisme graphique d'autre part, entraînent toutes sortes de conséquences sur les correspondances qu'il est possible d'établir entre les figures du code écrit et celles du code oral parallèle. Ces conséquences sont couramment mises au compte des imperfections et des illogismes propres à la plupart des systèmes d'écriture. Toutefois, il est possible de dégager certaines formes complexes de corrélation entre atomes graphiques et entités phoniques ou sémantiques correspondantes. Les unités non formelles, créées, récupérées ou bricolées à des fins techniques, portent en effet la trace d'un mode de structuration pragmatique du langage qui obéissent à des règles propres, en vertu de finalités spécifiques. Le grec mycénien et le hittite, consignés dans des syllabaires, nous en fourniront des exemples.

Au regard d'une analyse linguistique de type théorique, les graphies traditionnelles pèchent souvent soit par carence, soit par excès. Les deux tendances s'expriment simultanément en linéaire B, syllabaire de 89 signes qui, dans la seconde moitié du deuxième millénaire a. C., a servi à transcrire le grec mycénien. En effet, ce système graphique note les deux liquides *r* et *l* par une série unique de signes, ne distingue l'opposition sourde/sonore que pour les dentales, omet dans certaines conditions les sifflantes, liquides ou semi-consonnes "satellites" de la syllabe (ainsi *pe-ma* vaut pour *spér-ma*, cf. Lejeune, 1958: 321-330). Ces particularités font des syllabogrammes autant d'unités abondamment polyphoniques et sous-spécifiées, dont la valeur doit être récupérée contextuellement. Au contraire, certains groupes consonantiques, certaines implosives, ainsi que le *u* second élément de diphongue, sont assumés par la graphie. Celle-ci en devient obligatoirement redondante, un syllabogramme équivalant par approximation à une consonne simple, ainsi *ti*, *ke*, et *ko* dans:

ti-ri-po = *tripos* ("trépied")
ke-se-nu-wo = *xénwos* ("étranger")
te-ko-to = *tékton* ("charpentier").

Un exemple plus net encore de la polyvalences des atomes graphiques est fourni par le hittite, langue indo-européenne consignée un peu plus tôt dans un syllabaire de 137 signes lui-même adapté du cunéiforme sumérien. En hittite, un graphème peut théoriquement représenter (d'après Kronasser, 1966: 3 sq.):

1. un sumérogramme en valeur idéographique (il existe des sumérogrammes *tantum*, d'autres apparaissent aussi en valeur phonographique);
2. un déterminatif sémantique à valeur classificatoire, non prononcé, notant un trait de contenu (p. ex. GIS "bois");
3. un syllabogramme notant un groupe CV, VC, parfois CVC;

¹⁰ Cf. sur ce point Berrendonner & Reichler-Béguelin, 1989; Béguelin, dir., 2000.

4. dans certaines circonstances contextuelles, une consonne simple, implosive ou explosive (ainsi *pa-ra-a* = /pra/ "en avant");
5. une des voyelles simples *a, e, i, u* ;
6. une voyelle ou un groupe non prononcé, servant d'auxiliaire à l'élucidation d'un syllabogramme adjacent (quand il s'agit d'une voyelle, ce n'est pas l'indice d'une opposition quantitative).

Cette situation, à première vue anarchique, résulte de la façon dont un même matériau est réutilisé à des fins diverses, parfois pour satisfaire à des exigences externes, découlant directement des circonstances concrètes de production du message. Cela transparaît bien dans le cas n° 6, où un syllabogramme est exploité par certains scribes non, au départ, pour "représenter" directement la chaîne linguistique, mais pour faciliter la procédure de décodage en glosant par répétition la voyelle ou le groupe VC inhérents au syllabogramme précédent. Facultative, cette *scriptio plena* entraîne des variations dans l'orthographe, tout en ayant pour conséquence remarquable, mais sans doute fortuite, de faire tendre le système d'écriture vers une écriture phonologique. Comparer les translittérations pour /apas/ "celui-là":

- graphie simple: *a-pa-as*
- graphie "pleine": *a-pa-a-as* (la redondance du *a* entraîne, incidemment, une identification entre le nombre des graphèmes et celui des phonèmes de la séquence orale parallèle)

Dans sa composante idéographique résiduelle, dont témoignent les cas 1 et 2, l'écriture hittite cumule la référence à deux niveaux d'analyse différents du contenu: elle peut renvoyer au signifié d'un morphème, mais aussi à certains traits sémantiques plus ou moins assimilables à des classèmes. De même, les déterminatifs de catégorie du sumérien et les "clés" du chinois avaient pour fonction, à l'origine en tout cas, de distinguer des graphèmes polysémiques en les affectant à une catégorie donnée de signification. Ici encore, l'analyse n'est pas poursuivie comme un but en soi, elle intervient au hasard des besoins lorsque certaines désambiguïisations se révèlent nécessaires.

L'alternance entre la référence idéographique et la référence phonique syllabique (cas n° 1 et 3), héritée de la vieille technique du rébus, se rencontre également, à un stade ou à un autre de leur évolution, dans les systèmes graphiques sumérien, égyptien, chinois, et dans le linéaire B déjà rencontré plus haut. Celui-ci utilise en outre le procédé fort intéressant du "symbole acrophonique", en quelque sorte symétrique de celui du rébus. Ainsi, l'emploi du syllabogramme ZE valant pour l'unité lexicale *zeûgos* "paire", enlève au graphème le rôle de représentant exclusif de la phonie pour l'affecter également, par une sorte d'équivalence synecdochique, à la représentation du signe global. Alors que le rébus permet pratiquement le réemploi d'un équivalent de signe comme équivalent de syllabe, le symbole acrophonique, par un phénomène inverse, assure le passage, pour un graphème donné, de la représentation syllabique (ou phonologique) à une représentation de signe relevant de la première articulation (i.e. de l'articulation en morphèmes).

On trouve des exemples comparables dans certains systèmes sténographiques modernes (Chouvet, 1953), qui ne sont autres que des alphabets phonétiques, mais où divers symboles, dont la valeur primitive est de noter une consonne, servent aussi subsidiairement à transcrire un morphème commençant par cette consonne. Le même sténogramme est ainsi employé à plusieurs fins, pour représenter tantôt une unité distinctive, tantôt une unité significative:

- = [t] ou [tout, tous, toute, toutes]
| = [b] ou [bien], etc.

Soumis à des contraintes externes comme la rapidité et/ou l'étroitesse de l'écran disponible, les internautes et les adeptes des SMS (Short Message Service) sur téléphone portable recourent à des innovations graphiques nombreuses (cf. Anis, 1998), conditionnées de même par l'objectif pragmatique de "faire court et rapide". Concourent à cette visée abrégative des procédés très divers, dont certains renouent, le plus naturellement du monde, avec ceux qui caractérisent les syllabaires mycénien et hittite:

1. utilisation de lettres ou d'idéogrammes en valeur phonétique (rébus):
U2 = [you too]
Y A KELK1??? = [y a quelqu'un]
2. usage de symboles acrophoniques:
asv? = [âge sexe ville?]
mdr!!! = [mort de rire!]
Q = [quoi],. [qui] ou [que] p. ex. dans: CE Q JE PENSE
3. réemploi de phonogrammes en valeur syllabique:
CT = [c'était]
4. recours à l'idéographie:
@+ = [à plus] (prononcé "à pluss", troncation courante en français familier pour "à plus tard")
5. recyclage de divers signes en constituants d'icône:
8-)) :-p = signes non linguistique, évoquant un porteur de lunettes souriant ou un personnage tirant la langue, empruntés au répertoire des *smileys* ou *binettes*

La vitalité de la tendance idéographique, syllabographique ou acronymiques dans les usages graphiques les plus récents vient bel et bien démentir l'idée reçue selon laquelle les écritures à base phonologique constitueraient, forcément et en toute circonstance, un "progrès" sur les autres systèmes graphiques.

2.2. Le découpage en "mots"

2.2.1. Mot vs morphème

Le mot est une entité traditionnelle découlant de la pratique spontanée des sujets parlants, au moins dans les langues de type indo-européen dès leurs plus anciennes attestations. Après avoir été supplantée par le morphème, défini comme l'unité significative minimale, la notion de mot a retenu l'attention de nombreux linguistes qui ont cherché à la promouvoir au rang d'entité formelle, rigoureuse. Les

diverses définitions, proposées à partir de critères divers également, se heurtent malheureusement toutes à des difficultés:

- Le critère intonatif du mot comme unité accentuelle pose problème pour les langues dépourvues d'accent de mot; dans les langues qui en sont pourvues, elle laisse en suspens la question des formes clitiques (= inaccentuées).

- Le critère de l'autonomie syntaxique ne rend pas justice à certains éléments non autonomes, ou connaissant de fortes restrictions sélectives, que les sujets parlants considèrent néanmoins spontanément comme des mots (exemples: les prépositions, les articles, les pronoms conjoints).

- Le critère sémantico-référentiel (mot = unité de désignation) se heurte au cas des locutions du type *chemin de fer*, *pomme de terre*, où le fait qu'il y ait "unité de désignation", c'est-à-dire renvoi à un référent unique, n'enlève rien à l'aspect compositionnel du signifiant (ainsi, dans nos deux exemples, la métaphore originelle reste pleinement disponible).

À chaque fois que l'on cherche à radicaliser un point de vue et à définir la notion de mot en vertu d'une pertinence unique, des contre-exemples se présentent. Or, si l'on admet avec Martinet, 1966, que le mot n'est *pas* une notion formelle, on doit trouver naturel que les critères d'identification en soient multiples et que des phénomènes d'hésitation, de dissension et de chevauchement apparaissent dans l'analyse. Produit d'une construction non explicite, le mot est, comme le graphème examiné plus haut, hérité par tradition sociale, avec tout l'approximatif et l'arbitraire qui peuvent en découler, le critère de la démarcation graphique restant encore le plus solide.

De fait, la plupart des écritures séparent les mots, par un blanc ou intermot comme en hittite, par un tiret ou une interponction comme en grec mycénien. Cela ne veut pas dire que ce découpage soit universel et inné: les langues anciennes présentent des exemples de *scriptio continua*, et l'écriture devanâgarî qui, avec un haut degré de réalisme phonétique, a servi à consigner le sanskrit, n'utilise le blanc que quand survient une pause du sandhi (= ajustement phonétique de la finale d'un mot à l'initiale du mot suivant). Toutefois, quand il existe, l'intermot peut être pris pour indice de la réalité pragmatique et perceptive que constitue, aux yeux des usagers de la langue concernée, l'entité délimitée. Dès lors, et selon le principe appliqué plus haut à l'atome graphique, il devient possible de soumettre le mot, conçu comme unité naturelle hiérarchiquement supérieure au graphème, à une confrontation avec les unités formelles dégagées par la linguistique.

Dans les langues classiques, les cas de figure de la notion de mot graphique recoupent à quelques détails près les constatations qu'il est possible de faire pour les langues indo-européennes modernes¹¹. En latin par exemple, la pratique éditoriale couramment en vigueur, qui trouve ses sources chez les scribes irlandais et anglo-saxons du début du Moyen Âge, isole régulièrement par des blancs des unités qui relèvent de rangs syntaxiques différents:

1. Mot graphique = morphème (*semper, sic, ut, in...*).

¹¹ Pour plus de détails, voir Berrendonner & Reichler-Béguelin, 1989; Fruyt & Reichler-Béguelin 1990; Fruyt, 1992; [Reichler-]Béguelin, 1992, 1998, 2000.

2. Mot graphique = "syntagme étroit" composé de plusieurs morphèmes: forme fléchiée (*regem, oratorem, amabo*), dérivé (*orator*), composé (*suouetaurilia*), agglutiné (*agricultura*; mais voir sous 3), enfin groupe intonatif composé d'un tonique suivi de clitique(s) (*semperque, uenisne*).

3. Mot graphique = constituant de locution plus ou moins figée ou agglutinée, écrite parfois d'un seul tenant (*res publica/ respublica, agri cultura/ agricultura, bene dico/ benedico*), mais aussi souvent discontinu: *prius [...] quam* (on trouve aussi, en continu et en un seul mot graphique, *priusquam*).

4. Mot graphique = constituant de morphème discontinu: *praesto esse* "être à disposition", où *praesto*, opaque en synchronie au même titre que fr. *fur* dans la séquence *au fur et à mesure*, n'a semble-t-il pas le statut d'unité significative malgré son autonomie graphique et syntaxique.

Au-delà et en deçà du cas trivial où il correspond au signe minimal, le mot graphique latin possède donc une zone de recouvrement qui va du signe complexe (syntagme étroit) au constituant de lexie (i.e. de locution), voire à une séquence infra-morphématique, dépourvue de signifié à elle seule, "monème oblitéré" dans la terminologie de Frei (1962: 132), ou encore "kénomorphème" (cf. Grésillon, 1983: 95). Comme on l'a vu à propos du graphème, l'analyse non théorisée de la chaîne se satisfait donc de la cooccurrence d'unités ne relevant pas d'un même niveau.

2.2.2. Aléas dans l'usage de l'intermot

Quant aux segmentations graphiques non normalisées, en usage dans les documents anciens, elles peuvent différer notablement des standards en usage dans les éditions modernes. Pour une langue comme le grec ancien, qu'on a la chance de connaître à travers trois systèmes graphiques différents (syllabaire mycénien dit linéaire B, syllabaire chypriote, écriture alphabétique), le fait le plus remarquable est que l'articulation en mots est prise en charge de manière relativement conséquente par les syllabaires — où la question des frontières de signes se serait sinon posée de manière cruciale étant donné les phénomènes de concaténation syllabique — mais beaucoup moins par le système quasi phonologique que représente, à ses débuts au moins, l'alphabet adapté du phénicien. Il est trivial de constater que le premier alphabet grec, qui repose sur une analyse de deuxième articulation, a permis, par rapport aux syllabaires, à la fois de réaliser une économie notable sur le nombre de caractères et d'assurer, *grosso modo*, une référence phonique univoque à chaque graphème; mais il est plus rare de relever que ce gain a eu pour contrepartie un coût de décodage accru au niveau de la première articulation, les signaux de division n'étant que d'un usage sporadique et hautement variable autant dans les documents archaïques alphabétiques que dans les manuscrits jusqu'à l'époque alexandrine¹²).

Or, les travaux sur la physiologie de la lecture, ou sur la lecture dans l'Antiquité, ont mis en lumière les conséquences cognitives et les incidences sociales importantes de la *scriptio continua*¹³. Comme le montrent les expérimentations de laboratoire, lire un texte phonographique en continu, c'est d'abord être confronté à la tâche, intellectuellement coûteuse, d'identifier les limites de signes et de lever certaines

¹² Il en allait de même pour l'onciale et la minuscule latines; cf. Desbordes, 1990: 228

¹³ Cf. Saenger; 1989; Canfora; 1989; Marrou, 1965.

équivoques (par exemple, la séquence latine *incultoloco* peut se lire aussi bien *in culto loco* que *inculto loco*; "dans un lieu cultivé" ou "dans un lieu inculte"). On peut comprendre dès lors qu'en grec classique, "lire" se soit dit *anagnó:skein*, soit "reconnaître", "discerner". La lecture a longtemps été une tâche de spécialiste nécessitant le détour par une déclamation expressive ou par une activité préliminaire de vocalisation. L'utilisation d'une "ponctuation de mot" désambiguïsante (hyphen, diastole), puis la généralisation de l'accentuation de mots dans les textes grecs à partir d'Aristophane de Byzance, n'ont que tardivement facilité la technique de lecture des textes en permettant un accès plus direct au sens.

Dans la tradition graphique romaine, la ponctuation de mots, déjà connue de l'étrusque, a été beaucoup plus nettement systématisée qu'en grec, dans les inscriptions officielles de l'âge classique tout au moins. Otha Wingo (1972: 15) montre dans son étude que maints documents officiels ou littéraires soignés présentaient, outre les divisions de mots, une ponctuation de phrases et de groupes syntaxiques. Un très net déclin des séparateurs de mot, vers le 2^e siècle, est toutefois déploré par les épigraphistes, et attribué au désir d'imiter les livres grecs. À date classique, nous ferons l'hypothèse que la multiplication des abréviations en usage dans les documents officiels a favorisé la généralisation, au moins temporaire, de l'emploi des diviseurs en latin: elle rendait en effet nécessaire un marquage explicite de l'articulation en mots pour permettre au lecteur de distinguer d'emblée, dans la séquence de caractères, les symboles acrophoniques, valant pour un signe, des simples phonogrammes.

Si l'on veut bien admettre que les pratiques segmentatives de l'écrit renseignent sur les stratégies d'analyse spontanées des locuteurs, on peut distinguer deux tendances récurrentes dans les documents anciens qui présentent des signaux de division au niveau du mot:

I. La propension à démarquer les *groupes intonatifs* plutôt que les entités lexicales.

II. La propension à *faire interférer la segmentation syllabique avec la segmentation morphologique*.

La première tendance est sensible dans tous les types d'écriture qui font usage de l'intermot. Ainsi, beaucoup d'inscriptions grecques alphabétiques opèrent une sorte d'analyse en constituants syntaxico-prosodiques où l'entité-mot, de rang inférieur, ne se trouve délimitée que dans les occasions où elle se trouve coïncider avec un de ces constituants. Voici l'exemple d'une inscription de Locride (milieu du 5^e siècle), reproduite ci-dessous d'après la translittération de Guarducci, 1967: 301 (le diviseur de "mots" est signalé par une barre; le diviseur de fin de ligne, les blancs et l'accentuation ont été rajoutés). On remarquera la démarcation par le diviseur du nom propre sujet *Eúphamos*, et des groupes intonatifs *kaí toì sundamiorgoí* et *tô:i hé:ro:i*; à noter de surcroît le caractère morphologique, dans cette inscription, des coupes de fin de ligne, qui interviennent l'une après le préfixe *sun*, et l'autre après le verbe *anéthe:kan*, qui s'en trouvent également démarqués graphiquement:

Eúphamos / kaí toì sun-

damiorgoí / anéthe:kan
tô:i hé:ro:i
"Euphamos et ses compagnons
de démiurgie ont dressé
au héros".

Dans les textes alphabétiques de cette époque, il n'est pas exclu que certains raffinements de segmentation soient dus à un autre phénomène, d'ordre sémantique plutôt que phono- ou intonographique. Ainsi, dans une inscription d'Olympie de la seconde moitié du 6^e siècle, consignait un traité d'alliance (Guarducci, 1967: 502-503), les diviseurs n'isolent que des groupes syntaxiques comprenant plusieurs "mots" au sens actuel, à l'exception de deux unités lexicales où se trouve concrétisé un acte d'avertissement: *apotínoian* "qu'ils paient une amende (i.e. ceux qui ne respecteraient pas le traité)", *latreiómenon* "à titre de réparation" (Fruyt & Reichler-Béguelin, 1988). Tout se passe donc, en l'occurrence, comme si la démarcation graphique servait d'auxiliaire à une sorte de mise en relief informationnelle.

Bien des siècles plus tôt, le linéaire B, qui fait pourtant un usage régulier des signaux de division, tend aussi à démarquer des groupes accentuels plutôt que des entités lexicales. Dans ce système d'écriture à fondement syllabique, certains caractères ont conservé une valeur idéographique (ainsi myc. DA = "surface de terre cultivée"); d'autres sont par ailleurs, comme on l'a vu ci-dessus, utilisés en valeur acrophonique. Or, en partie par intonographie, en partie pour empêcher l'interprétation idéographique dont serait susceptible un graphème isolé, on n'écrit jamais séparément les monosyllabes (le plus souvent clitiques): ils sont reliés graphiquement au mot qui précède ou qui suit. S'il s'agit de monosyllabes toniques, ils peuvent être isolés à condition d'être artificiellement étoffés: ainsi *qo-o* (nom du boeuf, deux syllabogrammes) pour /g^wo:ns/ (une seule syllabe). Il s'ensuit qu'en écriture mycénienne, "un mot proclitique *peut* se trouver joint au mot suivant, mais "un enclitique est *toujours* joint au mot précédent" (M. Lejeune, 1958: 23, je souligne). Cette dissymétrie, qu'on retrouve d'ailleurs dans le système syllabique chypriote, est sans doute à mettre en rapport avec la constatation de Martinet, 1966: 43-44, concernant le statut respectif de ce qu'il appelle les "satellites postposés" et des "satellites préposés" dans les usages graphiques des langues indo-européennes. De manière historiquement récurrente, celles-ci tendent en effet à souder les premiers: par exemple, roumain *nasul* "le nez", sur *nas* "nez", mais à préserver plutôt l'autonomie des seconds: fr. *le nez*; en deux mots¹⁴.

En grec toujours, mais cette fois-ci dans le syllabaire chypriote, l'article connaît un rattachement inconstant au nom qui le suit. Or on a constaté que la répartition des formes soudées et non soudées dépendait en partie de facteurs phonétiques. Ainsi, on écrit *o-sa-ta-si-no-i-ko-ne* = *o Stasiwoíko:n* (en cas de soudure, il y aurait **o-so-ta...*), et *ta-se-a-te-na-se* = *tâ:s Athá:na:s*, avec une séparation signalée par l'emploi du syllabogramme *se*, spécialisé à cet effet. En revanche, le traitement est différent dans le cas de *n* final, qui est toujours relié à l'initiale du mot suivant: *ta-*

¹⁴ En l'occurrence, l'autonomie graphique de l'article préposé n'est pas arbitraire: elle va de pair avec la possibilité syntaxique d'insérer un adjectif (*le long nez*), qui n'existe pas quand l'article est postposé.

na-ta-ne = *tà:n Athá:na:n* (jamais **ta-ne-a-ta-ne*), *to-ni-ja-te-ra-ne* = *tòn ijatê:ran* (jamais **to-ne-i-ja-te-ra-ne*). Hermann, 1906, a montré que cette situation résultait d'une variation dans la frontière syllabique, qui passait avant l'*n* final de l'article, mais avant l'*s* : devant voyelle, position où il se trouvait fragilisé, celui-ci était en effet prononcé par hypercorrection, ce qui entraînait une rupture dans la concaténation syllabique.

Nous tenons ici un exemple de notre second phénomène caractéristique: l'interférence qui peut exister dans l'usage entre une segmentation de type morphologique, fondée sur un critère sémiotique (qui aboutit à la séparation de l'article), et une segmentation prioritairement syllabique (qui a ici pour effet sa soudure). La conscience de la syllabe, dont on connaît le rôle en métrique, explique probablement aussi l'apparition de phénomènes de division à l'intérieur du mot dans de nombreuses inscriptions latines (Dennison, 1906), dont les lapicides sont d'ailleurs jugés sévèrement par les épigraphistes: "...on voit [...] jusqu'à quel point certains lapicides avaient perdu la notion de la valeur primitive des signes séparatifs; ce n'était plus, à leurs yeux, que des éléments d'ornementation qui n'avaient même pas le mérite de l'élégance." Ces propos de Cagnat, 1964, s'appliquent à l'exemple de *C.I.L. VIII 4919*, inscription portant les noms propres *Papiria* et *Aufidianus*, et utilisant les diviseurs tantôt comme indices de frontières de syllabe, tantôt comme frontière entre deux mots abrégés (par exemple *P.V.* = *pie uixit*, *H.S.E* = *hic situs est*):

PA/PI/RI/A/A/U/F
I/DI/A/NUS/P/V/
AN/XVIII/H/S/E

L'histoire de l'orthographe, ainsi que les élèves apprenant l'orthographe, fournissent de nombreux exemples où le découpage proposé par les scripteurs révèle un conflit entre critère sémiologique, critère intonatif et critère syllabique. Pour illustrer cette dernière situation, revenons au cas des messages SMS en graphie non standard:

PQ, C DEJA LEUR = [pourquoi, c'est déjà l'heure]
TA PT 1 PLOMB OU Q? = [t'as péte un plomb¹⁵ ou quoi?]

On y observera cette fois-ci la soudure des articles et pronoms élidés (LEUR, TA), qui a pour effet d'ajuster, au niveau d'analyse considéré, l'unité graphique à l'unité syllabique, comme dans les graphies enfantines très courantes du type *da cor* (7 ans), *il ma lu listoire* (8 ans).

3. Pertinence linguistique vs pertinence utilitaire

Depuis le début du XXe siècle, beaucoup de linguistes se sont donné pour but avoué de rendre compte de la conscience ou de l'intuition des sujets parlants. En réalité, il semble plutôt que la linguistique ait cherché, avec plus ou moins de constance et de succès, à se débarrasser de certaines catégories transmises par la tradition, quitte à s'interroger périodiquement sur les raisons du rôle qu'elles

¹⁵ L'expression du fr. familier *péter les plombs* signifie "perdre son sang-froid".

semblent occuper dans l'activité épilinguistique des locuteurs. Ainsi en est-il de la notion peu homogène de mot, qui, dans un certain nombre de langues, se prête pourtant mieux que celle (univoque et scientifiquement fondée) de morphème, à fonctionner, *dans l'usage courant*, en tant qu'unité lexicale et graphique au niveau de la première articulation (cf. Galand, 1992 pour le cas du berbère).

Certes, depuis Saussure, la fameuse "intuition du locuteur naïf" serve d'argument privilégié dès que la validation d'une théorie est en jeu. Pourtant, les linguistes ont mis l'essentiel de leurs efforts à construire sur la langue des discours d'experts, distincts du savoir commun autant par l'ampleur du matériau recruté que par les contraintes imposées aux appareils descriptifs. Sans remettre en cause l'entreprise de théorisation ni contester sa nécessité, on peut relever qu'en elle-même, elle nous apprend peu sur les procédures d'analyse spontanément mises en oeuvre par les locuteurs, sur l'élaboration de leur métalinguistique intuitive.

Or, l'exploration des systèmes d'écriture permet d'éclairer quelque peu les mécanismes cognitifs qui orientent, en fonction des buts poursuivis, les analyses de la chaîne langagière. Une telle étude implique toutefois une certaine révision des objectifs de la modélisation linguistique. Celle-ci ne doit plus être cantonnée à la langue *stricto sensu* et mise au service d'une norme unificatrice, mais elle doit étendre son champ d'étude au comportement langagier lui-même, ainsi qu'aux ambiguïtés d'analyse et au flou catégoriel spécifiques du raisonnement quotidien.

La chaîne langagière se manifeste aux sujets parlants comme un continuum sur lequel ils perçoivent des discontinuités, à plusieurs niveaux hiérarchiques de structuration. L'émergence des micro- et des macro-unités relève d'opérations complexes, dont le résultat donne lieu à variations en fonction des points de vue mis en jeu. Telle qu'elle est opérée par les sujets parlants, la segmentation résulte généralement d'un compromis entre l'application de plusieurs critères: ainsi, comme on vient de le voir, l'usage spontané des séparateurs de "mots" à l'écrit témoigne d'interférences entre les critères d'analyse sémiologique, prosodique, articulatoire. Or, si l'on rapproche le fonctionnement du mot tel qu'il nous a été livré de celui d'autres unités "pratiques" comme le graphème ou la phrase graphique, un isomorphisme de structuration se fait jour. Rapportées aux entités théoriques construites par les linguistes, les unités langagières naturelles ont en effet pour spécificité de représenter la structure à deux ou plusieurs niveaux d'articulation théoriquement distincts, un même matériau servant à des fins diverses. Ce phénomène est à mettre au compte d'une logique utilitaire, proche du bricolage, qui construit et adapte ses unités en fonction de pertinences opératoires multiples. Les chevauchements constatés sont donc intrinsèquement liés à l'élaboration spontanée des outillages graphiques, dont ils garantissent l'efficacité et les qualités d'adaptation.

Pour compléter cette étude, il resterait à envisager le traitement, au rang supérieur de la structure, de l'unité discursive minimale. Qu'on la nomme *phrase, clause, énonciation, intervention, tour de parole, mouvement périodique*, etc..., le statut théorique n'en a été problématisé que très tardivement. Et si l'analyse en unités discursives minimales donne lieu aujourd'hui à quantité de controverses, c'est

qu'elle est la plupart du temps conduite sans qu'on ait pris la peine, au préalable, de clarifier le statut du "point final" délimitant la phrase graphique.

Bibliographie

- E. Alarcos Llorach, « Les représentations graphiques du langage », in *Le langage*, éd. par A. Martinet, Paris, Gallimard, 1973.
- J. Anis (éd.), « Le signifiant graphique », *Langue française* 59, sept. 1983.
- J. Anis, « La construction du graphème et ses enjeux théoriques », *Archives et documents de la Société d'Histoire et d'Epistémologie des Sciences du Langage* 5, 1984, 1-45.
- J. Anis, *L'écriture. Théories et descriptions*, avec la coll. de J.-L. Chiss & C. Puech, Bruxelles, De Boeck, 1988.
- M. Arrivé, *Réformer l'orthographe?*, Paris, PUF, 1993.
- M. Arrivé, F. Gadet, M. Galmiche, *La grammaire d'aujourd'hui: guide alphabétique de linguistique française*, Paris, Flammarion, 1973.
- S. Auroux (éd.), *Histoire des idées linguistiques*, 3 tomes, Bruxelles, Mardaga, 1990-2000.
- M. Baratin & F. Desbordes, *L'analyse linguistique dans l'Antiquité classique*, Paris, Klincksieck, 1981.
- M.-J. Béguelin, « Le rapport écrit-oral. Tendances dissimilatrices, tendances assimilatrices », *Cahiers de linguistique française* 20, 1998, 229-253.
- M.-J. Béguelin (sous la direction de), *De la phrase aux énoncés. Grammaire scolaire et descriptions linguistiques*, Bruxelles, De Boeck-Duculot, 2000.
- A. Berrendonner, « La phrase et les articulations du discours », *Le français dans le monde, Recherches et applications*, février-mars 1993, 20-26
- A. Berrendonner, « Périodes », in H. Parret (éd.), *Temps et discours*, Louvain, Univ. Press, 1993 (= 1993a), 47-61
- A. Berrendonner & M.-J. [Reichler-]Béguelin, « Décalages. Les niveaux de l'analyse linguistique », *Langue française* 81, février 1989, 99-125.
- C. Blanche-Benveniste & A. Chervel, *L'orthographe*, Paris, Maspéro, 1969.
- R. Cagnat, *Cours d'épigraphie latine*⁴, Roma, « L'Erma » di Bretschneider, 1964.
- L. Canfora, « Lire à Athènes et à Rome », *Annales ESC*, juillet-août 1989, n° 4, 925-937.
- N. Catach (éd.), *Pour une théorie de la langue écrite*, Paris, Editions du CNRS, 1988.
- Ch. Chouvet, *Cours et exercices gradués de sténographie*¹², Marseille, Ed. Ferrand, 1953
- F. Desbordes, « Ecriture et ambiguïté d'après les textes théoriques latins », *Modèles linguistiques* V/2, 1983, 13-37.
- F. Desbordes, *Idées romaines sur l'écriture*, Presses Universitaires de Lille, 1990.
- W. Dennison, « Syllabification in Latin Inscriptions », *Classical Philology* 1, 1906, 47-68.
- J. Derrida, *De la grammatologie*, Paris, Minuit, 1967.
- M. Détienne (éd.), *Les savoirs de l'écriture en Grèce ancienne*, Presses Universitaires de Lille, 1988.
- H. Frei, « L'unité linguistique complexe », *Lingua* XI, 1962, 128-140.
- J. Friedrich, *Elementarbuch der hethitischen Sprache* I², Heidelberg, 1960.

- M. Fruyt, « La délimitation des unités lexicales en latin », *Lalies* 10, 1992, 197-204.
- M. Fruyt & M.-J. Reichler-Béguelin, « La notion de mot en latin et dans d'autres langues indo-européennes anciennes », Colloque de Paris IV sur le Mot, novembre 1988, *Modèles linguistiques* 12/1, juin 1990, 21-46.
- M. Fruyt & M.-J. Reichler-Béguelin, « Le mot », atelier animé lors des sessions de linguistique et de littérature d'Aussois, *Lalies* 10, 1992, 113-214
- L. Galand, « Le problème du mot en berbère », *Lalies* 10, 1992, 197-204.
- I. J. Gelb, *A Study of Writing*³, London and Chicago, The University Chicago Press, 1963.
- J. Goody, *La raison graphique. La domestication de la pensée sauvage*, Paris, Minuit, 1979 (éd. anglaise: 1977).
- J. Goody, *La logique de l'écriture*, Paris, Armand Colin, 1986.
- A. Grésillon, « Mi-fugue mi-raison. Dévaliser les mots-valises », *DRLAV* 29, Université de Paris 8, 1983.
- M. Guarducci, *Epigrafia greca* I, Roma, Istituto poligrafico e zecca dello Stato, 1967 (ristampa 1995).
- E. Hermann, « Zur kyprischen Silbenschrift », *Indogermanische Forschungen* 19, 1906, 240-249.
- H. Kronasser, *Etymologie der hethitischen Sprache* I, Wiesbaden, O. Harrassowitz, 1966.
- M. Lejeune, *Mémoires de philologie mycénienne*, 1^e série, Paris, Centre national de la recherche scientifique, 1958.
- H. Lüdtke, "Zur Problematik des Wortbegriffs », in *Studia Linguistica Diachronica et Synchronica (...)* W. Winter sexagenario, U. Pieper et G. Stickel (éds), Berlin, W. de Gruyter, 1985, 527-536.
- H. Marrou, *Histoire de l'éducation dans l'antiquité* ⁶, Paris, Ed. du Seuil, 1965 (réimpr. 1981).
- A. Martinet, « Le mot », in *Problèmes du langage*, Paris, Gallimard, 1966, 39-53.
- A. Martinet, « Les choix du locuteur », *Revue philosophique*, t. CLVI, 1966 (= 1966a), 271-282.
- O. Masson, *Les inscriptions chypriotes syllabiques*, éd. augmentée, Paris, E. de Boccard, 1983.
- C. Metz, « Remarque sur le mot et sur le chiffre », *La Linguistique* 1967/2, 41-56.
- E. Otha Wingo, *Latin Punctuation in the Classical Age*, La Haye-Paris, Mouton, 1972.
- L. J. Prieto, *Etudes de linguistique et de sémiologie générales*, Genève-Paris, Droz, 1975.
- L. J. Prieto, "Lingue grafiche e lingue foniche », in C. Vallini (éd.), *Scrittura e scrittura. Seminario interdisciplinare su teoria e prassi della scrittura*, Napoli, Istituto Universitario Orientale, 1983, 5-20.
- M.-J. [Reichler-]Béguelin, « Perception du mot graphique dans quelques systèmes syllabiques et alphabétiques », *Lalies* 10, Actes des sessions de linguistique et de littérature d'Aussois, Publications de la Sorbonne nouvelle, 1992, 143-158
- S. Reinach, *Epigraphie grecque*, Paris, E. Leroux, 1885.
- P. Saenger, « Physiologie de la lecture et séparation des mots », *Annales ESC*, juillet-août 1989, n^o 4, 939-952.

- F. de Saussure, *Cours de linguistique générale*, éd. par C. Bally et A. Sechehaye, Paris, Payot 1916 (éd. critique par T. De Mauro, 1972; éd. critique et synoptique par R. Engler, Wiesbaden, Harrassowitz, dès 1967).
- J.-P. Seguin, *L'invention de la phrase au XVIIIe siècle : contribution à l'histoire du sentiment linguistique*, Louvain, Peeters ; Paris, Société pour l'information grammaticale.
- H. Seiler, « Zum Verhältnis von Wort und Satz in indogermanischen Sprachen », in *II. Fachtagung für indogermanische und allgemeine Sprachwissenschaft*, Innsbrucker Beiträge zur Kulturwissenschaft, 1962, 163-174.
- E. Schwyzer, *Griechische Grammatik I*, München, C.H. Beck'sche Verlagsbuchhandlung, 1939.
- J. Vachek, « Contribution au problème de la langue écrite », article de 1939 traduit par J. Caussat et paru dans *LINX* 12, 1985, 7-23, avec une introduction de J.-L. Chiss & C. Puech.