SPOILT FOR CHOICE? <THE> PROBLEM WITH <PE> IN <DE>
PETERBOROUGH CHRONICLE

ALEXANDER BERGS
Heinrich-Heine-Universität Düsseldorf

"Naturalmente, un manoscritto.

1 Introduction
The subject of this paper is the variation of <p>, <ð>, and <th> in the final continuation of the Peterborough Chronicle (PetC-2), i.e. the text commonly held to be the very first example of written Middle English (Lass 1993: 244). Traditionally, Middle English spelling has been regarded as chaotic and unreliable, and Benett & Smithers, editors of a standard anthology of early Middle English verse and prose, still insisted in 1974 that the

"philological value [of the PetC-2] is reduced by a slight admixture, not of scribal divergences from an 'author's' form of ME (since they are virtually of autograph status), but of forms from the standard written language of late OE, which was WS [...] and by a disordered system of spelling." (Benett & Smithers 1974: 374; my boldface)

To my mind, this verdict constitutes an unproductive simplification. 3 Therefore, one of the aims of this paper will be to show that spelling variation in the PetC-2 was in fact a system with structured heterogeneity in the sense of Weinreich, Labov, Herzog (1968) with respect to at least one graphemic variable

1 I would like to thank Derek Britton for his encouragement, his invaluable comments and criticism on an earlier version of this paper. Also, thanks are due to numerous colleagues at the University of Düsseldorf, and especially Kate Burridge for her inspiring and critical remarks on this version. They are all innocent when it comes to faults, of course. Moreover, I would like to thank all the people who commented on a presentation of this paper at the 10th ICEHL, Manchester, August 1998.

2 The following abbreviations will be used in this paper: PetC-2 - Final Continuation of the Peterborough Chronicle; PetC-1 - First Continuation of the Peterborough Chronicle; IAF - initial/final position in the word

3 In the present analysis I follow Milroy's argument that completely free spelling would have led to illegibility. "Therefore, the scribes did not spell 'at will', but according to variable (and historically mixed) conventions. It is our task to attempt to specify the constraints on spelling under which they were working, always admitting that even after we have done this, there may well be residues of apparent randomness that we cannot explain" (1992: 192)
(viz. below). Furthermore, I will try to show that this particular phenomenon opens interesting perspectives for a theory of speaker innovation and language change, as it has been developed by James Milroy (1992 b, c).

However, two simplifying assumptions are made in this paper:

- Although it can be found in Northumbrian early OE texts, <th> had not been used to such an extent before the PetC-2. Spellings of Anglo-Saxon personal names (e.g. <Cuthbert>) and place names (e.g. <norhamptun>) with <th> in Latin charters of the OE time will also be ignored for present purposes. A thorough examination of these spellings is still a desideratum.

- Although <o> probably had "some notion of voice" (Stockwell & Barritt 1961: 79) attributed to it originally, I assume that <o> and <d> were used in equal measures in OE, without regard to phonetic value.

2 The Manuscript, its scribe and our data
A comprehensive discussion of the manuscript in question seems unnecessary for my present purpose (viz. Clark's introduction (1970), Curzan's summary (1996: 301) and Home (this vol.) for this). Suffice it to say that we have to deal with the third part (the entries 1132-1154) of Text E of the Anglo-Saxon Chronicle, preserved in Bodleian MS Laud Misc 636. It can be shown that this part was written in a hand completely different from those of the first two sections and that the entries were made in Peterborough, by a scribe from Peterborough, as a single block in 1155 (Clark, 1970), Curzan (1996). The fact that the scribe was indeed an Englishman is of particular importance at this point. It has often been suggested that irregular spellings are due to Anglo-Norman scribes who were not familiar with English, i.e. 'regular', orthography. Their imperfect knowledge thus caused apparent randomness in ME orthography. Both Milroy (1992a: 193-194) and Clark (1992, esp. 121-125) discuss this problem. They stress the fact that the 'myth of the Anglo-Norman scribe' is an overgeneralized one originated in a prescriptive attitude towards spelling (as in Scrapp 1974) and lead to the neglect of structured variation in Middle English studies (Milroy 1992a: 193-194). Therefore, it is important to point out that the PetC-2 was, in all likelihood, written by a native Englishman.

The quantitative analysis is based upon this single text, the PetC-2, edited and transcribed by Cecily Clark (1970), with about 2700 words and exactly 384 occurrences in all of <p>, <d>, and <th>.

3 Phonology
In early ME <p>, <d>, and <th> were used for the two dental fricative allophones [θ] and [ð]. In determining the phonetic value of the graph in a specific word, I follow the allophonic rule put forward by Lass:

"If in a word were in most dialects voiceless except medially in the foot [...] Thus [v z θ] appeared in native words only preceded by a stressed vowel (followed by an optional liquid or nasal) and followed by an unstressed vowel: they were always voiceless initially, finally and in clusters [...]." (Lass 1992: 41). ¹

This can still be supplemented by John Anderson's analysis of OE fricatives (1988: 92). He claims that for OE [θ] in [θ] [voicelessX] _____ [θ] [voice] (there) only applied "within the word, and where X and Y are non-segmental and X ≠ morpse boundary and Y ≠ foot (or root-initial [...] boundary). This voicing rule, as far as we can say, still applied in 1155. Thus, <p>, <d>, and <th> can take either a voiced or voiceless value, according to their position in the word:

[θ]: <parof> (1137/53); <polf> (1135/23), <unpankes> (1140/52)
<dat> (1137/40), <wid> (1135/11), <norðhæf> (PetC-2 1131/2)
<thear> (1137/6), <warth> (1135/6), <withheld> (Havelock, line 88)
[ð]: <dor> (1137/50), <dæðer> (1135/11), <other> (1137/22)

Therefore, this seems to be a case of variation in "written-language features (W-features)", i.e. variation in features in a linguistic profile (LP) "which carry no contrasting phonetic implications" (McIntosh 1974: 46). Studies in W-features, although quite rewarding, have been much neglected, as McIntosh points out:

"...orthographic variation rarely attracts much interest unless it offers at least the possibility of clues about phonetic variation or unless [...] some variant form is at once striking and rare. It is important here to stress the fact that there is a great deal to be learnt from a thorough examination of those numerous cases of orthographic variation which have no phonetic implications, such as the examples (þth, ðz, ðh, ðv) [...]" (McIntosh 1963: 5; cf. also McIntosh 1974: 46).

While variation in W-features has been described as dependant on temporal and geographical factors (McIntosh 1956: 15, n.16; 1974: 49; Benskin 1982), this

¹ That implies, of course, that phonemisation has not yet taken place in 1155. Following Lass's argumentation, I assume the phonemisation process to have happened during the 13th century, and facing completion around 1450.

² For a very basic introduction to McIntosh's ideas see McIntosh (1956, 1963); his terminology and approach are probably best documented in McIntosh (1974, 1975). For a more recent introduction to the history of graphemics cf. also Bauer (1986: 199-200).
paper will investigate how other factors, e.g. language internal factors, can also influence the choice of graphemic variants.

4 The Data

4.1 Overall occurrences

Of all the 384 occurrences of <th>, <f> and <ð>, <th> is used 60 times, which equals 15.6%. <ð> is used 58 times, i.e. 15.1%. <f> is by far the most popular graph with our scribe, used in 69.3% (266) of all occurrences.

<table>
<thead>
<tr>
<th>Total</th>
<th>&lt;f&gt;</th>
<th>&lt;ð&gt;</th>
<th>&lt;th&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>384 (100%)</td>
<td>266 (69.3%)</td>
<td>58 (15.1%)</td>
<td>60 (15.6%)</td>
</tr>
</tbody>
</table>

Table 1: Total occurrences of <f>, <ð> and <th>

4.2 Specific distribution

The study of structured heterogeneity requires, *inter alia*, two different approaches to the phenomenon. First, the system of tokens has to be described, second, the motivation(s) for the specific structure(s), if at all possible, have to be analyzed and elucidated. Both of this will be done in the next paragraphs.

Here, the first thing to look for is the ‘classical case’ of phonological conditioning. Obviously, none of the variables is solely triggered by phonological principles (cf. the discussion and list of W-features on p. 3). Tables 2-4 show the specific distribution of <f>, <ð>, and <th> with respect to different criteria. The high percentage of <f> and <ð> for the voiceless dental fricative is due to their distribution in the single word. They occur almost exclusively in initial or final position (table 3, table 4). I will turn to the exceptions below. <f> is used word-medially in nineteen cases, fourteen of which also involve voicing. 73.7% of the latter instances of <f> and <ð>, which hints at some kind of lexical restriction or motivation.

<table>
<thead>
<tr>
<th></th>
<th>&lt;f&gt;</th>
<th>&lt;ð&gt;</th>
<th>&lt;th&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>[θ]</td>
<td>251 (94.4%)</td>
<td>53 (91.4%)</td>
<td>18 (30.0%)</td>
</tr>
<tr>
<td>[∅]</td>
<td>15 (5.6%)</td>
<td>5 (8.6%)</td>
<td>42 (70.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>266 (100%)</td>
<td>58 (100%)</td>
<td>60 (100%)</td>
</tr>
</tbody>
</table>

Table 2: Occurrences of <f>, <ð>, <th> representing [θ] and [∅]

<table>
<thead>
<tr>
<th></th>
<th>&lt;f&gt;</th>
<th>&lt;ð&gt;</th>
<th>&lt;th&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/F</td>
<td>247 (92.9%)</td>
<td>52 (80.7%)</td>
<td>18 (30.0%)</td>
</tr>
<tr>
<td>M</td>
<td>19 (7.1%)</td>
<td>6 (11.5%)</td>
<td>42 (70.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>266 (100%)</td>
<td>58 (100%)</td>
<td>60 (100%)</td>
</tr>
</tbody>
</table>

Table 3: Occurrences of <f>, <ð>, <th> in I(initial)/F(final) or M(medial) position.

<table>
<thead>
<tr>
<th></th>
<th>&lt;f&gt;</th>
<th>&lt;ð&gt;</th>
<th>&lt;th&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/F</td>
<td>247 (77%)</td>
<td>52 (78%)</td>
<td>19 (28%)</td>
</tr>
<tr>
<td>MEDIAL</td>
<td>53 (17%)</td>
<td>52 (16%)</td>
<td>6 (9%)</td>
</tr>
<tr>
<td>GRAM</td>
<td>18 (6%)</td>
<td>18 (6%)</td>
<td>42 (63%)</td>
</tr>
<tr>
<td>LEX</td>
<td>322 (100%)</td>
<td>317 (100%)</td>
<td>347 (100%)</td>
</tr>
</tbody>
</table>

Table 4: Individual occurrences of <f>, <ð>, <th> in their respective environments.

67.7% of all [θ]s are graphically represented by <th>, which makes up for only 15.6% of the total number of occurrences of all three graphs. This hints at some phonological conditioning for its use. Alternatively, the high frequency of <th> in word medial position (62.7%, table 4) might have functioned as a triggering principle. But of all the 42 occurrences of <th> in medial position, not a single one represents the voiceless fricative. All 42 tokens are voiced. The nineteen medial occurrences of <f>, on the other hand, include four in voiceless position: the six occurrences of <ð> include one (i.e. <sǐdōn>, which, as a geminate, probably should not be considered). Thus, statistically speaking, medial <th> should have voiceless value in at least three or four cases, if its medial occurrence was solely triggered by its position in the word. Since it does not follow this statistical prediction, we must assume a phonological motivation for the use of <th> in the first place. But how can we account for <th> in I/F, i.e. in voiceless position? The eighteen occurrences of <th> in I/F position are <th>(10), <thɛr>(3), <warh>(4), <warhɛr>(4), <ther>(4), <θeθɛr>(4), <θer>(4), <θɛr>(4).

First, <uwarh> as the only occurrence in final position may be left out of this discussion. One could only hypothesise that <warhɛr> is an analogical formation to <warhɛrɛn>. But this is, because of the lack of data, mere speculation.

Phonotactic conditioning for the rest is also negative. First, there is not a common denominator (e.g. <θɛr>/. . . [+front, +high]), second, all the combinations (i.e. [θ]/. . . [tʃ/ʃ/θ/θː/θːː ......]) can also be found with <f> or <ð>.

There is, however, one explanation that can be called ‘graphemo-pragmatic’ (N.B.: pragmatic in an extralinguistic sense!). In the text we find 34 instances of contracted [θæ], i.e. [θæ], mostly after <t> ([and]), but also with preceding [θæ], [æθ] or [mθ]. This is a fairly straightforward phonological phenomenon that
comes under the heading of ‘assimilation’. The voiceless dental fricative is assimilated to the preceding alveolar plosive and thereby loses its fricative quality. It becomes an alveolar plosive \( \tilde{l} \) It has to be noted, nevertheless, that this assimilation rule is not without exception and that it applies only for some (lexically/syntactically conditioned) items (e.g., it applies only occasionally for \(<\text{pat} >\) or \(<\text{par} >\), but never for \(<\text{parol} >\)). But as the details are not critical for the present argument, I will not go into them here. Suffice it to say, that the overall phenomenon has been described before (cf. Clark 1970: xv) for the \( \text{Ormulum} \), the \( \text{Katherine Group} \) and some OE documents.

The point of interest for us, however, lies in the fact that this phenomenon may have given rise to the digraph \(<\text{th} \). In \(<\text{th} \), just as in \(<\text{ph} \), the \(<\text{h} \) somehow represents a fricative quality that is added to the preceding plosive. In the case of assimilation, the digraph is stripped of this fricative quality, and we are left with the plosive \( [t] = \tilde{t} \). Therefore, I suggest that \(<\text{th} \) in \( \text{the} \) is just a graphemically ‘unpacked form’ of \( \tilde{t} \) in \( \text{the} \). With ‘unpacking’ I mean something very similar to the phonological equivalent, the opposite of assimilation. When the scribe used \( \tilde{t} \) in a non-assimilatory environment, and therefore pronounced it \( [\tilde{t}] \), not \( [t] \), he expressed this difference by using \( \text{th} \) in \( \text{I/F} \) position (i.e. he just added \( \tilde{t} \) as a fricative marker to the \( \text{th} \) as a plosive marker), instead of using the traditional \( \text{ep} \). In a way this is an analogy to the phonological process that links \( [t] \) and \( [\tilde{t}] \). There are, however, concessions to be made to this argument. First, this can also be interpreted as a ‘chicken and egg’ problem. What was first? \( \tilde{t} \) in \( \text{the} \) can be the result of the unpacking of an earlier form \( \tilde{t} \), or \( \tilde{t} \) can be the result of an assimilation of \( \tilde{t} \). As \( \tilde{t} \), however, is a much older form and might have originated in a phonetic writing system, i.e. without reference to \( \tilde{t} \) or \( \tilde{t} \), we can assume that \( \tilde{t} \) is a later development, an unpacked form of \( \tilde{t} \). Second, the development of \( \tilde{t} \) in \( \text{I/F} \) position is doubtless not mono-causal. Several factors might have contributed to it, and the unpacking mechanism is only one of them. I shall, however, entertain the idea that unpacking is a major factor during this development.

Let me now turn to the question: what happened to \( \tilde{t} \)? \( \tilde{t} \) is used in 58 cases. 49 of these are \( \tilde{t} \) (84.5%) - \( \tilde{t} \) occurs only eleven times. The five instances of \( \tilde{t} \) representing the voiced allophone, i.e. in medial position, are \( \text{wurde} \) (1135/23), \( \tilde{s} \text{uide} \) (1135/5), \( \tilde{s} \text{uyde} \) (1137/15), \( \tilde{c} \text{oder} \) (1135/20) and \( \tilde{c} \text{oder} \) (1135/17). The two occurrences in lexical words are \( \text{wurde} \)

\( (1135/23) \) and \( \text{unfrid} \) (1135/17). The rest are \( \text{wode} \) (1135/11) and \( \tilde{c} \text{es} \) (1132/1; 1137/1). All of the latter were also mostly, if not exclusively, spelt with \( \tilde{c} \) in the PeC-1. Probably we can assume a spelling habit at Peterborough in Hockett’s sense: “The word has a culturally prescribed spelling which the scribe has learned from others.” (Hockett [1959], cited in Stockwell & Barritt 1961: 77).²

The overwhelming majority of \( \tilde{c} \) seems to indicate that the use of \( \tilde{c} \) was also lexically motivated. For no obvious reason (since \( \tilde{c} \) was predominantly spelled \( \tilde{c} \) in the PeC-1 and similar \( [\tilde{c}] \) is mostly spelled \( \tilde{c} \)) our scribe preferred this spelling. That just a few other words are also written with \( \tilde{c} \) (viz. above) shows that \( \tilde{c} \) in general played only a marginal role in his orthographic system (or in other words ‘was no longer productive’). This is supported by the fact that \( \tilde{c} \) is not used at all, except in \( \tilde{c} \), after line 1137/15. It seems as if the scribe consciously tried to avoid this graph from that point onwards.³

Let me recap so far. \( \tilde{t} \) was introduced as a graph to mainly represent the voiceless dental fricative allophone \( \tilde{t} \) (at the same time was also current in the system for the same position, so that we can find some unexplained variation at this point (e.g. \( \tilde{c} \text{ep} \) vs \( \tilde{c} \text{ether} \)). It seems, however, that \( \tilde{t} \) was preferred for grammatical words, while \( \tilde{t} \) prevailed in lexical items (tables 4, 7, 10). Furthermore, \( \tilde{t} \) seemed to be used in certain cases in \( \text{I/F} \) position as a result of unpacking the assimilated sound in sequences such as \( \tilde{t} \text{and e} \). \( \tilde{t} \), meanwhile, was used only in very few lexically conditioned cases and culturally prescribed spellings. It was, literally, on the way out. My conclusion is that here we do not find an instance of ‘disordered spelling’, but rather a very orderly, even if idiosyncratic spelling system.

A different interpretation was offered by Betty Phillips in 1995. She investigated the \( \tilde{t} \text{ep} \) spellings in the Peterborough Chronicle and the \( \text{Ormulum} \). Her findings are that ‘the less frequent the word, the more likely it was to contain the innovative \( \tilde{t} \text{e} \) spelling; and word classes varied greatly in their perspexination of \( \tilde{c} \text{ep} \).

² We can assume that the scribe of the PeC-2 was greatly influenced by his predecessor, at least through reading the PeC-1, but maybe also through personal contact (the scribe of the PeC-1 wrote in 1131, our scribe in 1155; so there is ‘only’ 24 years separating them). It is also very likely that they came from the same area and were educated at the same monastery. Clark comments on this: “...as Peterborough monks, its writers were probably natives of the district; for in the late twelfth century and in the thirteenth, established abbeys recruited their monks locally, and this was probably the practice in the earlier twelfth century also.” (Clark 1970: xxvii; cf. also Blake (1992: 12-13) and Bauer (1986: 206-7) on the generation and abandoning of scribal conventions in monastic circles). Or as Lass put it: “Perhaps some significant part of linguistic structure has nothing at all to do with mind or sensibility, but derives simply from a kind of recipient’s inertia. Speakers just have to accept what’s there already”. (Lass 1997: 12).

³ Barchfield (1956: 66-69; 76) was able to describe a similar phenomenon in the \( \text{Ormulum} \).
The distinction between lexical and functional words has always been a very problematic task, and it is not possible to make a clear distinction in all cases. However, in general, lexical words are those that are used to convey meaning, while functional words are those that serve a grammatical function, such as pronouns, conjunctions, and auxiliaries. The distinction is useful in many contexts, but it is not always clear how to draw the line. In this case, the distinction is based on the fact that lexical words are used to convey meaning, while functional words are used to structure the sentence. This distinction is useful in many contexts, but it is not always clear how to draw the line. In this case, the distinction is based on the fact that lexical words are used to convey meaning, while functional words are used to structure the sentence.
speaker/writer. And, as we do not find <th> to such an extent and in such highly constrained environments before 1155, we must assume that this is indeed an act of speaker innovation. Furthermore, this act clearly was capable of influencing the linguistic system. <th> might well have been introduced as a general substitute for thorn and eth at that time. But it was not. <th> came into general use about three centuries later (e.g. Fisiak 1968: 15-16; Jordan 1975: 33 and every other traditional handbook on ME). So the speech (writing) community resisted this incipient change. The only question is: Why did actuation (the introduction of an innovation into the language system) not take place in 1155, but three hundred years later? There is reason to believe that Hugh Candidus or one of his pupils was the author of the Pet-C-2 (Jack Higham, p.c.). But still this cannot be proven, although this fits nicely into our network concept. Furthermore, there was no need for Middle English documents around 1200, which complicated the situation quite a bit and blurs our picture. But even if our data is quite patchy, we should still be able to trace back the development of this innovation somehow, if it had been successful. But no other documents show the same patterns, as we have seen. So we have to assume that the innovation was actually unsuccessful. Nevertheless, the study of speaker innovation implies the analysis of conditions in which such innovation was successful (and resulted in real language change, i.e. constituted actuation) and those conditions in which change or actuation did not take place, where such innovation was unsuccessful.

The first step presented here, a description and analysis of unsuccessful innovation, does not solve any of the major problems of linguistic change as listed in Labov, Weinreich, Herzog (1968). It is only when we have compared this with the successful innovation processes in about 1450 that we can get some more clues about the fundamental problem of actuation, that is the question, how, when, and why an innovation enters into the system and is implemented in an orderly fashion. At the moment we are still looking at the very tiny source of change within the single speaker.

This study was concerned with only one scribe and one text. This is admittedly not very much. But as Curzan mentions, “the examination of only one early English text necessarily imposes certain limitations on the scope of any insight into the language as a whole, the language of the Peterborough Chronicle nonetheless allows interesting and potentially significant conclusions [...]” (1996: 311) about the evolution of <th> spellings in English. Not only because of this, but also because of its fascinating and extremely progressive scribe, is the Peterborough Chronicle a text which truly deserves to be studied in further detail.

References
McIntosh, Angus. 1956. The Analysis of Written Middle English. As reprinted in Middle English Dialectology: Essays on some principles and problems - by Angus McIntosh, M. I. Samuels and Margaret Laing, ed. by Laird, Margaret (ed.), 1-22. Aberdeen: Aberdeen Uni-
versity Press.