

Etude de graphonomie appliquée à un texte du moyen français

par

Lise Lorentzen

1. Introduction

L'analyse graphémique décrite dans le présent article constitue, surtout pour les premières phases, un travail entrepris en commun par Lars Otto Grundt et moi-même.

Notre objectif a été d'examiner le système graphémique d'un texte scientifique du XV^e siècle, le *Traité de cosmographie*, écrit vers 1432 par l'astronome français Jean Fusoris (ms. B.N.fr. 9558). Paléographiquement, nous avons affaire à un manuscrit écrit en batarde, à longues lignes. Il s'agit vraisemblablement, non du manuscrit original de l'auteur, originaire de Reims, mais d'une copie, faite à Metz. L. O. Grundt a fait une édition préliminaire de ce manuscrit, en se servant des règles classiques (codifiées par P. Meyer, M. Roques et M. C. Brunel). Pour plus de détails, voir *Jean Fusoris, Traité de cosmographie, MS. B.N.fr. 9558*. Edition préliminaire par Lars Otto Grundt, Bergen, 1973, p. XV.

Nombreuses sont les discussions sur la valeur respective des deux codes du langage, l'oral et l'écrit, et sur leur interdépendance. Nous sommes d'avis que l'écrit constitue un système autonome en soi, et nous trouvons légitime et souhaitable de consacrer une étude à ce code seul, bien qu'une comparaison entre le système graphémique et un supposé système phonémique basé sur le même texte soit sûrement intéressant¹.

A notre connaissance, personne n'a abordé, jusqu'à présent, le sujet des graphèmes en moyen français; en fait, très peu de chercheurs se sont consacrés aux problèmes graphémiques en français.

Le projet de recherche s'est divisé en deux:

I. une partie, entreprise par L. O. Grundt, qui a consisté en une

1: Cf. par exemple Bjarne Westring Christensen (1964), où l'auteur a élaboré une étude comparative bien fondée des deux codes pour le français moderne.

analyse paléographique et un classement des traits graphiques distinctifs (mérismes) figurant dans le manuscrit²;

2. une partie où nous avons pris comme point de départ une version modifiée de l'édition préliminaire pour effectuer d'abord un classement des graphes (à l'aide de critères non-phonologiques), ensuite l'établissement d'un inventaire des graphèmes et une étude graphotaxique portant sur la distribution des graphèmes à l'intérieur des mots.

Pour le classement des graphes et l'étude de la graphotaxe, nous nous sommes servis d'un ordinateur. Sans ce moyen, nous n'aurions probablement pas été à même d'accomplir ces tâches; pour la graphotaxe au moins, cela est évident.

C'est cette deuxième partie du projet de recherche qui sera décrite dans les pages qui suivent.

2. Quelques définitions

Avant d'entrer dans le détail, précisons quelques points de terminologie. Nous venons de mentionner les termes *graphe*, *graphème* et *mot*.

Par *graphe*, nous comprenons le représentant des occurrences graphiques à forme identique telles qu'elles apparaissent dans le texte écrit à la machine.

A ce propos, nous nous voyons obligée d'introduire les termes techniques *type* et *token*. Ces deux termes, qui sont surtout employés dans la terminologie de la statistique lexicale, sont pertinents tant pour les graphes que pour les mots. Comme le dit très bien Lyons (1977, p. 14):

Tokens are unique physical entities, located at a particular place in space or time. They are identified as tokens of the same type by virtue of their similarity with other unique physical entities and by virtue of their conformity to the type that they instantiate.

Pour en revenir à notre terme *graphe*, il faut le considérer comme un graphe-type (défini comme un représentant de toutes les occurrences identiques) et non pas comme un graphe-token (défini comme une occurrence concrète). Ajoutons que l'écart entre le nombre de graphes-types et le nombre de graphes-tokens est considérable: notre corpus contient 22 graphes-types mais 54.400 graphes-tokens. Les

²: Voir Grundt et Lorentzen (1977), 1^{re} partie.

graphes sont tous candidats au statut de graphèmes, ce sont en quelque sorte des graphèmes provisoires ou des prégraphèmes.

La terminologie dans ce domaine varie souvent d'un auteur à l'autre. Ainsi, notre graphe correspond aux termes anglais «graph-class» (cf. McLaughlin (1963), p. 29; Crossland (1957), p. 15) et «graph-type» (cf. Francis (1962), p. 35) et au terme suédois «graftyp» (cf. Allén (1965), p. 80)³.

Les *graphèmes* sont les unités distinctives les plus petites qui ressortent d'une analyse interne de la langue écrite⁴. La reconnaissance des graphèmes, à la différence des allographes (voir infra) d'un même graphème, reposera sur l'étude des oppositions fonctionnellement distinctives vs. celles qui sont fonctionnellement non-distinctives. Pour dégager les graphèmes, il faut surtout faire appel aux relations paradigmatiques dans lesquelles entrent les graphes.

Nous parlerons d'*allographes* pour désigner deux ou plusieurs graphes qui, après analyse syntagmatique et paradigmatique, ne présentent aucune opposition distinctive entre eux. Les allographes sont donc des variantes – libres ou contextuelles – d'un même graphème.

Au besoin nous lierons les préfixes *auto-* et *syn-* aux deux termes graphe et graphème. Pour éviter un parallélisme poussé avec le système parlé, nous préférons le terme *autographème* à «graphème vocalique»⁵ ou «voyelle graphématique»⁶, ainsi que *syngraphème* à «graphème consonantique»⁵ ou «consonne graphématique»⁶. Nous reviendrons aux définitions des termes *autographe*, *autographème*, *syngraphe* et *syngraphème*.

Qu'est-ce qu'un mot? Le lecteur reconnaîtra sans doute le titre de l'article célèbre de Togeby. Or, malgré l'intérêt incontestable que cette question implique, nous n'avons pas l'intention d'entrer dans cette discussion sur le plan théorique.

Parallèlement à ce qui est le cas pour le graphe, le terme *mot* sera surtout employé dans le sens de mot-type. Dans les quelques cas où le mot est à comprendre comme un mot-token, nous le spécifierons explicitement.

Nous définissons le *mot* comme le mot-type graphique, c'est-à-dire la

3: Ces deux derniers ne doivent pas être confondus avec notre terme *graphe-type*.

4: Cf. par exemple Sture Allén (1965), p. 36-37.

5: Piirainen (1968, p. 36) parle de «vokalische Graphematik» et de «konsonantische Graphematik».

6: Termes utilisés par Westring Christensen (1967, p. 224).

chaîne de graphes comprise entre deux blancs ou entre un blanc et un signe non-alphabétique (de ponctuation).

Ceci nous conduit à dire quelques mots sur le statut du mot graphique dans nos matériaux. Nous avons gardé les délimitations des mots comme elles figurent dans le manuscrit. Ni accents ni apostrophes n'y sont employés. Ce qu'on écrirait aujourd'hui comme *l'année* – et où l'on compterait probablement deux mots – se trouve écrit en un mot: *lannee*.

La plupart des linguistes qui se sont occupés de questions graphémiques et plus particulièrement de questions de graphotaxe, ont surtout examiné les mots «monovocaliques», comme ils sont couramment appelés. (Cf. par exemple Spang-Hanssen (1959) et Westring Christensen (1964).) Sans recours à des outils automatiques, cela semble bien compréhensible. Or, disposant d'un ordinateur, nous avons trouvé opportun d'examiner tous les mots du corpus, qu'ils soient constitués d'un ou de plusieurs noyaux vocaliques (ou autographiques, pour nous en tenir à notre terminologie). Ajoutons toutefois que nous ne négligeons pas les avantages qu'il y aurait à ne traiter que les mots «mono-autographiques». On obtiendrait par là sûrement une plus grande régularité, entre autres choses.

Pourquoi ne pas prendre comme «unité de cadre»⁷ *la syllabe*? La syllabe est en effet considérée par plusieurs⁸ comme l'unité pertinente immédiatement supérieure aux graphèmes et aux phonèmes. Pour la phonologie, le statut de la syllabe nous semble clair et évident, mais cela n'est pas le cas en parlant de graphonomie. Qu'une syllabe soit ouverte ou fermée, atone ou tonique, etc., doit jouer un rôle capital dans la phonologie, mais ces notions ne s'appliquent pas de manière convaincante à la graphonomie. Comme le dit Einar Haugen (1956, p. 215):

If syllables exist, they must be defined independently of the morpheme, *in terms of purely phonological criteria*. (C'est nous qui soulignons.)

Pour compliquer encore la situation, les phonologues ont souvent du mal à se mettre d'accord en ce qui concerne la délimitation des syllabes. Il est facile de les compter à partir de leurs sommets vocaliques, mais il est beaucoup plus compliqué d'établir les frontières syllabiques. (Voir

7: L'expression a été empruntée à Troubetzkoy (1970, p. 270).

8: Cf. entre autres Westring Christensen (1964, p. 2-3), d'après Spang-Hanssen (1959, p. 122-124).

par exemple Haugen, 1956.) Selon Kurylowicz (1948, p. 80), et nous ne contestons nullement son affirmation,

La limite syllabique se place toujours devant le complexe» (c.-à-d. le complexe consonantique intermédiaire) «ou à son intérieur, jamais après le complexe, c.-à-d. immédiatement devant le centre syllabique.

Dans un exemple comme le mot *cristal*, on aurait donc deux découpages possibles, soit *cri + stal*, soit *cris + tal*, tandis que *crist + al* serait exclu. Mais en quoi est-ce pertinent pour notre analyse graphémique? *st* figure en effet comme séquence finale dans plusieurs mots et *al* peut figurer à l'initiale de mot. Dans un cas comme celui-ci, nous préférons laisser ouverte la question de l'appartenance de la séquence intermédiaire, et nous ne voyons pas en quoi le concept de syllabe serait utile.

Evidemment, dans les travaux de ceux qui ne s'occupent que de mots «monovocaliques», il y a identité entre mot et syllabe, et dans ces cas nous ne voyons pas l'utilité de la notion de syllabe.

3. Corpus

Donnons maintenant quelques renseignements sur notre corpus, appelé jusqu'ici la version modifiée de l'édition préliminaire. Notre point de départ a donc été cette édition préliminaire de L. O. Grundt, qui a été retouchée de diverses manières, dont quelques détails seront présentés par la suite.

On aurait pu, naturellement, utiliser une version plus proche du manuscrit, à savoir la version dite diplomatique. Cela aurait fourni des résultats quelque peu différents des nôtres, aussi bien en ce qui concerne l'inventaire des graphes et des graphèmes que dans la graphotaxe. Une version diplomatique de notre texte a été gardée sur bande magnétique pour être disponible dans d'éventuelles études à venir.

Nous avons déjà mentionné que les frontières de mot sont celles du manuscrit. On trouve par exemple *il ya* et *la quelle*, où l'on mettrait en français moderne (et aussi dans l'édition de L. O. Grundt) *il y a* et *laquelle*.

Les chiffres romains ont été supprimés pour éviter toute confusion avec des graphes alphabétiques.

Les abréviations du manuscrit ont été décomposées. Chaque occurrence de \tilde{q} en tant que mot graphique, par exemple (où le \sim superposé indique qu'il s'agit d'une abréviation), figure dans notre corpus comme

un *que* «normal». Il va de soi qu'une analyse des différentes abréviations a précédé cette décomposition. Ainsi, le mot *que* se trouve écrit *que* quand il n'est pas abrégé.

Nous avons supprimé tous les signes de ponctuation.

Nous ne distinguons pas minuscules et majuscules. A titre d'exemple, *L* et *l* figurent tous les deux comme un graphe, *l*, ce qui constitue une simplification avantageuse pour le traitement automatique. Cette décision nous empêche donc de considérer *L* et *l* comme deux allographes du même graphème ou comme deux graphèmes distincts. Elle nous empêche également de faire entrer en ligne de compte la possibilité d'opérer avec un «graphème» comparable à un phonème suprasegmental en phonologie, qui transformerait les minuscules en majuscules. (Voir Gleason (1966, p. 410), qui emploie le terme anglais «capitalization» pour ce genre spécifique de graphème.)

Les quelques parties du manuscrit écrites en latin ont été supprimées. Ceci pour n'avoir affaire qu'à un système «graphétique» et graphémique cohérent.

Parmi les modifications qui ont eu un impact substantiel sur l'inventaire des graphes figurent également les suivantes: Ayant constaté que dans le manuscrit *j* ne se trouve qu'en position initiale de mot, où il alterne avec *i* d'une manière non-distinctive, nous les avons ramenés sous un même graphe, *i*. La même chose a été constatée pour la paire *v* / *u*; *v* se trouvant uniquement en position initiale, alternant avec *u* d'une manière non-distinctive, nous les avons ramenés sous un même graphe, *u*.

Ces constatations, au lieu d'être faites a priori, auraient pu (et même dû, peut-être) découler de l'analyse faite par la suite. *i*, *j*, *u* et *v* auraient pu figurer tous les quatre dans l'inventaire des graphes, pour être ensuite réduits aux deux graphèmes <*i*> et <*u*>. (Voir 5.1.)

4. Inventaire et classement fonctionnel des graphes

Nous disposons d'un inventaire de 22 graphes alphabétiques: *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *i*, *l*, *m*, *n*, *o*, *p*, *q*, *r*, *s*, *t*, *u*, *x*, *y*, *z*⁹.

9: Cet inventaire est le résultat d'une réduction des 27 «graphes-types» dégagés par L. O. Grundt dans Grundt et Lorentzen (1977, p. 115-116). Deux facteurs ont influencé cette réduction: d'une part le fait que les trois variantes originelles de *s* et les deux variantes originelles de *r* ont été ramenées aux deux invariante *s* et *r* (voir *ibid.*, p. 116); d'autre part le fait, décrit ci-dessus, que nous avons réduit les paires *i* / *j* et *u* / *v* à *i* et *u* respectivement.

Est-il possible de faire un premier classement fonctionnel de ces graphes sans recours à des critères phonologiques?

Après avoir examiné les mots danois composés de deux graphèmes, Paul Diderichsen et Henning Spang-Hanssen sont arrivés à une répartition en *autographèmes* (éléments présupposés) et *syngraphèmes* (éléments présupposants). (Voir Diderichsen (1952), p. 10 et Spang-Hanssen (1952), p. 23.) Les autographèmes peuvent seuls, en combinaison mutuelle et en combinaison avec des membres de l'autre classe, former des expressions de mot minimales. Les syngraphèmes ne peuvent former des expressions de mot minimales quand ils figurent seuls ou qu'ils se combinent entre eux; ils le peuvent uniquement en combinaison avec un membre de la classe des autographèmes.

Nous nous inspirons de cette répartition pour essayer de l'appliquer, au niveau graphétique, à tous les mots de notre corpus. Après avoir établi la définition des autographes que voici: «Est à définir comme la classe des autographes la catégorie minimale de graphes dont au moins un membre figure dans tout mot graphique», nous avons fait examiner à l'ordinateur tous nos mots. L'ordinateur nous a fourni la liste suivante des graphes remplissant ces conditions: *a, e, i, o, u, y*¹⁰. Ces 6 graphes constituent donc la classe des autographes, et les 16 graphes restants celle des syngraphes.

Comme exemples de mots constitués uniquement d'autographes, on peut citer:

a, y;
au, ie, ya;
eue, iay;
eaue, ioeye.

5. Graphes et graphèmes

5.0. Transition de graphes à graphèmes

Comme nous l'avons déjà constaté, l'orthographe française était bien instable au XV^e siècle, fait dont notre texte offre de nombreux exemples. Il importe de savoir si la nature des variations graphiques est décisive ou non pour la constitution de l'inventaire des graphèmes.

10: Pour les détails de ce procédé, nous renvoyons le lecteur intéressé à notre petit article, «Bruk av datamaskin i en grafemundersøkelse», paru dans *EDB i språk- og litteraturforskning*. Rapport fra en konferanse på Geilo 4. og 5. januar 1979 arrangert av NAVFs EDB-senter for humanistisk forskning, Bergen, p. 99-109.

Contrairement à ce qui est pertinent pour l'analyse phonémique, les traits distinctifs de la substance graphique ne paraissent pas utilisables pour l'établissement de l'inventaire graphémique. (Cf. Grundt dans Grundt et Lorentzen (1977), p. 114.)

Quels sont donc les critères valables? Nous avons examiné les distributions des divers graphes pour voir si les relations syntagmatiques à l'intérieur du mot pouvaient nous aider. Les relations graphotaxiques qu'entretiennent les graphes entre eux, aussi bien à l'intérieur de chacune des deux classes qu'entre les membres des deux classes, semblent trop complexes pour qu'il soit possible de définir les graphèmes sur cette base. Mais en examinant les relations sur le plan paradigmatique, il est évident que la position, initiale, finale et médiane de mot, joue un rôle capital.

5.1. De *i* graphe à <*i*> graphème et de *u* graphe à <*u*> graphème

Rappelons que *i* et *u* ont été classés parmi les autographes. Cela n'empêche pourtant pas qu'ils ont des affinités distributionnelles avec les syngraphes dans quelques contextes précis. Dans des mots comme *il* et *un*, il est hors de doute que *i* et *u* font partie de la classe des autographes, ce qui découle de la définition même des autographes. Leur appartenance à cette classe est également évidente dans les mots où on les trouve après deux syngraphes initiaux incontestables, par exemple dans un mot comme *tristesse*. Mais les choses deviennent plus compliquées si l'on considère des paires minimales comme *ie* / *le* ou *uoz* / *noz*, où il est difficilement imaginable qu'un autre autographe puisse se substituer à *i* ou *u* respectivement. En somme, dans les cas où *i* et *u* figurent devant des autographes incontestables (comme dans *ianuier*, *ie*, *iehan*, *ueons*, *uoit*, *uous*, etc.), leur comportement semble plutôt syngraphique. Le plus juste serait peut-être de leur donner un double statut d'autographes et de syngraphes, où leur statut serait à déterminer dans chaque cas à partir de critères contextuels. Toutefois, puisque leur comportement semble être autographique dans la grande majorité des cas, nous trouvons pratique de continuer à les considérer comme des autographes, mais comme des autographes ayant parfois des emplois syngraphiques.

Mais le fait qu'ils sont à considérer comme des autographes ne leur donne pas automatiquement le statut d'autographèmes. Pour vérifier leur statut de graphèmes, il faut recourir à l'épreuve de commutation.

Cette épreuve montre qu'il y a opposition distinctive entre *i* et les quatre autographes *a*, *e*, *o* et *u* (nous traiterons de la distribution entre *i* et *y* ci-après), et qu'il y a également opposition distinctive entre *u* et les cinq autographes *a*, *e*, *i*, *o* et *y*. (Cf. la liste des paires minimales dressée au paragraphe 5.8.) Dès maintenant, nous parlerons des *graphèmes* <*i*> et <*u*>.

5.2. La variation *i/y*

Il s'est avéré que *i* et *y* ont souvent été employés dans les mêmes contextes graphiques, phénomène attesté dans les trois positions, initiale, finale et médiane.

Ex.: Position initiale : *icelle* / *ycelle*
 Position finale : *aussi* / *aussy*
 Position médiane : *nuit* / *nuyt*

En effet, *i* et *y* ne semblent présenter d'opposition distinctive en aucun contexte. Il y a ainsi de fortes raisons de croire que ce sont des allographes d'un même graphème.

Mais la variation entre *i* et *y* n'est pas toujours libre. *y* est la variante employée dans les cas où il forme un mot à lui seul. En ce sens, ils présentent une complémentarité partielle. Selon les termes de Troubetzkoy (1970, p. 52), il y a une relation combinatoire entre l'emploi de *y* et la variation *i/y*.

Comme nous avons déjà constaté qu'il n'y a pas d'opposition pertinente entre *i* et *j*, il ne peut pas y avoir d'opposition pertinente entre *j* et *y* non plus. Tout porte donc à croire que *i*, *j* et *y* sont des allographes d'un même graphème, que nous désignerons par <*i*>. Cette hypothèse se trouve en effet vérifiée par l'épreuve de commutation, qui montre une opposition distinctive entre <*i*> et les autres candidats au statut d'autographèmes. (Ex.: *mil* / *mal*.)

5.3. La variation *s/z*

En position médiane, et beaucoup plus fréquemment en position finale, nous avons trouvé *s* et *z* dans des contextes par ailleurs identiques.

Ex.: Position finale : *fois* / *foiz*
 bas / *baz*
 Position médiane : *philosophes* / *philozophes*

Leur distribution n'est pourtant pas toujours identique. *z* n'entre pas aussi facilement que *s* dans des groupes syngraphiques; il n'est par exemple jamais employé comme le premier de deux syngraphes consécutifs ni à l'initiale, ni à la finale. En fait, *z* est employé soit immédiatement devant un autographe, soit comme le tout dernier graphe du mot. Il y a donc une relation combinatoire également entre l'emploi de *s* et la variation *s / z*.

L'important, c'est qu'il ne semble pas y avoir de contextes où *s* et *z* soient en opposition distinctive, ce qui nous amène à les considérer comme des allographes d'un même graphème, que nous désignerons par <*s*>. Que <*s*> soit un graphème véritable ressort de l'épreuve de commutation. (Ex.: *me / se*)

5.4. La variation *z/x*

En position finale, *x* et *z* alternent dans les quatre mots suivants:

celestiaulz / celestiaulx
 cieulz / cieulx
 deulz / deulx
 principaulz / principaulx

On remarque que le contexte est assez particulier. Bien qu'il n'y ait pas d'opposition distinctive entre *x* et *z* dans ces cas précis, la variation semble trop peu généralisée pour qu'on puisse les considérer comme deux allographes d'un même graphème. Ce que nous pouvons conclure, néanmoins, c'est qu'il y a neutralisation entre eux après *ul* (ou plus précisément après *aul* et *eul*).

Le statut de *z* comme allographe de <*s*> reste inchangé.

5.5. La variation *s/x*

En position finale, il y a alternance *s / x* dans trois mots différents:

lieus / lieux
 aus / aux
 oyseaus / oyseaux

Ces cas sont assez semblables à ceux que nous venons de mentionner. Ils ne peuvent pas être considérés comme des allographes d'un même graphème. On voit qu'ici il s'agit d'une neutralisation après *u* (ou plus précisément après *eu* et *au*). Dans les cas de ce genre, on a affaire à un recouvrement graphémique partiel des deux graphèmes <*s*> (ayant les

deux allographes *s* et *z*) et <*x*>. Le statut de <*x*> en tant que graphème ressort de l'épreuve de commutation. (Ex.: *fixe* / *fine*)

5.6. Le statut du syngraphe *q*

Il s'est avéré que *q*, sauf dans le mot *cinq*, est toujours suivi de *u*, lui-même suivi d'un autographe. Cf. *quant*, *que*, *quil*, etc. On peut établir des paires minimales comme *tant* / *quant* et *sil* / *quil*, où il semble que ce soit la combinaison *qu* qui commute avec *t* et *s* respectivement; de l'autre côté, on trouve des paires minimales comme *que* / *rue* et *qui* / *lui*, où il semblerait que ce soit *q* seul qui commute avec *r* et *l* respectivement.

Ce qui importe, c'est qu'après *q*, *u* n'est pas commutable ni avec un autographe quelconque, ni avec zéro. *u* n'a donc pas de valeur distinctive dans ce contexte. Le *q*, lui, est commutable avec différents syngraphes.

Le fait que *q* ne figure qu'accompagné de *u* nous amène à conclure que *qu* constitue un graphème unique, <*qu*> se présentant sous son allographe *q* dans le cas du mot *cinq*. *u* garde son statut de graphème autonome dans tous les contextes autres que celui où il est précédé de *q*.

5.7. Variations accidentelles

A part les variations discutées jusqu'ici, le texte contient un certain nombre de variations accidentelles. Une variation accidentelle est à comprendre comme une variation non généralisée qui affecte seulement un nombre très restreint de cas.

Il s'agit souvent de gémiation vs. non-gémiation de syngraphes en position médiane (ex.: *accomplie* / *acomplie*). Dans d'autres cas, il s'agit de variations entre deux graphes différents, surtout en position médiane (ex.: *ardans* / *ardens*) ou entre une suite de graphes et un seul graphe (ex.: *ruisseaux* / *ruisseau(x)*).

Bien qu'il n'y ait pas d'opposition distinctive dans ces cas, il ne nous paraît pas légitime de parler d'allographes puisque, justement, ces variations ne sont pas assez généralisées.

5.8. Inventaire des graphèmes

En ce qui concerne les syngraphes dont le statut graphémique n'a pas été discuté jusqu'ici (*b*, *c*, *d*, *f*, *g*, *h*, *l*, *m*, *n*, *p*, *r* et *t*), il y a raison de

croire qu'ils sont tous des graphèmes. En utilisant l'épreuve de commutation, on peut établir des listes de paires minimales des types suivants:

Position initiale	Position finale	Position médiane
par / car	et / en	fine / fixe
pays / rays	fait / fais	laine / laide
ces / ses	tout / tour	dune / dure
des / tes	par / pas	pure / puce
mon / bon	or / on	pores / poles
hault / fault	nom / non	prime / prise
galer / aler		

Pour chaque paire, les oppositions des graphes sont distinctives, et on a affaire à deux graphèmes différents.

A partir des 16 syngraphes avec lesquels nous avons opéré provisoirement, nous avons maintenant les 15 syngraphèmes suivants: , <c>, <d>, <f>, <g>, <h>, <l>, <m>, <n>, <p>, <qu>, <r>, <s>, <t> et <x>.

Dans le cas des autographes, nous avons déjà discuté la relation graphémique entre *i* et *y*. A l'aide de l'épreuve de commutation, nous avons pu établir des unités de valeur distinctive. Exemples:

an / en / fin / on / un
 la / le / lune / ly
 mal / mil / mol
 mors / murs
 pert / port

L'inventaire des autographèmes semble donc contenir les 5 autographèmes suivants: <a>, <e>, <i> (avec les allographes *y* et *j*), <o> et <u> (avec l'allographe *v*).

6. Les graphèmes

6.0. Définition des graphèmes. Discussions et illustrations

Dans ce qui précède, nous avons tenté de faire une délimitation des graphèmes de notre corpus. Rappelons maintenant une remarque importante faite par Pierre Naert (1961, p. 51):

Tout revient à ne pas confondre délimitation, définition et description. - - - on *délimite* les phonèmes (ou graphèmes) à partir de commutations de mots, on les *définit* à partir de leur distribution dans le mot, et on les *décrit* eux aussi de tous les autres points de vue qui peuvent paraître intéressants. (C'est nous qui soulignons.)

Nous aimerions pouvoir partager ces points de vue de Naert. Pourtant, il ne faut pas oublier qu'en établissant l'inventaire graphémique, nous nous sommes déjà servie du critère de la distribution en quelque sorte. Nous trouvons que Lise Opdahl (1966, p. 51) a raison quand elle dit:

Graphotactical considerations must often be taken into account when graphemes are established and consequently influence this kind of classification. On the other hand the result of a graphotactical description largely depends on how the graphemic inventory has been set up.

Il semble donc que la distinction très nette qu'opère Naert entre délimitation et définition n'est pas aussi absolue qu'il nous le fait croire.

Quoi qu'il en soit, une discussion des critères de définition s'impose. Il est nécessaire d'opérer des choix pour que le comportement caractéristique de chaque graphème ressorte aussi clairement que possible dans un système total.

Au sujet du système des consonnes, Togeby (1965, p. 55) classe les consonnes de la manière suivante: d'abord les consonnes initiales en vocaliques (toujours voisines d'une voyelle) et consonantiques (séparables de la voyelle); ensuite les consonnes initiales-finales en consonantiques et vocaliques; ces dernières sont à leur tour réparties en consonnes incompatibles et en consonnes compatibles avec d'autres consonnes.

Pour les voyelles, il procède, d'une manière semblable, les caractérisant par leurs positions dans le noyau vocalique. (Ibid., p. 58.)

Kurylowicz (1949, p. 56), en parlant du consonantisme grec, avait déjà travaillé dans une direction un peu analogue, distinguant trois classes de consonnes. Citons-le:

Appartient à la classe I un élément consonantique qui peut être précédé, mais non suivi d'une consonne. La classe II est celle des consonnes pouvant être suivies et précédées d'autres consonnes. A la classe III appartiennent les consonnes admettant une consonne suivante, mais non pas précédente.

Nous sommes très en faveur de ce genre de classement fonctionnel, où l'on évite de faire appel à des critères phonétiques ou, plus généralement et en termes glossématiques, où l'on écarte tout recours à

la substance de l'expression. Aussi les graphématiciens ont-ils trop souvent eu tendance à définir les graphèmes sur des bases phonétiques ou phonologiques, cherchant les équivalents des graphèmes dans le système phonétique ou phonologique.

Il nous semble donc préférable d'employer des critères distributionnels, comme ceux employés par Togeby et Kurylowicz entre autres. Or, comment procéder pour analyser la distribution des graphèmes et ainsi arriver à les définir? Comme le dit Bengt Sigurd (1965, p. 10) en parlant de phonotaxe:

Although there are many studies of the distributional patterns of phonemes in various languages – and a considerable number of studies of English phonotactics – there exists no generally recognized method for systematically analysing and describing the phonemic arrangements of a language. The varieties in technique and method make it difficult to compare the results achieved.

Ajoutons entre parenthèses que le fait que nous nous servons souvent de citations qui portent sur la phonotaxe n'implique évidemment pas que nous tombions dans le piège que nous avons surtout voulu éviter, et qui consiste à se laisser influencer par le système phonologique. Mais puisque les œuvres traitant de la distribution des unités fonctionnelles les plus petites de la deuxième articulation portent beaucoup plus souvent sur les phonèmes que sur les graphèmes, il est naturel que nous trouvions là des idées utiles pour n'importe quelle étude distributionnelle.

Comme nous l'avons dit plus haut, nous avons choisi le mot comme le cadre à l'intérieur duquel nous étudions les relations distributionnelles. On aurait pu choisir une unité de cadre différente, par exemple le morphème, ce qui aurait sans doute abouti à des résultats tout aussi intéressants et sûrement différents. Car même si nous écartons des considérations morphologiques de notre étude graphotaxique, il est indéniable que beaucoup des combinaisons graphémiques à l'intérieur des mots sont dues à la concaténation de morphèmes.

En ce qui concerne le procédé à suivre, nous sommes d'accord avec ce que dit Troubetzkoy dans la citation suivante (1970, p. 272):

Il va de soi qu'on devra examiner séparément, d'un côté les positions à l'intérieur des unités de cadre en question (initiale, intérieur du mot, finale), et d'un autre côté les trois formes fondamentales des groupements de phonèmes (groupements de phonèmes vocaliques entre eux, groupements de phonèmes consonantiques entre eux, groupements de phonèmes vocaliques et de phonèmes consonantiques).

C'est en gros selon ces lignes que nous avons travaillé en étudiant la graphotaxe.

A la suite de Troubetzkoy (1970, p. 272-273), nous nous sommes posé les trois questions suivantes pour chacune des deux classes de graphèmes:

1. Quels graphèmes se combinent entre eux?
2. Dans quel ordre se combinent-ils?
3. Quel est le nombre de graphèmes qui font partie d'une combinaison?

Un relevé détaillé de la distribution complète de chaque graphème dépasserait les bornes d'un article comme celui-ci. Esquissons pourtant quelques résultats.

Illustrons cette démarche en exposant d'abord comment elle fonctionne pour les syngraphèmes en position initiale de mot.

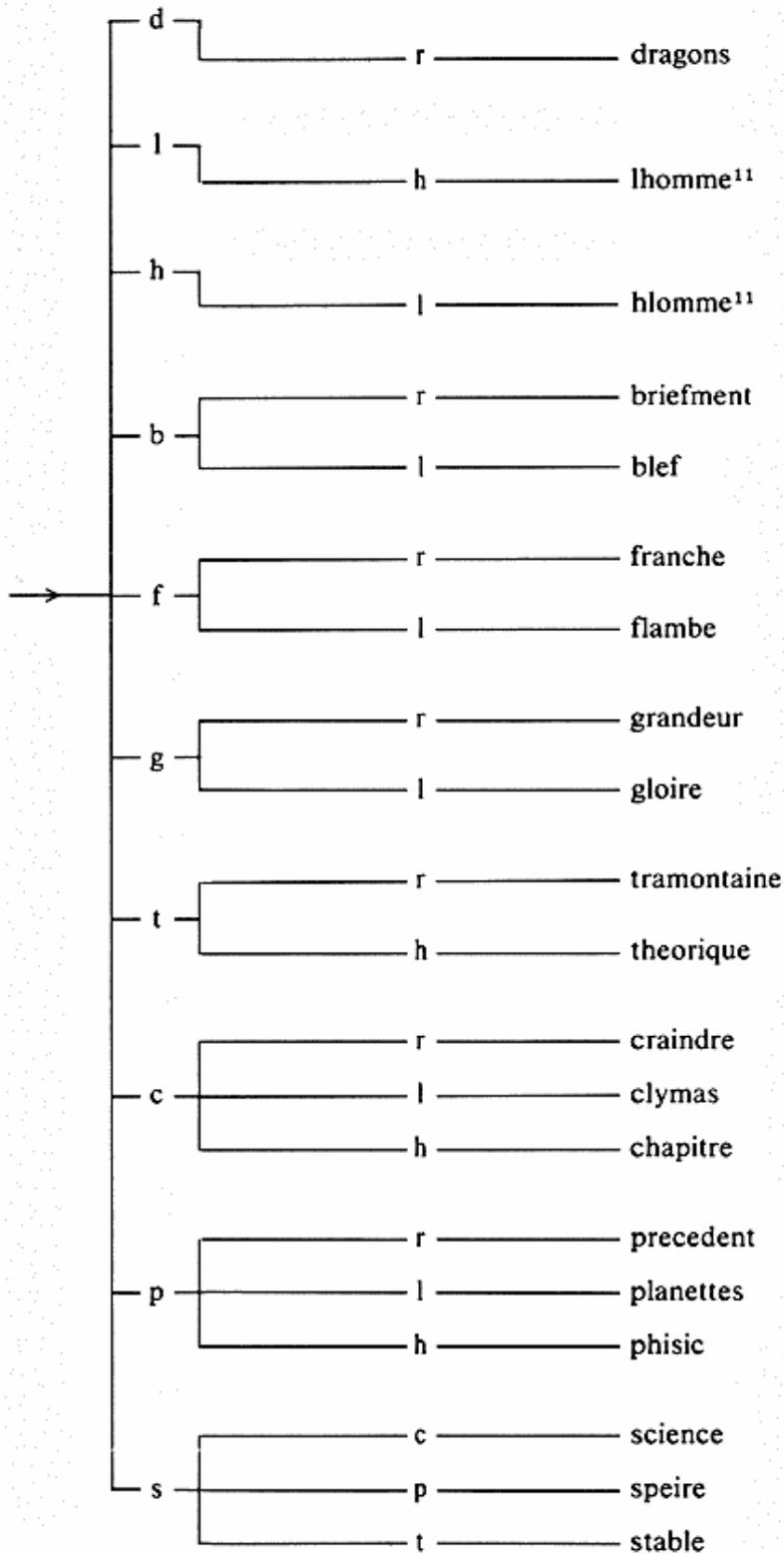
6.1. *Quels graphèmes se combinent entre eux?*

Les différentes combinaisons attestées sont exposées dans le tableau 1 (v. p. 113). Pour la manière de dresser ce tableau, nous nous sommes inspirée de B. Sigurd (1965, p. 42).

11: Les deux groupes <lh> et <hl> ne figurent que dans ces deux mots, et leur fréquence relative par rapport à l'ensemble des groupes de deux syngraphèmes consécutifs à l'initiale est très faible. Par conséquent, il serait préférable d'enlever ces deux combinaisons si nous dressions un tableau statistique; nous les gardons quand même dans notre tableau pour des raisons d'exhaustivité.

Ajoutons qu'intuitivement au moins, il serait naturel de considérer <hl> comme une variante graphique de <lh>, en dépit du fait que la forme *hlomme* est plus répandue dans le texte que la forme *lhomme*.

Tableau 1. Combinaisons entre sygraphèmes en position initiale



Nous tenons à signaler qu'on ne doit pas attacher d'importance au choix des autographèmes qui suivent ces groupes initiaux; il existe en effet des restrictions importantes sur la possibilité qu'ont les différents autographèmes d'apparaître après ces groupes.

Qu'est-ce qui ressort de ce tableau?

En ordre de compatibilité décroissante, et sans tenir compte de l'ordre des syngrophèmes, on constate que

<r>	est compatible avec	les sept syngrophèmes <d>, , <f>, <g>, <t>, <c>, <p>;
<l>	- " -	les six syngrophèmes <h>, , <f>, <g>, <c>, <p>;
<h>	- " -	les quatre syngrophèmes <l>, <t>, <c>, <p>;
<p> et <c>	sont compatibles avec	les quatre syngrophèmes <r>, <l>, <h>, <s>;
<s>	est compatible avec	les trois syngrophèmes <c>, <p>, <t>; ¹²
<t>	- " -	les trois syngrophèmes <r>, <h>, <s>;
, <f> et <g>	sont compatibles avec	les deux syngrophèmes <r> et <l>;
<d>	est compatible avec	un seul syngrophème, <r>.

Les quatre syngrophèmes <m>, <n>, <qu> et <x> n'entrent pas dans des combinaisons à l'initiale.

6.2. Dans quel ordre se combinent ces syngrophèmes?

<r> n'est jamais le premier élément de ces groupes; <d>, , <g>, <p> et <s>, au contraire, sont toujours les premiers éléments. À côté de <l> et <h>, qui sont les premiers éléments seulement dans les deux mots cités et où ils se suivent ou se succèdent l'un à l'autre mutuellement (voir la note 11), <c>, <p> et <t> figurent et comme les premiers et comme les derniers éléments.

12: L'allographe z du graphème <s> n'entre pas dans des combinaisons de syngrophèmes à l'initiale.

6.3. *Quel est le nombre de graphèmes qui font partie d'une combinaison?*

Comme il ressort du tableau 1, tous les groupes attestés en position initiale sont constitués de deux syngraphèmes. (Les groupes *pht* et *pth* ont été relevés dans les trois mots *ptholomee*, *ptholomeus* et *ptholomee*, mais puisqu'il s'agit de noms propres grecs, nous les avons écartés comme n'étant pas inclus dans le système français.)

Nous avons soumis les combinaisons de syngraphèmes à la finale et à l'intérieur de mot aux mêmes genres d'examen que ceux décrits ci-dessus, et nous avons procédé de façon semblable pour examiner la distribution des autographèmes. La description des différentes caractéristiques distributionnelles dépassent cependant les bornes de cet article, et nous nous voyons obligée de les laisser de côté.

6.4. *Combinaisons de syngraphèmes et d'autographèmes*

Considérons maintenant les combinaisons de syngraphèmes et d'autographèmes.

Tout d'abord, soulignons que tous les syngraphèmes peuvent figurer seuls, et à l'initiale, et à la finale, et en position médiane dans le mot¹³.

Ceci dit, considérons des traits de combinabilité plus généraux.

De la même façon que Togeby (1965, p. 55) opère une distinction entre les consonnes vocaliques (toujours voisines d'une voyelle) d'un côté, et les consonnes consonantiques (séparables de la voyelle) de l'autre, nous avons réparti les syngraphèmes en syngraphèmes autographémiques (toujours voisins d'un autographème) et syngraphèmes syngraphémiques (séparables de l'autographème).

Selon cette distinction, sont autographémiques: <m> et <r>; sont syngraphémiques: les 13 syngraphèmes restants.

Or, à l'intérieur du groupe des syngraphèmes syngraphémiques, il est possible de différencier encore selon que ces graphèmes sont séparables dans toutes les trois positions dans le mot, ou seulement dans une ou deux positions.

Précisons qu'en ce qui concerne les groupes de syngraphèmes à l'intérieur du mot (en position médiane), il est nécessaire qu'au moins trois syngraphèmes se suivent pour qu'on puisse parler de syngraphèmes syngraphémiques.

13: Une seule exception est fournie par <x> en position initiale.

Si nous employons cette dernière distinction, <c>, <d>, <g>, <p>, <s> et <t> sont constamment syngraphémiques, étant donné qu'ils sont séparables de l'autographème dans toutes les trois positions.

En ce qui concerne les autres, <n> et <qu> sont inséparables en position initiale et médiane, mais séparables en position finale. (Ajoutons que le <n> ne se sépare de l'autographème qu'au moyen d'un autre <n>, attesté dans le seul mot *environn*, et que le <qu> (*q*) ne se sépare de l'autographème qu'au moyen de <n>, attesté dans le seul mot *cinq*.)

 est inséparable en position finale et médiane, mais séparable en position initiale.

<f> et <l> sont inséparables en position finale, mais séparables en position initiale et médiane. (En position médiane, <l> est représenté uniquement dans le mot *pareillement*. En position initiale, <l> ne se sépare de l'autographème qu'au moyen de <h>, comme nous l'avons déjà dit.)

<h> est inséparable en position médiane, mais séparable en position initiale (représenté dans le seul mot *hhomme*) et en position finale (représenté dans le seul mot *maroch*).

<x> est inséparable en position médiane, mais séparable en position finale (représenté uniquement dans le groupe final <lx>). Il n'est pas attesté en position initiale.

En suivant toujours le procédé de Togeby (1965, p. 55), subdivisons finalement les syngraphèmes autographémiques en syngraphèmes incompatibles avec d'autres syngraphèmes (n'entrant jamais dans un groupe) et en syngraphèmes compatibles avec d'autres syngraphèmes. (Il va de soi que les syngraphèmes syngraphémiques sont tous compatibles avec d'autres syngraphèmes.)

A cet égard, les deux syngraphèmes autographémiques se comportent ainsi: alors que <r> est compatible avec d'autres syngraphèmes dans les trois positions, <m> est compatible avec d'autres syngraphèmes en position médiane et finale (même si en position finale il figure uniquement avec <p>, relevé dans un seul mot, *champ*), mais en position initiale il est incompatible avec d'autres syngraphèmes.

6.5. «Accidental gaps», fréquence, etc.

Notre étude est basée sur un corpus clos, et nous nous prononçons uniquement sur les combinaisons attestées. Parmi les nombreuses com-

binaisons non-attestées¹⁴, nous ne distinguons pas entre lacunes systématiques et lacunes contingentes. Nous n'avons donc pas résolu le problème épineux des «accidental gaps»¹⁵. Il est quand même probable que la plupart des combinaisons non-attestées sont exclues du système. Citons à ce propos un petit passage d'un article de Eli Fischer-Jørgensen (1952, p. 35):

The consonantal and vocalic clusters actually found in a language will normally be of a restricted number (compared to the theoretical possibilities), and the phonemes found in the different positions in these clusters will be still more restricted, so that the clusters found can normally be said to belong to a few frequently recurring types, and thus it will not be possible to maintain that the non-occurring clusters are simply accidental gaps.

(Elle parle donc de phonotaxe, mais il est à croire que ce qu'elle dit peut également s'appliquer à la graphotaxe.)

Si ce problème était résolu, ce serait un grand avantage, car dans ce cas nous pourrions formuler des règles générales, et nous pourrions nous exprimer sur la probabilité pour telle combinaison non-attestée de pouvoir figurer dans un corpus élargi.

Comme le lecteur l'aura remarqué, nous n'avons pratiquement jamais indiqué la fréquence de telle ou telle combinaison, sauf dans les cas où il s'est agi de la fréquence 1. Or, l'ordinateur nous a fourni, entre autres choses, la fréquence de chaque mot, et les matériaux pourraient assez facilement être soumis à un traitement statistique. On pourrait mesurer la quantité d'information offerte par chaque graphème dans les différents contextes, et des problèmes comme la probabilité et la redondance pourraient être attaqués.

N'ayant étudié en détail qu'un texte assez court (12.103 mots-tokens, qui se réduisent à 1.839 mots-types), nous ne sommes pas en mesure de nous prononcer sur la représentativité de nos résultats. Il serait souhaitable de faire de la graphonomie comparative, aussi bien synchroniquement que diachroniquement.

En somme, le champ reste ouvert pour une multitude de projets de recherche ultérieurs.

14: Pour les combinaisons décrites ci-dessus, sur un total théorique de 225 combinaisons, seulement 20 se trouvent exploitées (c'est-à-dire moins de 10 %).

15: Voir Eli Fischer-Jørgensen (1952, p. 32-38) pour une discussion du problème.

7. Remarques finales

Les différents classements que nous venons d'esquisser contribuent tous à donner à chaque graphème sa définition. Or, convenons du fait qu'il est difficile d'obtenir une vue d'ensemble à partir de constatations comme celles que nous venons de faire.

Pour illustrer la grande diversité entre les graphèmes, on pourrait par exemple exposer le comportement détaillé de deux graphèmes qui se différencient considérablement par leur capacité ou leur incapacité de faire partie de groupes.

Le but essentiel du présent article a cependant été plutôt d'exposer et de discuter une méthode que de décrire les résultats détaillés. Un objectif accessoire serait atteint si la graphonomie française commençait à être considérée comme un champ d'étude méritant un peu d'attention.

Lise Lorentzen
Trondheim

Résumé

L'article traite de quelques problèmes soulevés par une étude graphémique. Le texte examiné date du XV^e siècle, mais la plupart des questions posées sont indépendantes de l'état de la langue étudiée.

Nous avons surtout discuté des problèmes relatifs aux questions de méthode. Il s'est agi de trouver la meilleure façon de définir les graphèmes, et la méthode choisie s'insère dans la tradition distributionnaliste.

Nous n'avons inclus qu'une fraction des résultats concrets auxquels nous sommes arrivée, mais ces résultats sont présentés dans un manuscrit non publié, intitulé *Graphes et graphèmes dans* *Traité de cosmographie de Jean Fusoris*. (Trondheim, 1978.)

Bibliographie des travaux cités

- Allén, S. (1965), *Grafematisk analys som grundval för textedering*. Acta Universitatis Gothoburgensis, Göteborg.
- Crossland, R.A. (1957), «Graphic linguistics and its terminology.» *Proceedings of the University of Durham Philosophical Society*, p. 13-16. Durham.
- Diderichsen, P. (1952), «Nye bidrag til en analyse af det danske skriftsprogs struktur.» *Selskab for nordisk filologi 1951-52*, p. 7-22.
- Fischer-Jørgensen, E. (1952), «On the definition of phoneme categories on a distributional basis.» *Acta Linguistica Hafniensia*, VII., p. 8-39. Copenhagen.
- Gleason, H.A. (1966), *An Introduction to Descriptive Linguistics*. Holt, Rinehart and Winston, New York.
- Grundt, L.O. et Lorentzen, L. (1977), «Essai d'une description graphématique d'un manuscrit français du XV^e siècle.» *Actes du 6^e Congrès des Romanistes Scandinaves*, p. 111-124. Uppsala.

7. Remarques finales

Les différents classements que nous venons d'esquisser contribuent tous à donner à chaque graphème sa définition. Or, convenons du fait qu'il est difficile d'obtenir une vue d'ensemble à partir de constatations comme celles que nous venons de faire.

Pour illustrer la grande diversité entre les graphèmes, on pourrait par exemple exposer le comportement détaillé de deux graphèmes qui se différencient considérablement par leur capacité ou leur incapacité de faire partie de groupes.

Le but essentiel du présent article a cependant été plutôt d'exposer et de discuter une méthode que de décrire les résultats détaillés. Un objectif accessoire serait atteint si la graphonomie française commençait à être considérée comme un champ d'étude méritant un peu d'attention.

Lise Lorentzen
Trondheim

Résumé

L'article traite de quelques problèmes soulevés par une étude graphémique. Le texte examiné date du XV^e siècle, mais la plupart des questions posées sont indépendantes de l'état de la langue étudiée.

Nous avons surtout discuté des problèmes relatifs aux questions de méthode. Il s'est agi de trouver la meilleure façon de définir les graphèmes, et la méthode choisie s'insère dans la tradition distributionnaliste.

Nous n'avons inclus qu'une fraction des résultats concrets auxquels nous sommes arrivée, mais ces résultats sont présentés dans un manuscrit non publié, intitulé *Graphes et graphèmes dans* *Traité de cosmographie de Jean Fusoris*. (Trondheim, 1978.)

Bibliographie des travaux cités

- Allén, S. (1965), *Grafematisk analys som grundval för textedering*. Acta Universitatis Gothoburgensis, Göteborg.
- Crossland, R.A. (1957), «Graphic linguistics and its terminology.» *Proceedings of the University of Durham Philosophical Society*, p. 13-16. Durham.
- Diderichsen, P. (1952), «Nye bidrag til en analyse af det danske skriftsprogs struktur.» *Selskab for nordisk filologi 1951-52*, p. 7-22.
- Fischer-Jørgensen, E. (1952), «On the definition of phoneme categories on a distributional basis.» *Acta Linguistica Hafniensia*, VII., p. 8-39. Copenhagen.
- Gleason, H.A. (1966), *An Introduction to Descriptive Linguistics*. Holt, Rinehart and Winston, New York.
- Grundt, L.O. et Lorentzen, L. (1977), «Essai d'une description graphématique d'un manuscrit français du XV^e siècle.» *Actes du 6^e Congrès des Romanistes Scandinaves*, p. 111-124. Uppsala.

- Haugen, E. (1956), «The syllable in linguistic description.» *For Roman Jakobson*, p. 213-221. Mouton, The Hague.
- Kurylowicz, J. (1948), «Contribution à la théorie de la syllabe.» *Bull. Soc. pol. ling.*, VIII, p. 80-114. Krakow.
- Kurylowicz, J. (1959), «La notion de l'isomorphisme.» *TCLC*, V, p. 48-60. Copenhagen.
- Lyons, J. (1977), *Semantics*, vol 1. Cambridge University Press, Cambridge.
- McLaughlin, J. C. (1963), *A graphemic-phonemic study of a Middle English manuscript*. Mouton, The Hague.
- Naert, P. (1961), «Une définition et classification non-phonétique des graphèmes du vieil-islandais.» *Studia Linguistica*, XV, p. 29-51. Lund.
- Opdahl, L. (1966), *A Graphemic Analysis of The Equatorie of the Planetis*. Mémoire non publié, Université d'Oslo.
- Piirainen, I. T. (1968), *Graphematische Untersuchungen zum Frühneuhochdeutschen*. Walter de Gruyter & Co., Berlin.
- Sigurd, B. (1965), *Phonotactic Structures in Swedish*. Uniskol, Lund.
- Spang-Hanssen, H. (1952), «Analyse ved hjælp af kombinationsdiagrammer.» *Selskab for nordisk filologi 1951-52*, p. 22-26. Copenhagen.
- Spang-Hanssen, H. (1959), *Probability and structural classification in language description*. Rosenkilde et Bagger, Copenhagen.
- Togeby, K. (1965), *Structure immanente de la langue française*. Larousse, Paris.
- Troubetzkoy, N.S. (1970), *Principes de phonologie*. Trad. par J. Cantineau. Klincksieck, Paris.
- Westring Christensen, B. (1964), *En typologisk beskrivelse af fonem- og grafemforbindelser i monovokaliske ordudtryk på moderne fransk*. Mémoire non publié, Université de Copenhagen.
- Westring Christensen, B. (1967), «Phonèmes et graphèmes en français moderne. Quelques réflexions typologiques.» *Acta Linguistica Hafniensia*, X, p. 217-240, Copenhagen.