ON THE GRAPHEMIC REPRESENTATION OF ME /g/ IN MODERN ENGLISH

PETER H. ERDMANN
Hamburg University

In his recent monograph on the Structure of English orthography Venezky points out that the "graphemic differentiation of homophones was noted as an essential feature of English orthography as early as the beginning of seventeenth century" (Venezky 1970: 19–21). Venezky's reference is to Alexander Hume's 'Of the Orthographie and Congruite of the Britan Tongue' written about 1617 (Venezky 1970: 122). He goes on to say: "to what extent conscious efforts have been made to enforce this orthographic function is difficult to determine" (Venezky 1970: 122). In what follows I have nothing to say that might push the evidence Venezky has provided for a conscious effort to differentiate homophones graphemically sometime decades backward. I would rather like to point out that the phonological conditions for keeping homophones graphemically apart were given much earlier. Taking the Middle English fricative /g/ as an example one can demonstrate that the necessity to differentiate homophones in English arose with the development of new diphthongs in Early Middle English out of the sequence "vowel plus /g/". A further reason was the loss of the spirant /g/ in Late Middle English.

In Late Old English and Early Middle English new diphthongs arose from the sequence of a long or short vowel plus semivowel or fricative /g/. Etymologically LOE /g/ goes back to Common Germanic /z/ in some cases. We have argued elsewhere that the Germanic fricative /z/ has to be interpreted as underlying /g/ in Late Old English (Erdmann 1971: 3, 21). We will accept this interpretation for the rest of the paper, which is not decisive for our conclusions, but which simplifies the ensuing discussion of the problems involved. To give some examples for the sequence of 'vowel plus semivowel' first. The listed forms are underlying or base forms of the various periods, which are controlled by the derivational rules and were reconstructed on the evidence of Early Old English material:
Pre-OE /daujan/ “to die” > OE /digjan/ > LOE /dijen/ > /di-en/
Pre-OE /kljij/ “clay” > OE /klij/ > LOE /kli/ (Lauck 1994: § 278) > /kli/
OE /groman/ “to grow” > /grom-en/
OE /flowan/ “to flow” > /flow-en/.

The newly arising diphthongs are short diphthongs, i.e. bimoric units, according to a phonotactic rule of Late Old English and Early Middle English (Erdmann 1971: 3.21).

To turn to the sequence of “vowel plus fricative /j/”. In Late Old English the fricative /j/ is substituted by the palatal or velar semivowel /j/ or /w/ according to the quality of the immediately preceding vowel when followed by a voiced segment, i.e. another vowel or one of the resonants. Consider the following examples:

OE /segl/ “sail” > LOE /seggl/ = /segjl/ > /segjl/ > /sêl/
OE /regn/ “rain” > LOE /regn/ = /rejgn/ > /rejgn/ > /rejn/
OE /plecgn/ “play” > LOE /plecgn/ = /plecje/ > /plej-
OE /fowl/ “fowl” > LOE /fowl/ = /fowl/ > /fowl/
OE /gown/ “to own” > LOE /gown/ = /munen/ > /munen/
OE /drawn/ “to draw” > LOE /drawn/ = /drawn/ > /drawn/.

We have described the process as a replacement of /j/ by /j/ or /w/ (marked by a double arrow ➞) for reasons of orthography to be discussed below. We will take up the /j/ substitution first. Of the three front vowels of Late Old English, i.e. /i/, /e/ and /ê/, the sequences of mid front vowel plus /j/ and low front vowel plus /j/ coalesce in Early Middle English giving a diphthong /ae/ (Jordan 1968: § 95), e.g.

[LOE /wej “way”/] > /wej/ > EME /wej/.
[LOE /lej/ “lai”/] > /lej/ > EME /lej/.
[LOE /daej/ “day”/] > /daej/ > EME /daej/.
[LOE /maej/ “may”/] > /maej/ > EME /maej/.

As is apparent from Modern English spellings like /way/, /laid/, /day/, /may/ the substituted /j/ is nowhere spelled /gh/ in voiced environment after an original front vowel. The vocalized /j/ is represented graphemically as /y/ or /î/ according to the following segment: in case of a following consonant the grapheme /î/ is used, otherwise we find the spelling <y>. These rules will have to be refined, which is not the topic of the present paper. Marked deviations from the cited graphemic pattern occur only in cases of homophones, e.g. /way/ vs. /weigh/. Both forms read /wej/ vs. /wejen/ in Late Old English. The two forms /wej/ and /wejen/ coalesce in Late Middle English with the loss of /-n/ and unstressed /e/. Since that time the phonological conditions for differentiating the homophones graphemically are given. The grapheme

chosen to differentiate the Late Middle English homophone /weaj/ was obviously influenced by the noun /weigh/ “weight”, which preserved the graphemic reflex /gh/ of the fricative /j/ in voiceless environment to be discussed later on. The seemingly irregular /weigh/ “to weigh” therefore results from graphemic differentiation of the homophone /weaj/, which arose in Late Middle English. Another striking pattern can be found for the high front vowel /i/. Consider Modern English spellings like <dry> “dry”, <rye> “rye”, <sty> “stye”, which read /dryj/, /riye/ and /stij/ in Late Old English. The forms are contracted to /dry-i/, /ri-i/ and /stij-i/ in Middle English and lose their unstressed vowel around 1400 (Horn 1954: § 305). Graphemically the bimoric unit /i-i/ is either represented as /y/ or as /ye/. If the /y/ (e) is preceded by more than one grapheme, the unit /i-i/ is represented as /ye/, e.g. <dry>, <sty>, <rye>. Let us look at the following Modern English forms:

<die> “to die” vs. <die> “to dye”
<he> “to lie” vs. <lye> “lye”
<tie> “to tie” vs. <tye> “tye”

The Late Old English forms can be reconstructed as 1
/dyj/ vs. /dyj/  
/rije/ vs. /rije/  
/stij/ vs. /stij/  

The sequence /ej/ is shortened to /i/ in Middle English like /oc/ results in the bimoric unit /uui/ (Erdmann 1971: 3.23). The first pair coalesces with the shortening of diphthongs to bimoric units in (Early) Middle English, i.e. /jii-en/-. The two remaining pairs become identical with the loss of /-n/, i.e. /i-i-e/ and /i-i-e/. Since that time the phonological conditions for differentiating the homophones graphically are given. Concerning the graphemic representation we can take /ye/ to be the unmarked grapheme as opposed to /i/ as the marked graphemic sequence for the Late Middle English bimoric unit /i-i/. This opposition is not used in the case of /fly/ “fly” (2) because of etymological reasons. In the case of /shy/, /aly/, /eye/ we find the unmarked graphemic representation. The only exception to this pattern is Mod Eng /island/, an erudite spelling of the 16th century. Mod E /hie/ “to hasten” looks irregular. It should be spelled /* hie/ according to what we have said so far. Mod E /hie/ is probably the only surviving member of an original opposition, which can be found in Middle English, i.e. /hie/ “mind” vs. /hieh/n/ “to hasten” shows up as /hiei(e)/ vs. /hiei(e)/ after contraction. Mod E /sigh/ should likewise be spelled /* sigh/. As in the case of weigh, the Modern English spelling /sigh/ is influenced by a form of the paradigm in voiceless environment, e.g. by the weak preterite /sight/ “sighed”, which

1 These Late Old English forms postulated in the derivational history of the items in question should not be confused with attested forms.
is regularly spelled ⟨sight⟩. Modern E ⟨sigh⟩ is a back-formation of the weak preterite form.

The cases considered so far make the Modern English spelling look much less chaotic than it appears to be. The Late Old English fricative ⟨f⟩ which is substituted by a semivowel according to the quality of the immediately preceding vowel when followed by a voiced segment shows a definite pattern. We have to distinguish between marked and unmarked representations. The Middle English bimoric unit ⟨/i⟩ is spelled ⟨/i⟩. In cases of homophones arising in Middle English it is opposed to the grapheme ⟨<e⟩. This pattern is skewed by 1) erudite spellings like Modern E ⟨island⟩, 2) etymological considerations, e.g. Mod E ⟨fly⟩, 3) morphological factors, e.g. Mod E ⟨weigh⟩ and ⟨/sigh⟩ and 4) by the extinction of members of original pairs of opposition, which look irregular from the Modern English point of view, e.g. Mod E ⟨tie⟩.

Let us next take up the graphemic representation of LOE ⟨/f⟩ in voiceless environment following a front vowel, i.e. ⟨/i⟩, ⟨/e⟩ and ⟨/æ⟩. In Late Old English short ⟨/e⟩ was raised to ⟨/i⟩ before the checked fricative ⟨/f⟩. Consider the following correspondences:

| LOE /kni:ʃt/ 'knight' | Mod E ⟨knight⟩ |
| LOE /fliːt/ 'flight' | Mod E ⟨flight⟩ |
| LOE /riːʃt/ 'right' | Mod E ⟨right⟩ |

In contrast to the palatal variant of LOE ⟨/f⟩ in voiceless environment which is not represented graphemically but for the reasons mentioned above the same variant before a voiceless segment is written ⟨gh⟩ in Modern English. This strongly suggests that ⟨/f⟩ before a voiceless obstruent was not substituted by a semivowel, but remained a spirant until Late Middle English. The development seems to have been the intrusion of a palatal or velar glide in Late Old English or Early Middle English, which was subsequently phonemized (Vachek 1965: 3-13):

| LOE /kni:ʃt/ (→[kniʃt]) > /kniːʃt/ |
| LOE /fliːt/ (→[fliːt]) > /fliːʃt/ |
| LOE /riːʃt/ (→[riːʃt]) > /riːʃt/ |

The phonemized glide ⟨/iː⟩ was vocalized and contracted with the immediately preceding vowel to form a bimoric unit, i.e. ⟨/ii⟩ in our case:

| LOE /kniːʃt/ > ME ⟨kniːʃ⟩ |
| LOE /fliːʃt/ > ME ⟨fliːʃ⟩ |
| LOE /riːʃt/ > ME ⟨riːʃ⟩ |

The bimoric unit ⟨/ii⟩ is represented by ⟨i⟩ in Modern English, the following spirant ⟨/p⟩ by ⟨gh⟩.

To recapitulate the graphemic principles applied so far, the graphemic representation of the palatal variant of LOE ⟨/f⟩ is regulated by distribution. Before a voiced segment, i.e. a vowel or one of the resonants, the fricative ⟨/f⟩ was replaced by ⟨/ʃ⟩ which vocalized and contracted with the preceding voca- lic unit. The same applied to the palatal variant of ⟨/ʃ⟩ in final position which was replaced by ⟨/ʃ⟩ in Early Old English, since it did not participate in the general devoicing of voiced obstructions found otherwise (Brummel 1965: §214). Before a voiceless obstruent the palatal variant of ⟨/ʃ⟩ was not substituted by ⟨/ʃ⟩, but remained a spirant since Late Middle English. There developed a semivowel before it which vocalized and contracted with the preceding voca- lical nucleus. The fricative ⟨/ʃ⟩ was represented by ⟨gh⟩ in this environment. The mentioned principle of graphemic representation of the palatal variant of LOE ⟨/ʃ⟩ led to a further specification in the case of homophones. For the palatal variant of voiced ⟨/ʃ⟩ we cited the graphemic opposition ⟨/e⟩ vs. ⟨/i⟩, which is overlaid by other criteria in some cases. Similar graphemic oppositions can be found with the voiceless palatal variant.

A first pattern emerged within forms that contained LOE ⟨/ʃ⟩ before a voiceless obstruent following a front vowel. Its graphemic reflexes are found in two irregular spellings of etymologically related lexemes, i.e. ⟨/high⟩ vs. ⟨/height⟩, ⟨/sly⟩ vs. ⟨/slight⟩. The corresponding Late Old English forms read:

| LOE /hæɡ/ (→[hæʃ]) > /hæɡ/ |
| LOE /hæɡp/ 'height' (→[hæʃp]) > /hæɡp/ |
| ME /slæɡ/ 'sly' > LOE /slæɡ/ |
| LOE /slæɡp/ 'slight' (→[slæʃp]) > /slæɡp/ |

According to the mentioned phonotactic rule of the newly arising diphthong ⟨/ii⟩ is shortened to ME ⟨/i⟩:

| /hæɡ/ > ME ⟨hiiːɡ⟩ → [hiiːʃ] |
| /hæɡp/ > ME ⟨hiiːɡp⟩ → [hiiːʃp] |
| /slæɡ/ > ME ⟨sliːɡ⟩ → [sliːʃ] |
| /slæɡp/ > ME ⟨sliːɡp⟩ → [sliːʃp] |

In the case of ME ⟨hiiːɡ⟩ and ME ⟨sliːɡ⟩ the dental spirant ⟨/p⟩ of the nominal suffix is replaced by ⟨/ʃ⟩. Obviously a process of dissimilation between neighbouring spirants, i.e. ⟨/ɡ−p⟩ → ⟨/ɡ−ʃ⟩. Unstressed final ⟨/e⟩ drops in Late Middle English. The two nouns LME ⟨hiiːɡ⟩ and LME ⟨sliːɡ⟩ should

1 The arrow (→) marks a phonetic realization rule, the angle ⟨/⟩ stands for a sound-shift. Since only phonemes change, the realizational rule is put in parentheses to indicate that the sound-shift under consideration can be characterized as a phonemization of allophones.
be represented graphemically as \(<\text{hight}\rangle\) and \(<\text{slight}\rangle\). The Modern English spelling, however, is \(<\text{height}\rangle\) and \(<\text{slight}\rangle\). The reason for this irregularity again is the rise of homophones in Middle English.

The strong reduplicating verb OE /hūtum/ "be called" shows preterite forms which are spelled \(<\text{hæt}\rangle\) or \(<\text{het}\rangle\) (Brunner 1965: § 394). The spelling \(<\text{het}\rangle\) is regularly interpreted as \(<\text{hæt}\rangle\). The development to ENE /hæt/ "is called" requires a Late Old English form \(<\text{hæt}^\prime\rangle\) underlying graphemic \(<\text{hæt}\rangle\) with the raising of \(<\text{e}\rangle\) to \(<\text{i}\rangle\) before a checked spirant as in the case of LOE /kniht/ and LOE /kniht/.

Mod E \(<\text{slight}\rangle\) goes back to LOE /kniht/ developing regularly.

\[
\text{LOE} \quad \text{[kniht]} \quad \text{[slight]} \quad \text{[kniht]} \quad \text{[kniht]} \quad \text{[kniht]}
\]

The two pairs, i.e. EME /hægh\text{e}\text{e}/ "height" vs. /kniht/ "is called" and EME /kniht/ "slight" vs. /kniht/ "slight", coalesced in Late Middle English when \(<\text{e}\rangle\) dropped and the dental spirant \(<\text{f}\rangle\) was replaced by the corresponding obstruent. Since that time the phonological conditions for differentiating the homophones graphemically are given:

\[
\text{LME} \quad \text{[hætg]} \quad \text{"height"} \quad \text{< height >}
\]

\[
\text{LME} \quad \text{[hætg]} \quad \text{"is called"} \quad \text{< height >}
\]

\[
\text{LME} \quad \text{[kniht]} \quad \text{"slight"} \quad \text{< slight >}
\]

The substituted \(<\text{g}\rangle\) is not represented graphemically. The graphemes of Modern English mirror the newly arising binomic units, i.e. \(<\text{law}\rangle\), \(<\text{bow}\rangle\), \(<\text{fowl}\rangle\).

Before a voiceless segment we predict a representation of \(<\text{g}\rangle\) and the preceding binomic unit. Let us first consider the fricative \(<\text{g}\rangle\) before a following voiceless obstruent:

\[
\text{LOE} \quad \text{[friht]} \quad \text{"fought"} \quad > \quad \text{[friht]} \quad > \quad \text{ME} \quad \text{[friht]}
\]

\[
\text{LOE} \quad \text{[drægh\text{e}e\text{e}]} \quad \text{"drought"} \quad > \quad \text{[drægh\text{e}e\text{e}]} \quad > \quad \text{ME} \quad \text{[drægh\text{e}e\text{e}]}\]

An exception to the preconsonantal representation is LOE /kriht/ "trott" which is spelled without \(<\text{gh}\rangle\) in Middle English obviously in analogy to Old French (\<true\>, \<true\>, \<true\>) (Brunner 1960: 389). The velar variant of LOE /g/ which is written \(<\text{gh}\rangle\) before a voiceless segment like the palatal variant of LOE /g/ is not pronounced in Modern English with two exceptions. Besides the quoted instances we find Mod E /dright/ (\<draft\>, \<draft\>). The allophone /s/ is found since the 16th century to differentiate the polysemous lexical item ME /dright/ in Early New English (Brunner 1960: 389). In final position the Late Old English fricative \(<\text{g}\rangle\) is likewise represented graphemically by \(<\text{gh}\rangle\):

\[
\text{LOE} \quad \text{[t\Upsilon\text{g}]} \quad \text{"tough"} \quad > \quad \text{[t\Upsilon\text{g}]} \quad > \quad \text{ME} \quad \text{[t\Upsilon\text{g}]}\]
\[
\text{LOE} \quad \text{[r\Upsilon\text{g}]} \quad \text{"rough"} \quad > \quad \text{[r\Upsilon\text{g}]} \quad > \quad \text{ME} \quad \text{[r\Upsilon\text{g}]}\]
\[
\text{LOE} \quad \text{[t\Upsilon\text{g}]} \quad \text{"tough"} \quad > \quad \text{[t\Upsilon\text{g}]} \quad > \quad \text{ME} \quad \text{[t\Upsilon\text{g}]}\]

The final spirant \(<\text{g}\rangle\) is pronounced \(<\text{j}\rangle\) in Modern English. Along with the replacement of LOE /g/ by \(<\text{j}\rangle\) there was a shortening of the preceding binomic unit, i.e. the second mora dropped.

\[
\text{ME} \quad \text{[\Upsilon\text{g}]} \quad > \quad \text{LME} \quad \text{[\Upsilon\text{f}]}\]
\[
\text{ME} \quad \text{[\Upsilon\text{g}]} \quad > \quad \text{LME} \quad \text{[\Upsilon\text{r}]}\]
\[
\text{ME} \quad \text{[\Upsilon\text{g}]} \quad > \quad \text{LME} \quad \text{[\Upsilon\text{r}]}\]
There are, however, two groups of exceptions. The first comprises the two forms (though) and (through):

EME /bog/ *though* > /bun/ > ME /bun/
EME /bun/ *through* > /bun/ > ME /bun/

Both forms replace LOE /g/ by /f/ thereby shortening the preceding bimoric unit.

ME /bun/ > LME /bun/
ME /bun/ > LME /bun/
ME /bun/ > LME /bun/

There are, however, two groups of exceptions. The first comprises the two forms (though) and (through):

EME /bog/ *though* > /bun/ > ME /bun/
EME /bun/ *through* > /bun/ > ME /bun/

Both forms replace LOE /g/ by /f/ thereby shortening the preceding bimoric unit.

ME /bun/ > LME /bun/
ME /bun/ > LME /bun/
ME /bun/ > LME /bun/

The postulated forms /pol/ and /pruf/ are attested in Early New English and can be found in various Modern English dialects (Brunner 1960: 390). The Modern English pronunciations [bun] and [brun] show a loss of their final spirant. The preceding bimoric unit is retained.

ME /bun/ > LME /bun/
ME /bun/ > LME /bun/

Whereas LME /bun/ develops regularly to Mod E [bun], the LME /brun/ must have escaped the Great Vowel Shift. The loss of final /g/ in these two forms is probably due to matters of prosody.

The other group comprises nouns like Mod E <bough, plough, dough>. In Late Old English these words read /bog/, /plug/, and /dug/. If we apply the rules mentioned above, i.e. replacement of final /g/ by /f/ and concomitant shortening of the preceding bimoric unit, the three Late Old English forms should be pronounced in Modern English as *[baf], *[plaf], and *[daf]. The three nouns in question, however, are pronounced in [baf], [plau], [dou]. These Modern English pronunciations point to Mid Old English forms where the fricative /g/ stood between two voiced segments. Consider the singular and plural paradigm of /bog/ in Late Old English or Early Middle English.

SG PL
/bog/ /bog/
ground. It develops regularly in its intervocalic form into LME /slau/-, the
preinterrogative form should read /slaf/. In Early Middle English we find a noun
/slau/- 'outer skin shed by reptiles'. Its intervocalic and preinterrogative forms
likewise read LME /slau/- and /slaf/. In Modern English both nouns are
spelled /slough/. Concerning the pronunciation the one has generalized the
interrogative form, i.e. /slau/- 'bog', the other has codified the preinterrogative form,
i.e. /slaf/- 'skin'. One can ask why the spelling did not follow the pattern of
utilizing the graphemic differentiation of the two environments, e.g. <slow>
'bog' vs. <slough> 'skin'. The reason is obvious. Instead of homophones we
would have received homographs. OE /slau/- 'slow' would have been likewise
represented by <slow>. It is to be noticed that there exists a spelling variant
<slow> 'marshy, reedy pool' in the U. S. and Canada.

To summarize the foregoing discussion we can say that the graphemic
representation of ME /g/ in Modern English shows a conscious effort to distin-
guish the voiced and the voiceless environment of the spirant on the one hand,
and to prohibit homographs with forms which show no variation of intervocalic
and preinterrogative forms in their paradigm on the other hand. The palatal and
velar variant of ME /g/ between two voiced segments is not represented
graphically up to the differentiation of homophones in some, e.g. /way/ vs.
<weigh>. Between two voiceless segments the Middle English spirant /g/ is
spelled <gh> with one exception, i.e. <trout>. This consistent graphemic
differentiation points to a different development of LOE /g/ in voiced and voiceless
environment, which is supported by independent phonological reasons as
pointed out above. The representation of ME /g/ and its preceding bimoric
unit can be distinguished as marked vs. unmarked. This can clearly be shown
for the sequence /g/gh/, which is represented as <gh> in its unmarked and as
<ghg> in its marked form, e.g. in cases of homophones like <height> vs.
<night> and <sleight> vs. <slight>. This principle has not been used in the
case of ME /mug/, which is spelled as <ough> throughout. A consistent gra-
pemic representation of ME /g/ would have had to pay attention to the
environment of the spirant and to the immediately preceding vocalic nucleus. This
has been done fairly consistently for the sequence /VVVg/ when followed by a voices
less obstructant. In final position the graphemic representation has been hapha-
azard, and that is where we find the difficulties of relating sound and spelling-

REFERENCES

Erdmann, P. 1971. Tiefphonologische Lautgeschichte der englischen Vokale. Frankfurt:
Athenäum Verlag.
Horn, W. und Lehnert, M. 1954. Laut und Leben, englische Lautgeschichte in der neueren

Jordan, R. 1968. Handbuch der mittelenglischen Grammatik: Lautlehre. 3. Aufl. Heidelberg:
Carl Winter.
Taschius.
3 - 13.
7 - 108.