

Dora Riffer-Maček

The Graphemic Inventory of a Middle English Manuscript

1. The subject of the present paper is the description of the graphemic inventory of MS New College, Oxford, 95, folios 1 to 31a line 5. The MS consists of a complete set of Wyclif's Ferial Gospels. The particular part of the MS studied here forms a linguistically homogeneous piece of text, as it is written in the first of the four hands of the MS. After the pages 31b and 32 the text is resumed in the first hand for about 100 pages, but that part of the MS has not been included in the present study.

The purpose of this paper is to provide a description of the written signals of the above mentioned piece of text; this is envisaged as one step towards a detailed linguistic study of the MS. Such a study might also furnish interesting information relevant to research in the development of Standard English.

The analysis is based on the assumption that written language is one of the two expressions, namely the visual expression of language. Written language has its own system, the development of which is not necessarily parallel or even dependent on the other expression of language, the spoken language.¹ Thus written language conveys meaning directly

¹ For reference see: D. L. Bolinger, "Visual Morphemes", *Language*, Baltimore, 22/1946, no. 22, pp. 333—340; W. N. Francis, "Graphemic Analysis of Late Middle English Manuscripts", *Speculum*, Cambridge, Mass., XXXVII/1962, pp. 32—47; A. McIntosh, "Graphology and Meaning", *Archivum Linguisticum*, Glasgow, 13/1961, vol. 13, fasc. 2, pp. 107—120; A. McIntosh, "The Analysis of Written Middle English", *Transactions of the Philological Society*, Oxford, 1956, pp. 26—55; R. P. Stockwell, *Chaucerian Graphemics and Phonemics*, (dissertation in manuscript); H. J. Uldall, "Speech and Writing", *Transactions of the Philological Society*, Oxford, 1944, vol. 4, pp. 11—16; J. Vachek, "Two Chapters on Written English", *Brno Studies in English*, Praha, 1959, vol. I, pp. 7—34.

through its visual signals in some sense independently of their relations to the spoken language. The units of WL (written language) have to be established "on the Basis of visual criteria in combination with distinctiveness of value"². The units of the phonological system, with which the units of a graphemic system do not necessarily correlate in any simple way will be, for the present purpose, disregarded.

The text of a Middle English MS consists of letters and other visual signals such as non-alphabetical abbreviations and punctuation marks. All these signals will be called *graphs*. From the inventory of graphs a smaller inventory of graphemes will be finally derived. *Graphemes* are the smallest distinctive units of WL. The distinctiveness of their value can be arrived at by studying them in their immediate context, which is the word. In a passage of text words are divided by wide space, while narrow space or absence of space marks off graphemes within a word. The position of a grapheme after a wide space will be called *initial* and before it *final*. Within the word graphemes are in *medial* position. The absence of distinctively contrastive characteristics between two or more graphs is a feature of allographs belonging to one allographic set, that is, a characteristic of graphs which do not stand to each other in opposition, but in complementary distribution or free variation.

Beside graphemes and allographs there are a number of graphs which are non-alphabetic and which stand mostly in free variation with one or more graphemes, or a sequence of graphemes, sometimes whole morphemes. Also, there is another set of graphs, which function in the same way as the punctuation in modern written or printed texts.

The list of graphs is given below in alphabetical order, the non-alphabetical graphs following the alphabetical ones. The differences in shape reproduced here are only the most obvious ones, while the various elongations of the free ends of graphs, loops added, or ornaments, and distinctions inevitable in a MS, and obviously due to chance variation of handwriting, are not taken into account. The list does not include written signals of the Latin text at the beginning of each sermon, nor the written signals of the glosses; it consists only of the graphs found in the book hand of the English text.

² A. McIntosh, "Graphology and Meaning", *Archivum Linguisticum*, Glasgow, 13/1961, vol. 13, fasc. 2, p. 110.

1	a	9	e	17	l	25	R	33	v	41	f	49	g	57	≠
2	ɹ	10	f	18	ɣ	26	r	34	zβ	42	ɳ	50	o	58	=
3	d	11	g	19	ɱz	27	z	35	ɣ	43	ɸ	51	ɸ		
4	B	12	h	20	n	28	S	36	y	44	t	52	ɸ		
5	b	13	ɣ	21	ɔ	29	ɟ	37	ɣ	45	t	53	ɸ		
6	c	14	i	22	o	30	ɔ	38	ɣ	46	ɸ	54	/		
7	C	15	ɣ	23	ɸ	31	t	39	ɣ	47	ɸ	55	//		
8	d	16	L	24	g	32	u	40	y	48	ɸ	56	.		

2. From the inventory of graphs we proceed to derive a smaller inventory of graphemes by determining the contrastiveness of graphs in the context. The procedure of contrasting the above graphs in their immediate context has given the following graphemic set.

GRAPHEMES	represented by	GRAPHS
$a_{1,2,3-n^3}$	" "	1, 2, 3
$b_{1,2-n}$	" "	4, 5
$c_{1,2-n}$	" "	6, 7
d_n	" "	8
e_n	" "	9
f_n	" "	10
g_n	" "	11
h_n	" "	12
$i_{1,2,3-n}$	" "	13, 14, 16
k_n	" "	15
$l_{1,2-n}$	" "	16, 17
$m_{1,2-n}$	" "	18, 19
n_n	" "	20
$o_{1,2-n}$	" "	21, 22
p_n	" "	23
q_n	" "	24
$r_{1,2,3-n}$	" "	25, 26, 27
$s_{1,2,3-n}$	" "	28, 29, 30
t_n	" "	31
$u_{1,2-n}$	" "	32, 33
w_n	" "	34
x_n	" "	35
z_n	" "	37
β_n	" "	38

³ n denotes the infinite number of variations of the graphs representing the allographs of an allographic set.

Graphs 39 to 51 are abbreviations used for various graphemes or sequences of graphemes:

39 — and	43 — pro-	47 — ur
40 — con-, -is/, -es -us	44 — at	48 — a, ra
41 — -is/, -es	45 — i + one or more other graphemes	49 — m, n
	46 — re, er, ur, -es/-is	50 — par-, per-
		51 — er

The remaining graphs (52 to 58) differ from the above mentioned graphemes in that they have a syntactic function similar to the function of punctuation.

To the above list one more graphemes could be added, that is, any alphabetic grapheme when doubled, as will be shown later.

Abbreviations, "punctuation" graphs and double alphabetic graphs can all be regarded as graphemes because of their distinctiveness of value. That means that they cannot be freely replaced or omitted if the meaning is not to suffer. The context for all graphemes except "punctuation" graphemes is the word:

an : in, as : is, at : it
 bad : had, : ben
 came : same, come : dome, cope : pope
 day : may, don : zon, and : any
 en : an, es : is, men : man
 face : lace, foule : poule
 gete : mete, goo : loo, gost : most, long : lond
 he : be, his : is : pis,
 is : as
 take : tale
 lawe : sawe, lif : zif, lord : word, holy : hony
 make : take, mot : not, hem : her
 ne : be : we, no : so
 of : if, on : an, lost : last : lust
 pope : cope : hope
 quick
 worde : wolde
 same : came
 to : do, at : as : al
 loue : lore
 we : ne, wo : to, lowe, : loue
 fox : for
 worþe : worde, deþ : ded
 nazareþ, elizabeþ, eiez, rizt, lizt

Graphemes 39 to 51 will be shown here, as far as examples from the text permit, not in opposition, but in free variation with alphabetic graphemes.

39 — and

40concei₃u₁ed — co50cei₃u₁e — concei₃u₁ed, mann40 — manni₃s,

pu40 — pu₃s,

king41 — kingi₃s₃

42alme

43fited, 43fitis

44phete, 44phecie

p45 — pat

c46er — crier,

g47ce — g49ce — grace, p47stis — prestis, bataille47, barakie47

c47se — cursid

p48chase

abrah49m, thom49s,

ma50nes — mannis, da50pned — dampned

p51te, p51ceiue

des₃52t — des₃er₂t

a) There is an undefinable number of allographs in an allographic set, and they all represent one and the same grapheme. Most of the allographs, in the above table marked n , can be freely varied though there are some limitations: allographs whose free ends are extended above the line (d , b) or below it (h), occur only in the first and last line of a page. Similarly, allographs with free ends extending to the right (s_2) occur only in final position, especially at the end of line. The angular and rounded varieties of some graphs (e. g. β), as well as graphs with various additional strokes or loops (r_2 , l_2) stand in free variation to the simple forms of the grapheme they represent.

b) The following allographs stand in complementary distribution: r_2 and r_3 , s_2 and s_3 , u_1 and u_2 .

r regularly occurs after the grapheme o , where r_2 has been found only in one example, possibly a case of scribal error: *nor₂ische*. After a grapheme whose exponents have a round shape (like b , d , c , a) r_2 and r_3 can be freely varied. It is to note that r_3 never occurs in initial position.

s_2 never occurs finally except as the variant with the extended upper end. On the contrary, s_3 is found only in final, and sometimes in initial position. Only once it occurs medially, which again may be a mistake: *fals₃e*.

u_1 occurs as a rule medially and finally, but there are two instances in our text where it is found in initial position, the normal position for u_2 . The examples are: u_1p and u_1nto , which are in the rest of the text spellt only u_2p and u_2nto .

c) A number of allographs ($a_{1,2}$, b_1 , c_1 , i_1 , l_1 , m_1 , o_1 , r_1 , s_1 and the angular variety of β) stand in complementary distribution with other allographs of their respective allographic sets at the beginning of a sermon, or in free variation after punctuation graphemes. An exception is a_1 which occurs only at the beginning of sermons. Thus these allographs are syntactical signals for the beginning of a new paragraph and function in the same way as punctuation graphemes. This feature is called capitalization and is also present in two instances where the grapheme is doubled: ff after a punctuation grapheme, and the line-end variant of s_2 at the beginning of a sermon in the word $a_1s_2s_2$.

d) In other cases double graphemes occur mainly medially and do not have a syntactic value. They normally stand in opposition to single graphemes, but there are also cases of synonymy between them. Not all alphabetic graphemes occur doubled, nor do double graphemes occur in all the three positions. Graphemes occurring medially:

a (baal, balaam, caas)

b (robben)

c (occasiou50, grucche, leccherie)

d (myddel, biddingis, haddde, fledde, fedde, dredde). This grapheme occurs only in front of the graphemes e and i and is therefore common in the past tense.

e (scheep, heed, freel)

f (suffre, suffisen, diffine)

g (beggerd)

l (killinge, gospelle, fulfillide, fallas). It is normal in front of the graphemes e , i and a . If in similar graphemic sequences these graphemes do not occur, doubling is left out: *killinge*, but *kild*, *fulfillide*, but *fulfild*, *fallas*, but *fals*.

n is also doubled only in front of e and i , and in one instance before a (annas). Even in these positions it seems to be synonymous with single n : *begining* — *beginninge* — *beginniþ*.

o if doubled occurs frequently with ligature.

p when doubled has as its first component the abbreviation grapheme no. 44, and as the second the grapheme p .

r can be doubled either by repeating r_2 (ver_2r_2ei , myr_2r_2e) or, when the position after o demands it, r_2 is preceded by r_3 (cor_3r_2upte).

s_2 is frequently doubled, especially in the morpheme *-nesse*, but not so the other allographs of s .

t (witte, wittili, better, grette, gretter, putte, letten)

The following graphemes occur doubled in final position as well:

e (flee, 3ee)

- o (loo)
- t (witt, grett, putt, writt), and
- l when followed by an abbreviation grapheme as in *gospel148*.
In initial position graphemes are doubled only if they are in combination with an abbreviation:
- g (*g44gorie*; here *gg* does not function as a capital)
- c (*c47cu51cidid*).

Other cases of synonymy between double and simple graphemes are: *staate* — *state*, *citees* — *cites*, *steffli* — *stefli*, *ioon* — *ion*, *matthew* — *mathew*. In the words *godis* — *goddis* the double and single grapheme are not synonymous but opposed, the first being the genitive plural of “good”, the second the genitive singular of “God”.

e) Besides the above mentioned kind of synonymy there are words which differ in one single grapheme:

- any — eny
- mannes — mannis, sire — siri
- helizabeth — helizabet — elizabet *zhe* — *ze*, nazarep̄ — nazareth — nazaret, zacharie — zacarie, ihon — ion — zon
- lickli — licli
- herde — herd
- nexste — next
- hous — hows
- muche — miche, suche — siche

For the moment it is not our concern that the reasons for the synonymous status of the cited words are different. It is assumed that as soon as two different graphemes stand in the same context, or are omitted from the same context without changing its meaning synonymy is in question.

f) As regards the distribution of the graphemes the following characteristics are of interest:

1. Only the graphemes *a*, *i*, *oo*, and *p̄* and *h* in combination with an abbreviation occur as a sole grapheme in a word.
2. Some graphemes are rare or even exceptional in final position:
 - a* (only in *bethsaida*, *locusta*, *tetrarca*, *roma*, and in combination with the abbreviation no. 49: *ama50*, *noma50*).
 - b* (*iacob*, *moab*, *iob*) only in foreign names.
 - c* (*patriarc*, *milc*, *abac*). In the word *patriarc* *c* varies with *k* when it stands in a form of the word where it does not occupy the final position (*patriarkis*).
 - h* (*nazareth*, *helizabeth*, *elizabeth*, *ioseph*) like *b* only in foreign names.

p (only in: *scheep*, *archebischop*, *up*, *philip*, *lordschip*, *sleep*).
 q never occurs finally.
 w (only in: *matthew*, *now*, *saw*, *vow*, *shadow*),
 x (*fox*),
 з (*niz*, *iuз*, *eiez*, *porouз*).

3. The initial position is rarely or never occupied by the following graphemes:
 q (*quick*, *quikid*, a *quei50tid*)
 x (never in final position).
4. The graphemes *q* and *x* are rare even medially, *q* occurring only in the word *required*, and *x* in merely 16 cases.
5. There are some graphemic sequences which occur with a much greater frequency than other sequences.
q is always followed by *u*₁,
зt is a common sequence when *з* occurs medially.
 Other frequent sequences are *ch*, *ck*, *sch*.
 More statistical data would be required for determining the functions of such sequences.
6. A special case are graphemes whose exponents consist of a sequence of vertical strokes linked either at the top or at the bottom. By this are meant the allographs *m*₂, *n*, *u*₁ and *i*₂. The linking strokes are their main visual distinctive features, but in the handwriting they are not always clear enough, so that the mentioned graphemes can sometimes be confused. This is frequently prevented by other signals, such as space or distribution.
 - a) Space is the main feature that distinguishes combinations of more than four strokes.
 - b) Doubling of graphemes does not occur either initially or finally so that in these positions the combination of several vertical strokes indicates that it must be a sequence of various graphemes, or in the case of three strokes, the grapheme *m*.
 - c) Even medially *n* is the only grapheme that occurs doubled, mainly in front of *e* or *i*.
 - d) The initial position is also a negative signal for *u*₁.
 - e) In collocations of the graphemes *i* and *m*, *i*₃ is much more frequent than *i*₂ with which it occasionally stands in free variation.

The mathematically infinite number of combinations can thus be greatly reduced, and a more detailed study of the collocations would give some more precise data about the distribution of the particular graphemes.

g) *Abbreviations* are used as free variants of one grapheme or a sequence of graphemes, which enables us to draw conclu-

sions about what they correspond to. According to their visual characteristics they can be divided into those equaling ordinary alphabetic graphemes in size, and others which are smaller and normally placed above the line. In this paper the grapheme numbers given after an alphabetic grapheme represent abbreviations placed above the grapheme they follow. In the case of 51 it means that the abbreviation is written below the line, and 52 is the second exception as it is placed on the line, following s_2 . Graphemes 39 to 44 belong to the first group, and 45 to 52 to the second group.

39 is the variant of the graphemic sequence *and* when it is used as a free morpheme.

40 initially varies with *con-* (40*cei₂u₁ed*), finally with *-us* or the morpheme *-es/-is* (king40).

41 is used only finally,

42 and 43 only initially, also

44 which frequently appears in the sequence 43*ph* in the word "prophete". It also appears in ligature with *p* as a sign for *pp*. The grapheme no 42 is probably also a ligature of 44 and s_2 used in the word "psalm".

Abbreviations no. 45 to 52 occur always above, below or after an alphabetic grapheme as mentioned earlier. Some of these abbreviations are more frequent with some alphabetic graphemes and less frequent with others. 45 occurs mostly with *p*, 51 only with an initial *p*, and 52 with initial or medial s_2 .

Many of the abbreviations are limited to a certain position in the word;

initial only 42, 43, 51

only final, 41

40 occurs both finally and initially;

other graphemes change the position more freely which depends on the alphabetical grapheme they are linked to.

49 stands both for a single grapheme (*a*) and a sequence of graphemes containing *a*. 46 and 47 represent various sequences of graphemes, 46 always containing an *i*, 47 a *r* grapheme. Moreover, 46 can also represent several non-consecutive graphemes in the word *i₂er₂l46m* (Jerusalem), and 47 in final position stands for the morpheme *-es/-is*.

It is interesting to note that there are several abbreviations used for the morpheme *-es/-is* (40, 41, 47), and it should also be observed that certain abbreviations are used mostly in a limited number of words:

39 only for *and*

42 only in 42alm

43 in 43fit, 43fitable etc.

44 mostly in 44phete and other derivatives of the same stem.

h) The graphemes which are listed in the table under numbers 53 to 58 are functioning syntactically by marking off units of text larger than a word. An exception is no. 58 which links the sequence of graphemes at the end of a line with the sequence of graphemes at the beginning of the next line into one word.

The main point about the graphemes 52 to 57 is that their use is not compulsory. They are derived from liturgical writings and are primarily designed to mark intonation. Graphemically they are important as signals of capitalization, word division where space is too narrow, and word linking (58).

52 occurs at the beginning of large divisions of text (in our case only twice). 53, 54, 55 and 56 signal divisions between groups of words which form breath groups or sense groups.

*

3. Here should be stressed that the above described graphemic inventory is liable to correction after the remaining text of the first hand of MS New College 95 has been studied, as well as that the results obtained so far are limited to the text studied, and cannot be extended to other Middle English manuscripts without thorough comparison. An extensive statistical investigation of the distribution of various graphemes and their collocations, and a detailed study of the abbreviations and "punctuation" graphemes, would give a more accurate picture of the graphemic system of this particular corpus, and from a necessary prelude would lead on to the study of its morphology and syntax. Such a study in its turn would help us to form a clearer picture not only of the system of the written language of the text, but of its underlying language system as well.