

**The Replacement of  
Indigenous Old English Third Person Plural Pronouns  
by Scandinavian Forms:  
A Diachronic Investigation**

Seminar paper

for the *Hauptseminar*

**“Corpus linguistics and the history of English”**

at Justus-Liebig-Universität Gießen

Prof. Dr. Magnus Huber

FirstName SecondName

Street

Postcode Place

Phone

Email

Matr.Nr.: 12345678

Semester (Line of studies)

This term paper was written for one of my seminars in SS 2006. It is excellent both in form and contents and received the mark 1.0 (sehr gut). The term paper is made available here as an example of a very good paper by consent of its author, whose intellectual property it remains.

## Contents

1	Introduction.....	1
2	Theoretical part.....	2
2.1	The relationship between Old English and Old Norse.....	2
2.2	The subject matter.....	2
2.3	Hypotheses.....	4
3	Practical part.....	5
3.1	The research.....	5
3.2	The results.....	7
3.2.1	Absolute numbers.....	7
3.2.2	Percentages.....	9
4	Conclusion.....	12
	List of References.....	14

## List of figures:

Figure 1: <i>Hie, hem</i> and <i>hir</i> in absolute numbers.....	8
Figure 2: <i>They, them</i> and <i>their</i> in absolute numbers.....	8
Figure 3: <i>Hie</i> vs. <i>they</i> in percentages.....	10
Figure 4: <i>Hem</i> vs. <i>them</i> in percentages.....	10
Figure 5: <i>Hir</i> vs. <i>their</i> in percentages.....	11
Figure 6: <i>Hie, hem, hir</i> vs. <i>they, them, their</i> in percentages.....	12

## List of tables:

Table 1: The sub-periods of the HC.....	5
Table 2: The findings of <i>hi, hem, hir</i> and <i>they, them, their</i> in absolute numbers....	7
Table 3: The numbers of words in the sub-periods of the HC.....	9
Table 4: The findings of <i>hie, hem, hir</i> and <i>they, them, their</i> in percentages .....	9
Table 5: The total findings of <i>hie, hem, hir</i> vs. <i>they, them, their</i> in percentages...	11

## List of abbreviations:

EModE	Early Modern English
HC	Helsinki Corpus
ME	Middle English
OE	Old English
<i>OED</i>	<i>Oxford English Dictionary</i> , online edition (see list of references for complete bibliographical reference)
ON	Old Norse
PDE	Present-Day English

## 1 Introduction

In English texts from the Middle Ages, we often find forms like *hie*, *hem* and *hir* used as third person plural pronouns. As we know, these forms have died out completely and are no longer part of present-day Standard English. The third person plural pronouns that are used today – *they*, *them* and *their* – actually go back to Old Norse, the North Germanic language that is the common ancestor of Danish, Norwegian and Swedish. The presence of these forms as well as many other ON borrowings in the English language is the result of the so-called “Viking Age”, which from England’s point of view began with a series of plunders and raids of small villages along the east coast by Danish and Norwegian Vikings in the late 8<sup>th</sup> century A.D., and ended with Danish kings ruling all of England for about 25 years in the 11<sup>th</sup> century.<sup>1</sup> From the year 878, the Danes and Norwegians that had come to England were allowed to settle in an area north of a line running roughly through Chester and London, and in this area (known as the Danelaw) a lot of cultural mixing took place (cf. Jucker 2000: 19). It is probably around this time and in this place that the *th-* forms of the third person plural pronouns were borrowed from ON into Old English.

The aim of this paper is to investigate the relationship between the different forms of the third person plural pronouns in English throughout the Middle Ages. I have tried to find out when the shift from the indigenous *h-* forms to the borrowed *th-* (or *þ-*) forms occurred, using a corpus linguistic approach: Parts of the Helsinki Corpus (HC) were analyzed with the help of the concordance program WordSmith. According to the nature of this research, I have decided to take into account only those parts of the HC that contained texts written before 1500 A.D.

Before we turn towards the investigation proper (chapter 3 of this paper), it will be necessary to have a look at the basic relationship between OE and ON and the nature of the specific borrowing that we are dealing with (chapters 2.1 and 2.2), in order to formulate our working hypotheses.

---

<sup>1</sup> See Baugh and Cable 2002: 92 ff. for a detailed outline of historical events during the Viking Age.

## 2 Theoretical part

### 2.1 The relationship between Old English and Old Norse

The first fact we need to be aware of when comparing the related languages OE and ON is that “[t]he relation between the two languages in the district settled by the Danes is a matter of inference rather than exact knowledge” (Baugh and Cable 2002: 96). Apart from the few historical facts mentioned above, there is in fact very little from which we could gather what cultural life in the Danelaw was really like. The only traces of what happened on the linguistic level are changes in the language itself, and given the relatively small number of surviving OE texts, it is generally difficult to arrive at any satisfying conclusions.

Graddol et al. (1996 : 119) stress the fact that “[i]t is [...] impossible to say whether the Anglo-Saxons and the Vikings found their respective languages mutually intelligible”. This fact has given room to a lot of controversy among experts. Freeborn (1998: 46), for example, thinks it likely that “the two languages were similar enough in vocabulary for OE speakers to understand common ON words, and vice versa, so that the English and Norsemen could communicate”. Björkman (1900: 4f.) takes the idea of mutual intelligibility even a step further by conjecturing that over the course of time ON and OE “were gradually amalgamated into one language which was chiefly of English character but very rich in Scandinavian elements”, by which he of course refers to OE in its later stages. In his introduction to the history of English, Jucker (2000: 24) even briefly refers to OE and ON as “dialects” (implying that the Anglo-Saxons and the Vikings spoke basically the same language). On the other hand, Kastovsky (1992: 328 f.) is convinced that, while a “certain amount of mutual intelligibility probably existed” between ON and OE, the two distinct languages must have been preserved as such, and that “the main vehicle for linguistic influence of the kind at issue here” must have been a fairly widespread bilingualism.

### 2.2 The subject matter

When we consider the nature of the influence of Old Norse on Old English, it is striking that the list of loanwords<sup>2</sup> contains not only lexical items such as nouns,

---

<sup>2</sup> See Baugh and Cable 2002: 99 ff. for a list of words of Scandinavian origin that are still in use in PDE.

verbs and adjectives, but also grammatical items such as pronouns, prepositions and adverbs. Geipel (1971: 14) refers to these as “structure words”, whereas Baugh and Cable (2002: 102) as well as Kastovsky (1992: 320) prefer the label “form words”. However we might call these items, it is a well-known fact that they are usually not borrowed from other languages (cf. Baugh and Cable 2002: 102). This is obviously because their function is highly grammatical, which means that they belong to the basic grammatical “construct” of a language. “Form words” from one language are not often borrowed into another because their semantic correspondents either already exist in the other language,<sup>3</sup> or if they do not, the other language would have to change its grammatical “character” in order to borrow them.

However little we might know about the relationship between the Anglo-Saxons and the Vikings, the fact that grammatical items from ON were borrowed into OE suggests that the two peoples must have had close contact for a long time, as Geipel (1971: 14) conjectures:

The assimilation by one language of such intimate items as 'structure words' (i.e. conjunctions, prepositions and personal pronouns) and even of phonological features from another language, can [...] only take place when the speakers of the two languages live together for a prolonged period as a single, closeknit community.

Moreover, this borrowing of words with grammatical functions also strengthens the hypothesis mentioned above, namely that the two languages were mutually intelligible at least to some extent. Smith (1996: 131) concludes that the loanwords from ON must have been “treated sociolinguistically as equivalent to items of English lexis, and thus available for use within the core vocabulary of the language”, and this is indeed only possible if there is a certain degree of intelligibility.

Now, if this kind of borrowing so rarely takes place, what actually made it happen in this case? One possible reason for the replacement of *hie* by *they* could be the fact that *hie* is fairly similar in form to the OE third person singular pronouns (usually spelled *he* ‘he’ and *heo* or *hio* ‘she’), and this similarity extends to the other cases as well (e.g. the plural genitive *hir* and its spelling variants are easy to confuse with the female singular genitive *hire*, and the plural dative *hem* and its variants look similar to if not the same as the male singular dative form *him*). Thus the Scandinavian *th*- forms were “possibly [...] felt to be less subject to confusion with forms of the singular”, as Baugh and Cable (2002: 102) put it. This theory of the

---

<sup>3</sup> This is clearly the case here, as *hie* and *they* are completely synonymous.

borrowing of Norse pronouns as a means of disambiguation also plays an important role in Smith's study (1996: 130f.), who claims that the "problem" of potentially ambiguous OE third-person pronouns was "solved by the adoption of [...] more distinctive variants" from ON.

One of Smith's further observations is that "*they* advanced [i.e. became widely used in written English] much more rapidly than *their*, *them*" (Smith 1996: 132). Of course, wherever linguistic changes take place over time, a certain amount of synchronic variation is to be expected. In our case, this means that in a lot of mediaeval texts we may expect to find *they* used in the nominative case and the indigenous forms in the other cases. In fact, Blake (1996: 150) states that

[b]y the end of the fourteenth century the use of *they* as the subject was fairly regular, and *their* was the dominant, but not the exclusive, form of the genitive, whereas in the oblique case (*hem*) was as common as *them*.

We should not, by the way, be surprised to find the period of linguistic change extending even far beyond the Norman Conquest, by which time the period of actual contact between English and ON was over. We must remember that we are dealing only with written texts, and that we do not know how different spoken OE actually already was from the language used in the texts when they were written. However, a common feature of linguistic changes is that they first occur in the spoken language and do not make their way into written language until later. Another reason for the surprisingly little number of *th*- pronouns in OE is the fact that most surviving texts from this period were written in the West Saxon dialect, which was spoken in the region of England that was geographically the most distant from the Danelaw, where the *th*- forms probably occurred first.<sup>4</sup>

## 2.3 Hypotheses

According to Meyer (2002: 100), a common reproach against corpus analyses is that they "do little more than simply 'count' linguistic features in a corpus, paying little attention to the significance" of their results. Therefore, what we need to do first is to frame research questions or hypotheses on which to ground our analysis.

---

<sup>4</sup> The studies of Blake (1996: 34; 157) and Smith (1996: 131 f.) have shown that the earliest instances of *th*- forms actually do occur in texts from the northern and eastern regions of England; Smith speaks of a gradual "southward drift" of Scandinavian forms. The inclusion of geographical issues could have been made in a longer paper.

On the basis of chapters 1 through 2.2 of this paper, we may now formulate the following hypotheses: In the texts from the earliest stages of Old English, we will probably find no traces of the *th*- forms. They will probably start occurring sometime after 878 A.D., during the Danelaw period, or possibly even some hundred years later, for the reasons described in subchapter 2.2. Then, over the course of late OE and throughout ME times, the *h*- forms will probably be gradually replaced by the *th*- forms. In the case of *hie* vs. *they*, this process will probably long have been completed by the end of the ME period (as Blake states that *they* was already “common” around 1400 A.D., see the quotation above). *Them* and *their*, on the other hand, will be expected not to climb as rapidly as *they* (again, if we believe Blake, in both cases the *th*- forms will possibly have reached slightly more than 50% of the scale by the end of the Middle Ages).

### 3 Practical part

#### 3.1 The research

The Helsinki Corpus is well-suited for diachronic research, as its texts are subdivided into eleven categories according to the period of time in which the texts were written. Table 1 gives an overview of these sub-periods (from Kytö 1996).

Sub-period	Years covered
OE1	-850
OE2	850-950
OE3	950-1050
OE4	1050-1150
ME1	1150-1250
ME2	1250-1350
ME3	1350-1420
ME4	1420-1500
EmodE1	1500-1570
EModE2	1570-1640
EModE3	1640-1710

Table 1: The sub-periods of the HC

As mentioned in the introduction, this research will be limited to the time up to 1500 A.D., which leaves us with eight sub-periods (OE1 through ME4) to investigate. Now, simply typing “hie” or “they” into WordSmith would have yielded inaccurate

results, since we are not looking for frequencies of spellings, but for frequencies of words that can occur in various different spellings. Michell and Robinson (2001: 18), for example, provide two different spelling variants for each form in their *Guide to Old English*. However, a glance into the *OED* shows that there are in fact between ten and thirty different historical spelling variants for all six of our words.

As a practical example, the twenty-four possible spelling variants of the word *them* found in the *OED* were as follows: <þeZZm>, <þeym>, <þeim>, <theym>, <theyme>, <theim>, <theime>, <þaime>, <þaym>, <þaim>, <þaem>, <taim>, <thaim>, <thaym>, <thayme>, <thaim>, <þæm>, <þam>, <þame>, <thame>, <tham>, <þem>, <them>, and <theme>. In order to make all of these variants searchable, I had to replace all characters not available in ASCII coding (namely <æ>, <Z>, <ð> and <þ>) by their surrogates used in the HC: <+a> for <æ>, <+g> for <Z>, <+d> for <ð> and <+t> for <þ>. The next step was to create a text file containing all the spelling variants in their modified “ASCII spellings” with each variant on a new line, which meant that in the case of *them*, the contents of the text file looked like this:

```
+te+g+gm
+teym
+teim
theym
theyme
theim
theime
+taime
+taym
+taim
+taem
taim
thaim
thaym
thayme
thaime
+t+am
+tam
+tame
thame
tham
+tem
them
theme
```

Creating such text files saves a lot of time: Instead of typing each spelling variant individually into WordSmith and then adding up all the results, WordSmith can simply be told to search for any word in the text file, which in the case of the word *they* means that only one search will have to be run instead of twenty-four.

Now, all of the sub-periods were searched through using one of the six text files (for *hie*, *hem*, *hir*, *they*, *them* and *their*) after the other. During this process I came upon another problem: Some of the instances that were found in the corpus were in fact completely different words and not the variants of the words looked for – e.g. the file for *hir* contained the rare spellings <her> and <here>, and both forms were found in the corpus. However, in nearly all cases, the context of the findings confirmed my suspicion that these represented the third person singular pronoun *her* (the same as in PDE) and the adverb *here*, respectively. In order to attain more accurate results, I decided to eliminate these ambiguous spellings from the text files and have WordSmith work with the now slightly abbreviated word lists.

## 3.2 The results

### 3.2.1 Absolute numbers

The results of these searches were as follows:

	OE1	OE2	OE3	OE4	ME1	ME2	ME3	ME4
<i>Hie</i>	13	1362	2205	677	1248	87	0	0
<i>Hem</i>	6	22	451	255	642	400	726	372
<i>Hir</i>	30	591	1214	362	996	416	344	553
<b>TOTAL</b>	<b>49</b>	<b>1975</b>	<b>3870</b>	<b>1294</b>	<b>2886</b>	<b>903</b>	<b>1070</b>	<b>925</b>
<i>They</i>	0	0	0	0	75	87	1736	1334
<i>Them</i>	0	0	0	4	3	1	56	522
<i>Their</i>	0	0	0	0	29	0	427	436
<b>TOTAL</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>	<b>107</b>	<b>88</b>	<b>2219</b>	<b>2292</b>

Table 2: The findings of *hie*, *hem*, *hir* and *they*, *them*, *their* in absolute numbers

A first glance at table 2 already shows that, as was to be expected, there are a lot of zeroes in the lower left corner of the table, which means that the search for the Scandinavian *th*- forms in early OE has yielded no results. In fact, the *th*- forms do not occur in high numbers before ME1 (i.e. around 1150), which is even later than expected. Furthermore, if we take a look at the upper right corner, we will see that of the three *h*- forms, at least *hie* had been completely replaced by *they* by the end of ME times.

These trends become more easily visible if we turn our data into graphs, compare figures 1 and 2 below.

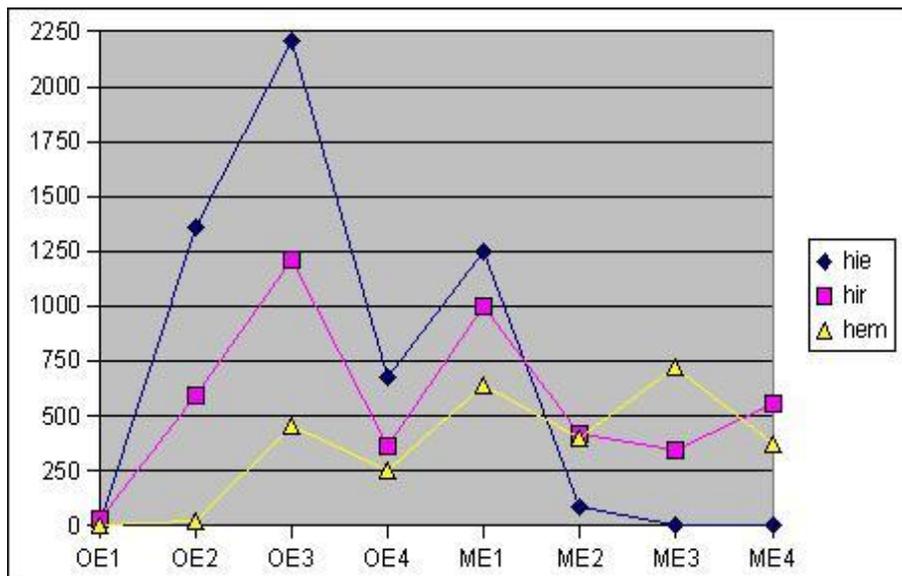


Figure 1: Hie, hem and hir in absolute numbers

It is easy to see that the curves for the *h*- forms in figure 1 mostly reach their peaks in the left half of the chart (the *hem* curve being an exception here; its peak is reached at ME3), whereas the curves for the *th*- forms in figure 2 stay fairly close to 0 on the y-axis and then dart up after ME2. Additionally, the two charts nicely show the relatively quick replacement of the nominative form *hie* through *they*: The blue curve in figure 1 is the only one to reach the 0 line on the y-axis, and in figure 2, the blue curve is clearly the one that contains the steepest rise.

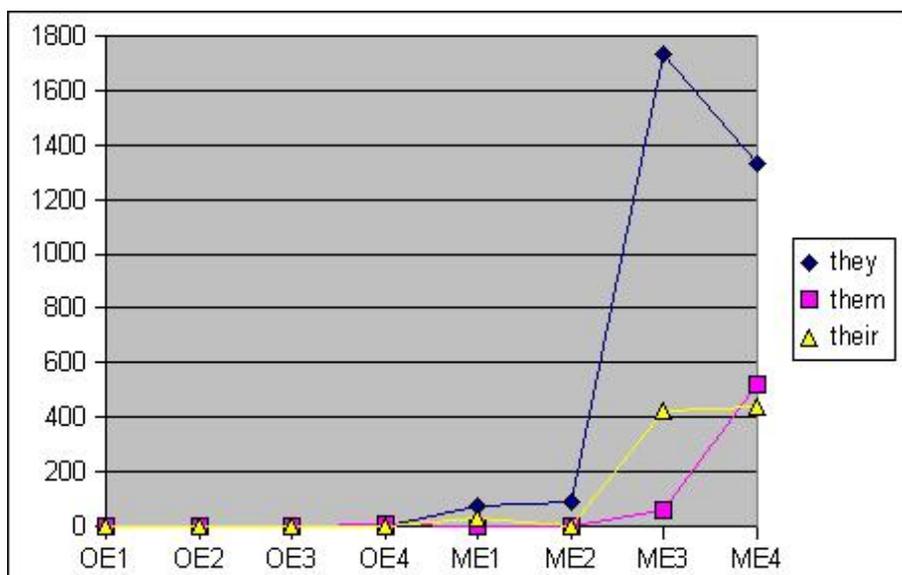


Figure 2: They, them and their in absolute numbers

Another fact that strikes the eye is that, especially in figure 1, the quantities of instances seem to fluctuate from one time period to the next, e.g. all three curves drop from OE3 to OE4, then rise again from OE4 to ME1, and then drop again. This has to do with the fact that we are dealing with the absolute numbers of instances

found in the time periods. From the manual to the HC (Kytö 1996), we learn that the overall number of words in the texts from the different sub-periods is not always the same; in fact, the excerpts from the earlier periods are actually quite unequal in length:

Sub-period	Words
OE1	2 190
OE2	92 050
OE3	251 630
OE4	67 380
ME1	113 010
ME2	97 480
ME3	184 230
ME4	213 850

Table 3: The numbers of words in the sub-periods of the HC

### 3.2.2 Percentages

The only way for us to get around the problem of the unequal amounts of words in the sub-periods (note in table 3 that OE3 contains about four times as many words as OE4 and over a hundred times as many as OE1!), is to shift from absolute numbers to percentages. Table 4 contains the data from table 2 changed into percentages, meaning that for each sub-period the pairs *hie* and *they*, *hem* and *them* and *hir* and *their* always add up to 100%.

	OE1	OE2	OE3	OE4	ME1	ME2	ME3	ME4
<i>Hie</i>	100	100	100	100	94,33	50	0	0
<i>They</i>	0	0	0	0	5,67	50	100	100
<i>Hem</i>	100	100	100	98,46	99,53	100	92,84	41,61
<i>Them</i>	0	0	0	1,54	0,47	0	7,16	58,39
<i>Hir</i>	100	100	100	100	97,17	100	44,62	55,92
<i>Their</i>	0	0	0	0	2,83	0	55,38	44,08

Table 4: The findings of *hie*, *hem*, *hir* and *they*, *them*, *their* in percentages

Based on these percentages we are able to create charts whose graphs will not fluctuate with the length of the different sub-periods, and will therefore reflect the linguistic changes more accurately. In the following figures 3 to 5, purple always represents the now extinct forms while blue represents the forms adopted from ON.

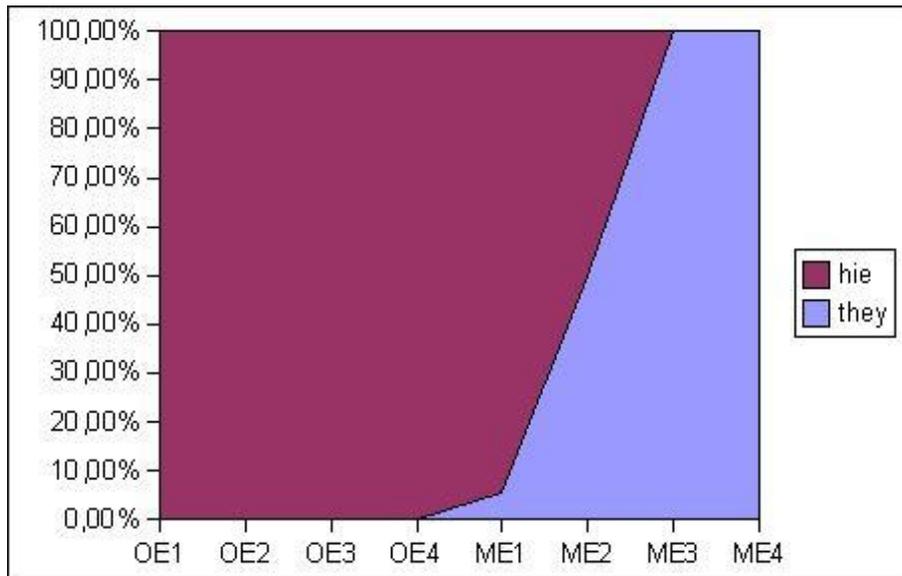


Figure 3: Hie vs. they in percentages

In the *hie* vs. *they* chart (figure 3), both the steep drop of the blue line from figure 1 and the equally steep rise of the blue line from figure 2 combine to form a steady replacement of purple by blue from about ME1 (c. 5%) to ME3 (100%). Note that even though the replacement happens rather quickly, there is an initial phase with only little change happening (OE4 - ME1).

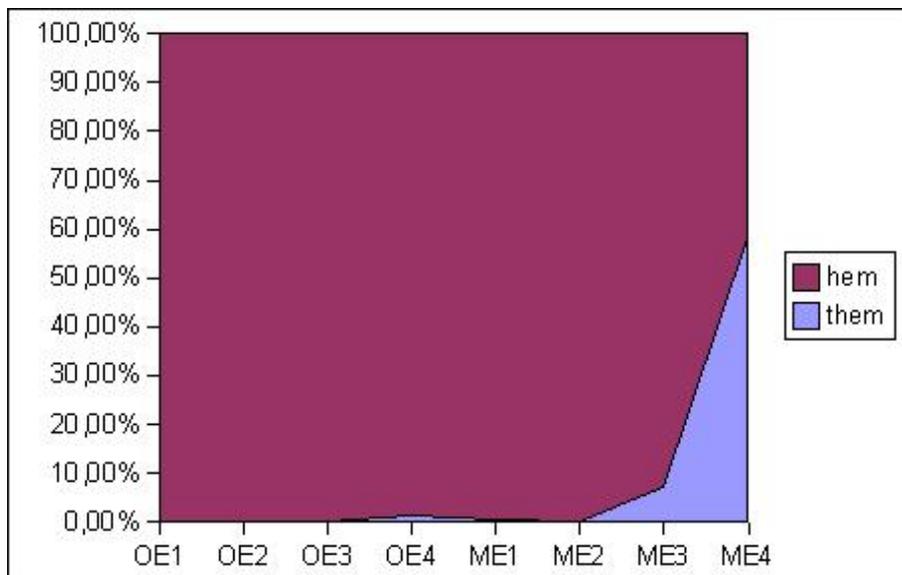


Figure 4: Hem vs. them in percentages

The percentage data for *them* rises just as steeply as that for *they*, and again the change starts taking place slowly and then gains velocity; it almost looks like the curve from figure 3 has been copied into figure 4. The only significant difference between the two charts is that in figure 4 the changes happen later – the replacement of *hem* through *them* in written English actually extends into EModE times.

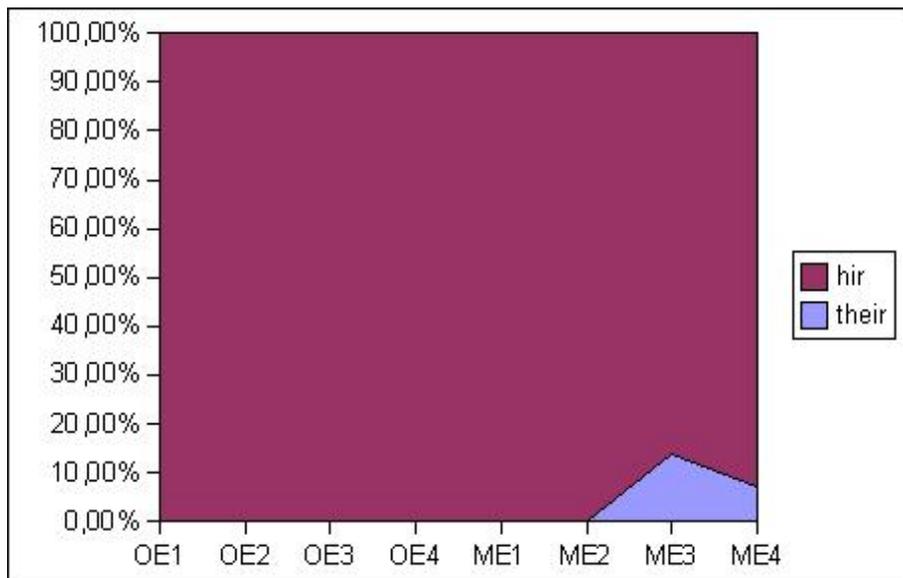


Figure 5: Hir vs. their in percentages

Figure 5 is the least conclusive of all: Though it is clear to see that the changes at first (ME2 - ME3) resemble those in figure 4, one would expect the amount of blue to increase, and not to decline, from ME3 to ME4. An investigation into the following three sub-periods would no doubt shed more light on the problem. It is possible that the replacement in the genitive case took place even slightly later than the replacement in the oblique case.

So far, the most interesting data from table 2 has not been shown in percentages yet. In the following, we will see what happens when we directly compare the total of *h-* forms to the total of *th-* forms in every sub-period (cf. the lines marked “TOTAL” in table 2).

	OE1	OE2	OE3	OE4	ME1	ME2	ME3	ME4
<i>h-</i> forms	100	100	100	99,69	96,22	91,12	32,53	28,75
<i>th-</i> forms	0	0	0	0,31	3,78	8,88	67,47	71,25

Table 5: The total findings of *hie*, *hem*, *hir* vs. *they*, *them*, *their* in percentages

As it is plain to see, these overall results are the most straightforward of all: From OE1 to OE3, the indigenous *h-* forms make up 100% of the findings, and beginning from OE4, there is a continuous shift from *h-* forms to *th-* forms; this shift begins slowly (with a difference of less than one per cent from OE3 to OE4), then increases its rate, and finally slows down again. This behavior is typical of linguistic changes, and if we turn this data into a chart, we get what is probably the main bulk of a characteristic s-curve (see figure 6).

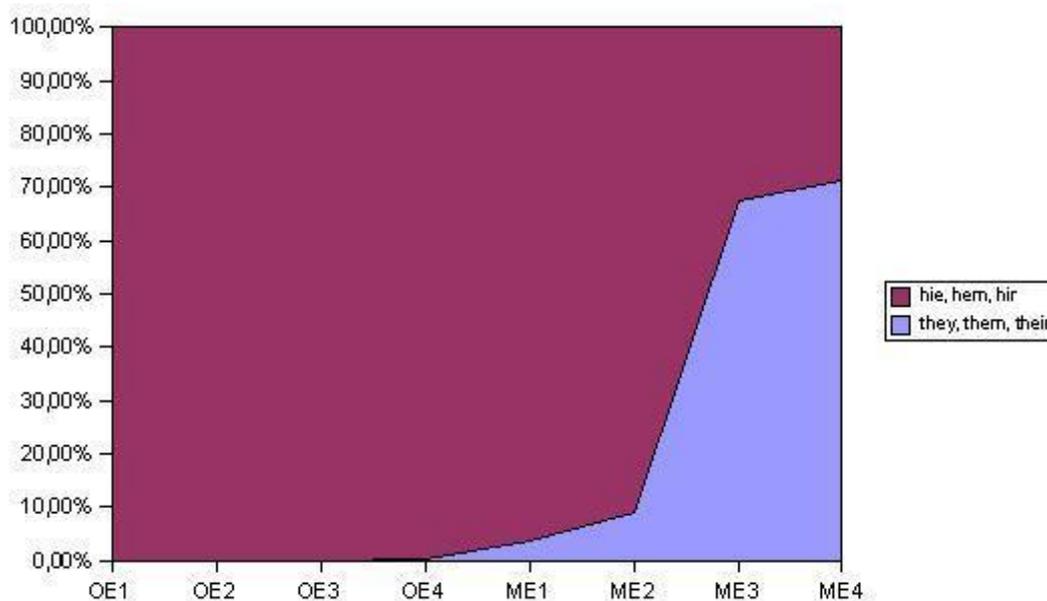


Figure 6: Hie, hem, hir vs. they, them, their in percentages

#### 4 Conclusion

This paper has shown that, although the period of the highest influence of Scandinavian on the English language obviously lasted from c. 900 to c. 1100 A.D., the actual change from *hie, hem, hir* to *they, them, their* in written English largely took place throughout the period of Middle English. However, as table 5 and figure 6 show, this process was not yet completed by the end of the Middle Ages, but over 25% of all third person plural pronouns written in the sub-period ME4 were still indigenous forms.

This, however, is not true for the nominative case (*hie* vs. *they*): As figure 3 nicely demonstrates, *hie* had been replaced by *they* by about 1400 A.D. (ME3). The curve in figure 3 as well as the other percentage curves behave roughly as expected; Blake (1996: 50) seems to have been right in claiming that around 1400, *hem* and *them* were about equally common. The *hir* vs. *their* chart is the only one that has been found not to behave according to Blake's predictions: It seems that around 1400 *their* is in fact not yet the dominant form used in the genitive case.

If we compare these results with the outline of historical events during the Viking Age briefly touched upon in the introduction, we cannot but notice the temporal discrepancies: From the 9<sup>th</sup> century to the 11<sup>th</sup> century, there are hardly any traces of the *th*- forms in the written language yet. In fact, about 300 years passed between the period of actual contact with the Norsemen and the first significant

changes of form in the nominative case (from c. 1150), and with the genitive and dative cases, about 500 years passed before the Scandinavian forms became the standard forms of written English. These discrepancies are partly accounted for by the fact that we are dealing mostly with West Saxon texts (see chapter 2.2 above).

We may conclude with the statement that the replacement of indigenous OE third person plural pronouns by the corresponding Scandinavian forms is not only in itself a surprising fact, but that, given the socio-historical background, it also seems to have taken place surprisingly late, at least in written English. It would be interesting to investigate the chronological development of this specific borrowing in relation that of other words borrowed from ON, as well as to take geographical issues into account, in future papers.

## List of References

- Baugh, Albert C. and Thomas Cable (2002 [<sup>1</sup>1951]). *A History of the English Language*. 5<sup>th</sup> ed. London: Routledge.
- Björkman, Erik (1900). *Scandinavian Loan-words in Middle English*. Part I. Studien zur englischen Philologie 7. Gen. ed.: Lorenz Morsbach. Halle: Max Niemeyer.
- Blake, N. F. (1996). *A History of the English Language*. London: Macmillan.
- Freeborn, Dennis (1998 [<sup>1</sup>1992]). *From Old English to Standard English. A Course Book in Language Variation across Time*. 2<sup>nd</sup> ed. London: Macmillan.
- Geipel, John (1971). *The Viking Legacy. The Scandinavian Influence on the English and Gaelic Languages*. Devon: David & Charles.
- Graddol, David, Dick Leith and Joan Swann (eds.) (1996). *English History, Diversity and Change*. London: Routledge.
- Jucker, Andreas H. (2000). *History of English and English Historical Linguistics*. Stuttgart: Klett.
- Kastovski, Dieter (1992). "Semantics and Vocabulary". In: Hogg, Richard M. (ed.): *The Cambridge History of the English Language. Volume 1: The Beginnings to 1066*. Cambridge: Cambridge UP. 290-408.
- Kytö, Merja (1996). "Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts."  
<<http://khnt.hit.uib.no/icame/manuals/HC/INDEX.HTM>> (16. 8. 2006).
- Meyer, Charles F. (2002). *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge UP.
- Mitchell, Bruce and Fred C. Robinson (2001 [<sup>1</sup>1964]). *A Guide to Old English*. 6<sup>th</sup> ed. Oxford et al.: Blackwell.

*The Oxford English Dictionary* (1989). 2<sup>nd</sup> ed. *OED Online*. Oxford UP.  
<<http://www.oed.com>> (16. 8. 2006).

Smith, Jeremy (1996). *An Historical Study of English. Function, Form and Change*.  
London: Routledge.