The consolidation of *þat* as an invariable relativizer in the history of English[1]

Cristina Suárez, University of the Balearic Islands

Two different invariable relative markers were in use in early English, *þe* and *þat/that*. This paper aims to answer the question of how and why *þat* replaced *þe* as an invariable relativizer in Middle English. To this end I analyse the distribution of invariable relativizers in the relevant periods of the English language (from Old English to late Middle English) as represented in *The Helsinki Corpus of English Texts*. The following variables are examined: (i) the syntactic function of the relativizer, which determines the progression and recession of relativization strategies, following the Accessibility Hierarchy proposed by Keenan and Comrie (1977); (ii) the type of relative clause (whether restrictive or non-restrictive), which conditions the distribution of relativization strategies in particular; (iii) the type of NP antecedent, which also plays a role in the selection of relativizer; (iv) the text type; and (v) dialect. The analysis reveals that *þat/that* starts to replace *þe* very slowly, occupying the environments less favoured by *þe*, that is, those of object and resuming inanimate antecedents. Moreover, I will show that this slow, progressive introduction suddenly evolves into a dramatic change, with *þat* quickly becoming the only invariable relativizer available.

1. Introduction

Different relativization strategies have coexisted over the course of the English language. These include the pronominal relativization strategy, the zero relativization strategy and the invariable relativization strategy.

The most frequently used of these has been the invariable relativizer, represented in early English (Old and early Middle English) by *þe* (example (1)), which from Middle English was wholly replaced by *þat*[2] (example (2)), precursor of Present-day English *that*, as in (3), which was itself already in existence in Old English, but used with considerable less frequency.

(1) Gemunon we symle þæt we þa god   don [rc þe us Godes bec lærað]
    Remind   we ever that we the good things do   rel us God's books teach
    "Let us be always mindful that we do those good things that God's books teach us"
                                                      [Q O2/3 IR HOM BLICK 6: 139]

(2) Morgan wolde haue hade alle þe lande fram bigende Humber, [rc þat
    Morgan would have had all the land from beyond Humber   that
    Conedag helde]
    Coneday held
    'Morgan wanted to have had all the land from beyond de River which Coneday
    held.'
                                                      [Q M3 NN HIST BRUT3 21]

(3) I am reading the book [rc that I've always wanted to read].

Much previous work in this area has focussed on the distribution of relativization strategies in different periods of English (Quirk 1957; Jack 1975, 1978; Dekeyser 1984, 1986; Rissanen 1991; Aarts 1993; Yamashita 1994; Guy and Bayley 1995; Ball 1996; Stein 1998 to mention a few) and different varieties of the language (Ihalainen 1980; Van den Eynden 1993; Sigley 1997; Tottie and Rey 1997; Alsagoff and Lick 1998; Gisborne 2000; Tottie and Harvie 2000; Geisler, 2002; Peitsara, 2002; Tagliamonte, Smith and Lawrence 2005; among many others), as well as on the expansion of *wh-* pronominal relativizers from late Middle English onwards at the expense of Old English pronominal relativizers *se/seo/þæt* (Meier 1967; Rydén 1983; Nevalainen and Raumolin-Brunberg 2002; Suárez 2008a; etc.). Less often addressed has been the consolidation of *that* as the only invariable relativizer in late Middle English, ousting *þe*, the favourite relativizer in Old English, from the relativization system (McIntosh 1947-48; Kivimaa 1966).

---

[2] The invariable relativizer of early English will be spelled *þat* throughout the paper, although it was recorded with many different spellings throughout the history of English.

From Old English onwards, *þat* existed as an invariable relativizer but was very sporadically used as such. A common view is that relativizer *þe* was phonologically weaker than the already existing *þat*, and this led to its disappearance from the system earlier (Fischer et al. 2000: 91). Additionally, the displacement of *þe* by *þat* as the invariable relativizer has been linked to the use of *þat* as the general subordinator of the English language, already used in OE as a subordinating particle in non-relative dependent clauses (especially complement or content clauses and clauses of purpose and result) (Fischer 1992: 294; Rissanen 1997).

My purpose here is to shed light on the consolidation of *that* as the only invariable relativizer in English and to provide an answer to the question of how the consolidation of *þat* as an invariable relativizer proceeded. In order to answer this question, I report here on a study set up to investigate the use and evolution of *þe* and *þat/that* as invariable relativizers, from Old English to late Middle English (1420), when *that* had become the only invariable relativizer available in the English language. I will analyze the distribution of these relativizers, taking into account a number of linguistic factors (both internal and external) that have been reported in the literature to influence the choice of the relative word. The ultimate goal of this paper is to reconstruct the conditions under which *þe* came to be replaced by *þat/that* in the history of English.

2. Data and methods

The elements under examination here are invariable relativizers in Old and Middle English. During these periods, two relative words were used in this context, namely *þe*, as in (1) and *þat* as in (2), their distribution resulting from personal choice and/or linguistic factors which may have influenced the selection of one over the other. In this study I will consider the two elements as variants of the same relativization strategy, since both fulfil the requisites expected of invariant relativizers.

The 'invariable relativization strategy', also known as 'relative subordinator strategy' (Comrie 1981: 151, Givón 1993, Comrie 1998), represents one of the major relativization strategies typologically speaking, since it is among the most frequent means of marking

relativization cross-linguistically.[3] Items used for this relativization strategy are characterized by being indeclinable, and not marked for gender, number or case. Unlike pronominal relativizers, they lack genitive forms (*contra* Seppänen and Kjellmer (1995) who report examples such as *The dog that's leg was run over* as examples of *that* explicitly marked for the genitive) and are not marked for animacy. Additionally, these elements are usually found as complementizers or subordinators in the language, as is the case with *that* and also *þe* in earlier English, the former used to introduce complement clauses throughout the history of the language and as a pleonastic marker of subordination up to early Modern English (Rissanen 1997) and the latter frequently used to mark complement clauses and subordinate clauses of purpose and result in Old English (Fischer 1992: 294).

The data used in the present study have been extracted from the *Helsinki Corpus of English Texts*. As Table 1 shows, the sample is taken from texts from Old English (950-1150) (O3, O4) and Middle English (1150-1420) (M1, M2 and M3).

Table 1. Description of the corpus

Period	Sub-periods	No of words	No of tokens
OE	O3 (950-1050)	36,630	539
	O4 (1050-1150)	47,445	576
ME	M1 (1150-1250)	75,800	1,184
	M2 (1250-1350)	4,489[4]	81
	M3 (1350-1420)	57,774	741
	TOTAL	222,238	3,121

---

[3] The other major types of relativization strategies include the 'non-reduction stratety', 'anaphoric pronoun' or 'pronoun-retention strategy', 'relative pronoun strategy' and 'gap strategy' (see Keenan 1985, Comrie 1998).

[4] The low number of words in this period has to do with how the little material has been preserved from early Middle English. In the *The Helsinki Corpus* only 3 texts (*Select English Historical Documents of the Ninth and Tenth Centuries, The Bee and the Stork* and the *Kentish Sermons*) from this period are classified as non-translations. Also, one of these (the *Kentish Sermons*), although not classified as a translation in *The Helsinki Corpus*, is considered in the literature to be a translation from French.

The corpus comprises approximately 222,000 words and has rendered 3,121 examples of relative clauses introduced by an invariable relativizer. The last period included in the analysis is late Middle English (M3), since by then (1350-1420) the relativizer *þat/that* was the only remaining invariable relativizer in use, thus removing the need to examine texts dating from any later than this time. In order to keep the corpus as homogeneous as possible, only original prose texts have been selected. Translations (as classified in the *Helsinki Corpus*) were discarded, so as to avoid any potential influences from language contact.[5] This same selection criterion, however, had the additional effect of upsetting the balance between sub-samples, a problem which I have corrected by normalizing the frequencies per one-thousand words.

   Invariable relativizers have always been present in the English language and have in general been used more frequently than any other relativization strategy, as Figure 1 below (based on Table 2) illustrates.

Table 2. Distribution of relativizers (adapted from Suárez 2004: 216, 224, 226)

	950-1050	1150-1250	1420-1500
Invariable	539 (81.2%)	1184 (95.1%)	741 (83.4%)
Pronominal	124 (18.7%)	54 (4.3%)	143 (16.1%)
Zero	1 (0.1%)	7 (0.6%)	4 (0.4%)
TOTAL	664	1245	888

---

[5] See Taylor (2008) for a recent analysis on the effects of translation on Old English texts.
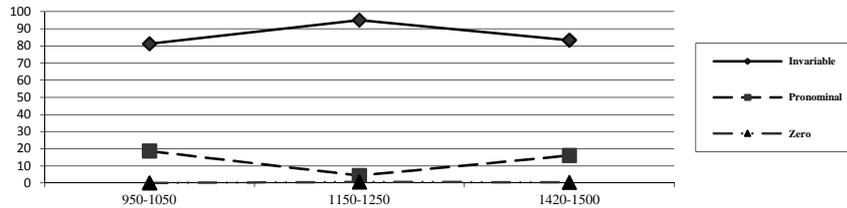
Figure 1. Distribution of relativizers

Recent studies on the distribution of relativizers in different British dialects (Herrman 2005: 24-28) confirm that this is still the norm, and is also becoming the trend in New Englishes, particularly in spoken language, as reported by Gut (2009) with respect to Kenyan, Jamaican and Singapore Englishes.

Table 3 presents the distribution of invariable relativizers in the corpus under analysis: it includes raw numbers and their normalized frequencies per one thousand words.

Table 3. Chronological distribution of pronominal relativizers (normalized frequencies per 1000 words)

	O3 (950-1050)	O4 (1050-1150)	M1 (1150-1250)	M2 (1250-1350)	M3 (1350-1420)	TOTAL
Þe	526 (14.3)	551 (11.6)	846 (11.2)	1 (0.2)	–	1924
Þat	13 (0.3)	25 (0.5)	338 (4.5)	80 (17.8)	741 (12.8)	1197
TOTAL	539	576	1184	81	741	3,121

(χ^2, p <.0001, in the comparison between Old English and M1)

Þe-relative clauses are by far the most numerous group of relative clauses in late Old English (O3, O4) and early Middle English (M1). Nevertheless, by early Middle English (M1), there were already signs of a decrease in the frequency of *þe* relative clauses in favour of

relative clauses introduced by *þat* (the same tendency is also found sporadically in English earlier than this). By early Middle English, the frequency with which *þat* was used as an invariant relativizer had increased considerably (Kivimaa 1966: 133-136), an increase that was to become dramatic in late Middle English (1150-1250), when *þat* was by far the more important relativizer (McIntosh 1947-1948: 73; Kivimaa 1966: 134; Fischer 1992: 196; Fischer et al. 2000: 93). This sudden increase is depicted in Figure 2 (based on Table 3), which demonstrates graphically the progress of *þat* to become the most important relativizer from the second half of the thirteenth century on; and, similarly, how *þe*, in recession from the twelfth century onwards, became almost invisible. By the fifteenth century, *þat* was the only invariable relativizer.
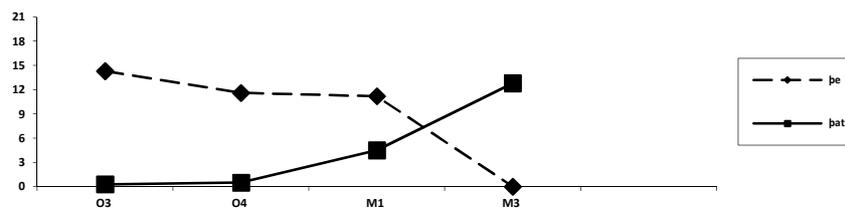

Figure 2. Competition between *þe* and *þat* in the history of English (normalized frequencies per 1000 words)[6]

Here the use of an innovative invariable relativizer in the English language—in this case *þat*—grew in frequency and progressively pushed *þe* away, as demonstrated by the decline of *þe* in the 12th century, being compensated by the increase of *þat*. As happened in the case of *wh*-relative pronouns, the use of which rose as that of the demonstrative relative pronouns receded (Suárez 2008a), the gap left here by the declining invariable relativizer *þe* started to be occupied by *þat* in early Middle English, eventually becoming the only invariant

---

[6] M2 has been excluded from the graphs because the low number of examples—related to the scarcity of available material—hides the clear competition between the two relativizers (see footnote 4).

relativizer by late Middle English. This ultimately led to the complete disappearance of *þe* as a relativizer and hence the prevailing state-of-affairs in Modern Standard English. Around the thirteenth century, invariant *þat* was the norm everywhere, potentially occurring with any kind of antecedent and in any context, in most dialectal areas, with both restrictive and non-restrictive relative clauses, and with both animate and inanimate antecedents. The specialization of invariant *that* in restrictive relative clauses did not take place until around the fourteenth century, when *wh*-relative pronouns began to occur in the English language with some frequency (Mustanoja 1960: 197; Fischer 1992: 297; Fischer et al. 2000: 93). Before this consolidation, there existed variation between *þe* and *þat*; the factors which may have conditioned this variation will be addressed in the following section.

3. Factors conditioning variation

3.1. Syntactic function of the relativizer
The syntactic function of the relativizer has generally been considered one of the main factors governing the choice of relativizer. Syntactic function, in fact, is the variable most frequently mentioned in the literature on the expansion of *wh-* relativizers to occupy the space left by pronominal relativizers in earlier English (Suárez 2008a: 345-346) and may also be of importance in the case of invariant relativizers. For my analysis I have adopted (and adapted) Keenan and Comrie's 'Accessibility Hierarchy' (1977) so that three different categories are distinguished. The 'Accessibility Hierarchy' is a hierarchical ordering of noun phrase positions illustrated as follows (Keenan and Comrie 1977: 66):

SU > DO > IO > OBL > GEN > OCOMP
(where SU stands for ''subject'', DO for ''direct object'', IO for ''indirect object'', OBL for ''oblique'', GEN for ''genitive'' and OCOMP for ''object of comparison'')

The syntactic positions included in the hierarchy are ordered from the most to the least accessible, that is, the easiest and hence most

frequently relativized on the left, and the most difficult and less frequently relativized syntactic functions on the right.

For the present paper, the Accessibility Hierarchy was simplified and the following syntactic functions were distinguished: subject (S), as in example (4), object (examples (1), repeated here as (5)), and oblique (Obl), which refers to relativizers that function as prepositional phrases (all of which are stranded, as must be the case with invariable relativizers in the periods under analysis), as in example (6). The categories GEN and OCOMP were not included since no examples were found of invariable relativizers being used in either of these types of positions.

(4) se [_RC_ þe ne can   þa beorhtnesse þæs   ecan   leohtes], se bið blind
he   rel neg knows the brightness   of the eternal light   he is blind
'he who does not knows the brightness of the eternal light is blind'
                                          [Q O2/3 IR HOM BLICK 2: 62]

(5) Gemunon we symle þæt we þa god   don [_RC_ þe us Godes bec lærað]
Remember we ever   that we the good things do   rel us God's books teach
"Let us be always mindful that we do those good things that God's books teach us"
                                          [Q O2/3 IR HOM BLICK 6: 139]

(6) We agen to understonden hwer boð þe   wepne [_RC_ þet adam wes mide forwunded]
We ought to understand   what are the weapons   rel adam was with wounded
"We ought to understand what the weapons are that Adam has been wounded with."
                                          [Q M1 IR HOM LAMB8: 83]

The question of which positions in a language may be subjected to relativization is closely linked to the degree of explicitness of the relativization strategy. Relative clauses which are introduced by

invariable relativizers allow a narrower range of positions to be relativized in comparison with those introduced by pronominal relativizers. With this in mind, we would expect invariable relativizers to be able to relativize only the most accessible positions in the hierarchy. Table 4 shows the distribution of invariable relativizers in the periods under analysis:

Table 4. Invariable relativizers and syntactic function (normalized frequencies per 1000 words)

		S	O	Obl	TOTAL
O3 (950-1050)	*þe*	369 (10.1)	133 (3.6)	24 (0.7)	526
	þat	8 (0.2)	1 (0.02)	4 (0.1)	13
O4 (1050-1150)	*þe*	352 (7.4)	138 (2.9)	61 (1.3)	551
	þat	11 (0.2)	11 (0.2)	3 (0.02)	25
M1 (1150-1250)	*þe*	593 (7.8)	183 (2.4)	71 (0.9)	846
	þat	173 (2.4)	135 (1.8)	30 (0.4)	338
M2 (1250-1350)	*þe*	1 (0.2)	-	-	1
	þat	58 (12.9)	19 (4.2)	3 (0.7)	80
M3 (1350-1420)	*þat*	487 (8.4)	191 (3.3)	63 (1.1)	741
TOTAL		2,051 (9.2)	811 (3.6)	259 (1.2)	3,121

S=subject; O=object; Obl=oblique.

The results show that, with respect to late Old English, the indeclinable relativizer *þe* is most frequently used when it appears as the subject or the object of the relative clause, as Traugott (1992: 226) and Kivimaa (1966: 28) have previously shown. In other words, the most frequently relativized positions are those which correspond to the roles highest on the 'Accessibility Hierarchy', as the behaviour of *þe* demonstrates. The

same situation is observed in the distribution of *þat* (subjects and objects being the most commonly relativized positions), although the number of examples in this case is extremely low.[7] The most notable aspect of *þat*'s behaviour, however, is that the expected hierarchy is not respected: objects are as frequently relativized as subjects. In relation to early Middle English, Kivimaa (1966: 135) and Jack (1975[8]: 104; 1988[9]: 49) have observed that the invariable relativizer *þe* is used almost exclusively in the subject slot; such exclusivity was not observed in the case of *þat*. The correlation between the syntactic function of the relativizer and the choice of relativizer has also produced some interesting and significant results (χ^2, p <.0001) in relation to my corpus of early Middle English (see Table 4). All syntactic roles are more frequently performed by the invariable relativizer *þe*, in particular, though not exclusively, that of the subject (just as Kivimaa and Jack had previously found). The most notable development in this period (1150-1250) involves *þat* whose importance as an invariable relativizer rises and which, although more commonly used as subject, also achieves an extremely high frequency as an object (in comparison with the other periods).

In its expansion *þat* is stronger—and ousts *þe* earlier—as object, a position in which *þe* is proportionally less frequent, as illustrated in O4 and M1, illustrating a similar process to that observed in the substitution of *se* pronominal relativizers by *wh-* words. The unavailability of relevant data from M2 noted above does not allow us to confirm this tentative hypothesis.

3.2. Restrictiveness of the relative clause
Another important variable affecting the distribution of relativizers is that of the restrictiveness of the relative clause, according to which relative clauses are divided into restrictive and non-restrictive types. The hypothesis being examined here is whether the distribution of invariable relativizers correlates significantly with the distinction

---

[7] No significance tests were applied to Old English due to the low number of examples in some of the *þat* cells.

[8] The conclusions are based on an analysis of only one text, the *Ancrene Wisse*.

[9] The conclusions are based on an analysis of only one text, *Layamon*.

between restrictive and non-restrictive relative clauses. The distribution of relativizers in Present-day English changes depending on this variable: restrictive relative clauses can be introduced by any of the available items, namely *zero*, *that* and *wh-* pronouns, and non-restrictive relative clauses favour (indeed almost require) a *wh-* word; only very rarely are they introduced by the relativizer *that* (Jacobsson 1994; Huddleston and Pullum 2002: 1059),[10] and never by the relativizer *zero*. Throughout the history of the English language, invariable relativizers (as well as other relativization strategies) have been used to introduce both restrictive and non-restrictive relative clauses, though more commonly in relation to restrictive relative clauses. Scholars such as McIntosh (1947-1948: 79 fn. 19), Mitchell (1985: §2283) and Traugott (1992: 223) are not as categorical on this point but, broadly speaking, are of much the same opinion. Jack (1975: 106-107, 1988: 52-53), on the other hand, denies that the choice between the two major relativizers—*þe* and *þat*—is a function of the type of relative clause; nonetheless, his analysis still indicates a higher proportion of invariable *þe* in non-restrictive relative clauses, and the prevalence of *þat* in restrictive relative clauses (Jack 1988: 53). Table 5 sets out the distribution of relativizers according to the type of relative clause.

---

[10] See example "He wants less freedom, that he always thought was outward motion, turns out to be this inner dwindling" (Jacobsson 1994: 186).

Table 5. Invariable relativizers and type of relative clause (normalized frequencies per 1000 words)

		RRC	NRRC	TOTAL
O3	*þe*	422 (11.5)	104 (2.6)	526
(950-1050)	*þat*	5 (0.1)	8 (0.2)	112
O4	*þe*	360 (7.6)	191 (4.0)	551
(1050-1150)	*þat*	23 (0.5)	2 (0.04)	25
M1	*þe*	694 (9.1)	279 (3.7)	973
(1150-1250)	*þat*	152 (2.0)	59 (0.8)	211
M2	*þe*	1 (0.2)	-	1
(1250-1350)	*þat*	66 (14.7)	14 (3.1)	80
M3 (1350-1420)	*þat*	587 (10.2)	154 (2.7)	741
TOTAL		2,310 (10.4)	811 (3.6)	3,121

The results show that the choice of *þe* does not relate to any type of relative clause in particular, but is simply a reflection of the general distribution of higher frequency in restrictive than in non-restrictive relative clauses across all the periods, and especially in late Old and early Middle English. During these periods of coexistence, *þat* exhibits a higher tendency to introduce restrictive relative clauses, and this is statisticially significant in Old English (χ^2, p <.0001). However, from M2 onwards, when *þat* becomes the only invariable relativizer, it inherits the behaviour of *þe* in O4 and M1, so that the number of non-restrictive relative clauses introduced by *þat* also increases. This new distribution leads us to the tentative conclusion that when *þat* substituted *þe*, it adopted *þe*'s distribution in relation to the type of relative clause, thereby reinforcing the substitution process already observed in terms of syntactic function.

3.3. The NP antecedent

One of the most complex variables hypothesized to affect the distribution of relativizers is the NP antecedent. According to Jack, the preference of a relativizer, in particular *þe* and *þat* in early Middle English, is to be interpreted in light of the nature of the antecedent (Jack 1988: 53). The influence of the antecedent on the distribution of relativization strategies, however, is denied by Romaine, who states that "the effect of different types of antecedent (grouped according to certain characteristics or features of the modification which precedes the head noun) is in most cases negligible" (Romaine 1982: 143), at least in her investigation of relative clauses in Middle Scots (c. 1530-1550).

In the current study, NP antecedents have been analyzed for the following two characteristics:

- Category/form of antecedent
- Type of antecedent: animate vs inanimate

Regarding the category of the NP antecedent, the aim is to observe whether there exists any correlation between the category of the NP antecedent and the selection of *þe* and *þat* in the periods under analysis. For the analysis of this variable I reduced the form of the NP antecedent to the following six categories: count nouns in the singular (example 7), count nouns in the plural (example 6), non-count nouns (example 8), pronominal items (example 4), proper names (example 9) and a miscellaneous category which would comprise every other antecedent (basically nominalized adjectives (example 5) and coordinate NPs (example 10).

(7) þis is þe miracle [$_{RC}$ þet þet godspel of te dai us telþ]
    this is the miracle   rel the gospel   of today us tells
    "This is the Miracle that today's gospel tells us."
                        [Q M2 IR HOM KSERM: 218]

(8) Þis   sceal   wyð   eagena dymnysse, [rc þt grecas   nemniað
      glaucomata]
      This   schall against   eyes   dimness   rel Greeks   call
      glaucomata
      'This shall be used against the eyes' dimness, which Greeks call
      glaucomata.'

                                                    [Q MX/1 IS HANDM PERI 96]

(9) Iohannes eac, se   fulluhtere, [rc þe Crist   gefullode]
      John   also the Baptist   rel Christ baptized
      'Also John the Baptist, who baptized Christ.'

                                                    [Q O3 IR RELT LSIGEF 20]

(10) Somtyme   hym þink   þat it is paradis or heuen, for diuerse
        wonderful
        Sometimes him seems   that it is paradise or heaven for diverse
        wonderful
        swetnes   and counfortes, ioyes & blessid vertewes [rc þat he
        findeþ þer-in]
        sweetness and comforts   joys & blessed virtues   rel he
        finds   therein
        'Sometimes it seems to him that it is paradise or heaven because of
        the diverse wonderful sweet things and comforts, joys and blessed
        virtues which he finds in it.'

                                                    [Q M3 IR RELT CLOUD 68: 123]

The bearing of the antecedent on the choice of relativizer in Old
English has been studied by Kivimaa (1966), Mitchell (1985) and
Traugott (1992), and among the few descriptive notes we find that the
invariable relativizer *þe* is favoured when the antecedent is modified by
a pronominal element (Kivimaa 1966: 44; Mitchell 1985: §2270,
§2270; Traugott 1992: 226), especially in the singular (Traugott 1992:
226) (with the exception of the indefinite pronoun *eall*, which when
modified by a relative clause, is invariably introduced by *þat* (see
Mitchell 1985: §2263)). The distribution of relativizers *þe* and *þat*
according to the antecedent they resume is shown in Table 6:

Table 6. Invariable relativizers and category of NP antecedent (normalized frequencies per 1000 words)

		Count (sg)	Pronoun	Count (pl)	Non-count	Proper name	Other	TOTAL
O3 (950-1050)	*þe*	130 (35.5)	202 (55.1)	145 (39.6)	26 (7.1)	16 (4.4)	7 (1.9)	526
	þat	6 (4.0)	2 (0.5)	1 (0.3)	4 (1.1)	-	-	13
O4 (1050-1150)	*þe*	188 (39.6)	90 (19.0)	159 (33.5)	61 (12.8)	47 (9.9)	6 1.3)	551
	þat	4 (0.8)	16 (3.4)	2 (0.4)	2 (0.4)	1 (0.2)	-	25
M1 (1150-1250)	*þe*	364 (48.0)	214 (28.2)	180 (23.7)	58 (7.6)	21 (2.8)	9 (1.2)	846
	þat	178 (23.5)	94 (12.4)	36 (4.7)	22 (2.9)	4 (0.5)	4 (0.5)	338
M2 (1250-1350)	*þe*	-	1 (2.2)	-	-	-	-	1
	þat	18 (40.1)	25 (55.7)	23 (51.2)	12 (26.7)	1 (2.2)	1 (2.2)	80
M3 (1350-1420)	*þat*	355 (61.4)	172 (29.8)	111 (19.2)	56 (9.7)	40 (6.9)	7 (1.2)	741
TOTAL		1,243 (55.9)	816 (36.7)	657 (29.6)	241 (10.8)	130 (5.8)	34 (1.5)	3,121

The results in Table 6 show those differences in the two periods of Old English and early Middle English (M1). In O3 (950-1050), although invariable relativizer *þe* is used in all environments, it is clearly favoured when it resumes pronominal items and count nouns in the plural. Relativizer *þat* is preferred in combination with count nouns in the singular (either masculine or feminine),[11] but we must bear in mind that the low number of examples here does not allow us for definite conclusions. Regarding O4 (1050-1150), relevant differences are also

---

[11] Otherwise it could be considered a pronominal relativizer gender-agreeing with the antecedent.

observed (χ^2, p <.0001). *Þe* is favoured with count nouns, both in the singular and in the plural, and is much less frequently used with pronominal items, which becomes the favourite NP antecedent of the emergent *þat*. This is reinforced by the results shown for *þat* in M2 (1250-1350). Regarding M1 (1150-1250), a very similar distribution is observed in *þe* and *þat*, both favoured with count nouns in the singular and with pronouns, thus mirroring the general distribution of the category of antecedents in the whole corpus. The distribution in this case is also significant (χ^2, p <.0001).

In terms of the animacy of the antecedent, in the codification of the results from my corpus, 'animate' (example 9) and 'inanimate' (example 7) antecedents were distinguished.

There is no consensus in the literature as to the influence of the animacy of the NP antecedent on the choice of relativizer. As noted above, from the results of her analysis of relativizers in Middle Scots (c. 1530-1550), Romaine concludes that the influence of the antecedent "is in most cases negligible" (1982: 143), and, more categorically, that "the animacy of the antecedent has virtually no effect in determining which form of the relative will occur" (1982: 142).

Table 7. Invariable relativizers and animacy of the antecedent (normalized frequencies per 1000 words)

		Animate	Inanimate	TOTAL
O3 (950-1050)	*þe*	321 (8.8)	205 (6.0)	526
	þat	1 (0.03)	12 (0.3)	13
O4 (1050-1150)	*þe*	323 (6.8)	228 (4.8)	551
	þat	4 (0.8)	21 (0.4)	25
M1 (1150-1250)	*þe*	429 (5.6)	417 (5.5)	846
	þat	117 (1.5)	221 (2.9)	338
M2 (1250-1350)	*þe*	1 (0.2)	-	1
	þat	42 (9.3)	38 (8.5)	80
M3 (1350-1420)	*þat*	329 (5.7)	412 (7.1)	741
TOTAL		1,567 (7.1)	1,554 (7.0)	3,121

The results from Table 7 demonstrate that at least some relativizers are sensitive to the animacy of the antecedent. The results show that *þat* is preferred with inanimate antecedents both in Old English and early Middle English; less frequently, it is also found resuming animate ones in O4 and M2 (the latter somehow reflecting the general distribution of animate vs inanimate antecedents). By contrast, no such clear preference regarding the animacy of the antecedent has been observed with invariable *þe*. It is used as frequently with animate as with inanimate antecedents, contra McIntosh (1947-48: 74), Kivimaa (1966: 135) and Jack (1975: 101; 1988: 58), who all observed a preference for *þe* to be used with animate antecedents. The distribution in M1, the most representative period, confirms that, when relativizer *þat* becomes the favoured relativizer, it tends to be used with animate antecedents, occupying the space dominated by *þe* (Fischer 1992: 295), and also behaving as the only invariable relativizer used with both animate and inanimate antecedents.

Figures 5 and 6 illustrate the findings discussed. Invariable relativizer *þe* shows a clear preference for combining with animate antecedents, while *þat* favours inanimate ones. This could be related to the fact that the pronominal relativizer *þat* in Old English was used with nouns marked for the neuter gender, and these are very frequently inanimate.[12]

Figures from Table 7 above confirm that the choice of *þe* and *þat* is sensitive to the animacy of the antecedent, both in Old English and in early Middle English (χ^2, p <.0001): *þe* is more frequently used with animate antecedents than with inanimate ones and *þat* is more frequently used with inanimate antecedents than with animate nouns. However, when *þat* becomes the only invariable relativizer it is used very frequently both with animate and inanimate antecedents, adopting the roles covered earlier by *þe*.

In sum, interesting tendencies emerge from the analysis of the NP antecedent. In Old English, the influence of the antecedent on the relativizer's choice has to do with the form. It is concluded that the use of the invariable relativizer *þe* is favoured when the antecedent is pronominal or singular. In early Middle English, animacy seems to be

---

[12] As opposed to the majority of male and female referent nouns which belong to the masculine and feminine gender groups.

the most relevant factor. On these grounds we can conclude that *þe* is mostly found with animate and inanimate plural antecedents and *þat* with inanimate antecedents and some very specific animate nouns.

3.4. Position of the relative clause

The last intralinguistic variable analyzed to explain the distribution of invariable relativizers is the position of the relative clause with respect to the antecedent. Unlike Present-day English, this variable is of relevance in earlier English because there was a strong tendency in the language to avoid non-clause-final embedded structures (Carkeet 1976). Therefore, relative clauses were usually separated from the antecedent they depend on.

Relative clauses have been coded as 'intraposed' (if the antecedent and the relativizer were adjoined), as illustrated by example (11), or 'non-intraposed' (if the antecedent and the relativizer were separate), represented by so-called 'extraposed' (example 5) and 'left-dislocated' (example 4) relative clauses. An important aspect that needs to be clarified in relation to this variable is that it is only relevant for a reduced number of examples, that is, only those instances in which the separation of the antecedent is possible. Instances such as (2) and (7), in which the entire NP antecedent is clause-final, were excluded from the analysis.

(11) Forþon þære burge nama [rc þe is nemned Gerusalem] is gereht sibbe
Because the city name rel is called Gerusalem means of peace
gesyhþ, forþon þe halige sawla þær restæþ.
sight because holy soul there rest
'For the name of the city which is called Jerusalem signifies 'sight of peace', because the holy souls rest there.'
[Q O2/3 IR HOM BLICK 6: 25]

The results are included in Table 8:

Table 8. Invariable relativizers and position of the antecedent (normalized frequencies per 1000 words)

		[+ intraposed]	[- intraposed]	TOTAL
O3 (950-1050)	*þe*	127 (3.5)	53 (1.4)	180
	Þat	-	-	-
O4 (1050-1150)	*þe*	64 (1.3)	115 (2.4)	179
	þat	1 (0.03)	3 (0.06)	4
M1 (1150-1250)	*þe*	131 (1.7)	75 (1.0)	206
	þat	24 (0.3)	5 (0.06)	29
M2 (1250-1350)	*þe*	-	1 (0.2)	1
	þat	9 (2.0)	15 (3.3)	24
M3 (1350-1420)	*þat*	178 (3.1)	25 (0.4)	203
TOTAL		534 (2.4)	292 (1.3)	826

The data from Table 8 show that, with the exception of O4 (1050-1150), both *þe* and *þat* show a strong preference for adjacency, which might lend support to the hypothesis that *þat* enters the language to occupy the slots left empty by *þe*.[13] The data of M3 (1350-1420) are irrelevant not only because it is the period when *þat* consolidates as the only invariable relativizer, but also because by this time the language had already fixed the word-order and intraposition was the norm in the case of adnominal relative clauses.

3.5. Type of text
In addition to the intralinguistic variables, two extralinguistic variables were analysed, namely the type of text and the dialect. Although four types of texts were initially included (medicine handbooks, history, religious treatises and homilies), in the end only religious treatises,

[13] No significance tests were applied here due to the low number of examples in some of the cells.

homiletic texts[14] and history were analysed, since these text types were available for all the different periods covered by the corpus. The relevance of this variable lies in the fact that the choice of relativizer can be stylistically determined, as demonstrated by Romaine (1982) in her study of relative markers in Middle Scots. Similarly, in the present study the label 'style' will be used to compare the distribution of the invariable relativizer used to introduce the relative clause. Taking into account that *þe* is earlier than *þat*, it is then expected that, when coexisting, a higher presence of *þe* would render a more conservative text type and, on the contrary, a higher presence of *þat* a more innovative text type. The question under investigation here is whether the text types show a different distribution of invariant relativizers. Table 9 includes the distribution of *þe* and *þat* in religious treatises, history and homiletic texts in the different periods covered in the present study.

Table 9. Invariable relativizers and text type (normalized frequencies per 1000 words)

		Religious Treatises	History	Homiletic Texts	TOTAL
O3 (950-1050)	*þe*	178 (19.2)	17 (8.6)	292 (10.3)	487
	þat	8 (0.9)	-	5 (0.2)	13
O4 (1050-1150)	*þe*	208 (10.4)	151 (7.7)	75 (12.8)	434
	þat	4 (0.01)	12 (0.6)	4 (0.7)	20
M1 (1150-1250)	*þe*	494 (14)	11 (4.2)	280 (9.2)	785
	þat	168 (4.8)	11 (4.2)	149 (4.9)	328
M3[15] (1350-1420)	*þat*	207 (10.4)	187 (13.2)	280 (15.8)	674
TOTAL		1267	389	1085	2741[16]

---

[14] Under the label 'homiletic texts', I included texts classified as 'homily', 'rule' and 'sermon' in the corresponding COCOA header.

[15] M2 has been excluded because there was no variation in terms of text type.

The comparison of different text types in Table 9 shows a significant correlation between the choice of invariable relativizer and text type, especially in early Middle English (χ^2, p <.005). The general trend shows a decrease of relative clauses introduced by *þe* from Old English to Middle English in all text types. In Old English, *þat* is clearly disfavoured. Diachronically, however, history shows a different distribution from religious treatises and homiletic texts, in that it evinces an important earlier increase of *þat*. In homiletic texts, the consolidation of *þat* takes place later, indicating that these texts are more conservative texts in nature than in the case of history texts.

3.6. Dialect

The final variable included in the analysis is the dialect. Since Old English texts are mostly in the West-Saxon dialect, texts are very homogeneous in terms of dialect, making an analysis of variation pointless. In Middle English, however, geographic diversity is the norm and thus it is from this period that dialectal variation was analyzed (1150-1250), and indeed the significance of this variable will be seen clearly in Middle English. Although five main dialects are distinguished in Middle English, the most revealing regional indicators group them into two macro-dialects: (i) Northern, which comprises the inheritors of Old English Northumbrian and Mercian, namely, Middle English Northern, East-Midland, and West-Midland; and (ii) Southern, which comprises the descendants of Old English West-Saxon and Kentish, namely Middle English Southwestern and Southeastern (or Kentish). Table 10 shows the distribution of invariant relativizers in early Middle English:

---

[16] The total number of tokens differs from that of other tables because only those text types that are represented in all periods are included.

Table 10. Invariable relativizers and dialect in early Middle English (1150-1250) (normalized frequencies per 1000 words)

	Þe	Þat	TOTAL
West-Midlands	236 (6.7)	252 (7.1)	488
East-Midlands	354 (10)	50 (1.4)	404
South	226 (6.4)	31 (0.9)	257
Kentish	30 (0.8)	5 (0.1)	35
TOTAL	846	338	1184

The distribution of invariable relativizers in early Middle English is highly significant (χ^2, p <.0001). Invariable *þe*, the favourite relativizer in late Old English and in early Middle English, turns out to be preferred in the South and Kent, the areas which were less affected by the Scandinavian invasions, and therefore more conservative linguistically speaking (Milroy 1992: 181). More surprisingly, it is also favoured in the East-Midlands, showing an even higher frequency than in the other two areas. This contradicts Kivimaa's claim that this relativizer levelled out first in this dialectal area (and the North). Taking into account that the southern dialect is the direct inheritor of Old English West-Saxon, the overuse of *þe* looks like a conscious strategy by the scribes in order to preserve this early form and resist the loss of a tradition which had already been lost in other areas. Much more difficult is to account for the high frequency of this relativizer in the East-Midlands, a linguistically advanced area which, accordingly, should integrate the innovations and abandon the old traditions earlier. The West-Midlands warrants special notice, for it is here that invariable *þat* has become the preferred relativizer, being even more frequent than invariable *þe*, and thus becoming an innovation. As mentioned before, this invariable relativizer was already present in late Old English, but it was very rarely used. It is important to remember that invariable *þat* scores higher in all dialects from early Middle English than in Old English, but the West-Midlands dialect warrants special notice, for it is here that invariable *þat* has become the preferred relativizer, more

frequent even than invariable *þe*, and thus becoming an innovative feature.

4. Summary and conclusions

Invariable relativizer *þe* was the predominant relativizer in late Old and early Middle English, but was replaced by *þat* in Middle English. Early Middle English marks the turning point in the status of *þat*, as its use expanded into a growing number of contexts, irrespective of its agreement with the antecedent. Emerging in late Old English, its frequency then rose slightly in early Middle English, before increasing still further until, by Middle English, it had become the only remaining invariable relativizer (13th century). In the 13th century, there is a (weak) tendency for *þat* to be used more frequently as subject and object, introducing intraposed restrictive relative clauses and resuming inanimate antecedents, mostly realized by singular count nouns in the West-Midlands and in history writing, that is, in most of the environments where *þe* is less prevalent. However, the notion of a complementary distribution is unsustainable; instead, *þat*'s emergence is one of almost immediate substitution (from c. 1150, where *þe* was favoured, to c. 1200, where *þat* was almost the only choice). By the end, it had become possible for *þat* to be used with any kind of antecedent and in any kind of context. In response to the question posed at the beginning of this paper, then, it can be concluded that *that* spreads by occupying the gaps occupied by *þe*, and as a consequence finally ousts it from the system of the invariable relativizers.

Why? The weakening and disappearance of inflectional markers which took place in the English language from late Old English onwards and the progressive levelling of declensions which was completed by Middle English may well account for the sudden nature of the substitution. The overuse of *þe* probably favoured the introduction of, or rather the preference for, an already existing, if yet infrequent, element in the paradigm. Though initially occupying the environments less favoured by *þe, þat* rapidly became the only choice in all environments. The replacement of *þe*, likewise, coincides to some extent with the period during which *þat* was gaining ground as the general subordinator. The critical period of analysis in this regard is

probably M2, but with so little material available from early Middle English, such hypotheses are no more than tentative.

References

Aarts, Flor. 1993. "*Who, whom, that*, and 0 in two corpora of spoken English." *English Today* 9: 19-21.

Alsagoff, Lubna and Ho Ckee Lick. 1998. "The relative clause in colloquial Singapore English." *World Englishes* 17: 127-138.

Ball, Catherine N. 1996. "A diachronic study of relative markers in spoken and written English." *Language Variation and Change* 8: 227-258.

Carkeet, David. 1976. "Old English correlatives. An exercise in internal reconstruction." *Glossa* 10(1): 44-63.

Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. Oxford: Basil Blackwell.

Comrie, Bernard. 1998. "Rethinking the typology of relative clauses." *Language Design: Journal of Theoretical and Experimental Linguistics* 1: 59-85.

Dekeyser, Xavier. 1984. "Relativizers in Early Modern English: A dynamic quantitative study." In Fisiak, Jaček (ed.). *Historical Syntax*. Berlin: Mouton de Gruyter: 61-88.

Dekeyser, Xavier. 1986. "Relative Markers in the *Peterborough Chronicle*: 1070-1154 or Linguistic Change Exemplified." *Folia Linguistica Historica* 7: 93-105.

Dekeyser, Xavier. 1988. "Preposition stranding and relative complementizer deletion. Implicational tendencies in English and the other Germanic languages." *Leuvense Bijdragen* 77: 161-181.

Fischer, Olga. 1992. "Syntax." In Blake, Norman (ed.). *The Cambridge History of the English Language, vol II*. Cambridge: Cambridge University Press: 207-408.

Fischer, Olga, Ans van Kemenade, Willem Koopman and Wim van der Wurff. 2000. *The Syntax of Early English*. Cambridge: Cambridge University Press.

Geisler, Christer. 2002. "Relativization in Ulster English." In Poussa, Patricia (ed.). *Relativization on the North Sea Littoral*. Munich: Lincom Europa: 135-146.

Gisborne, Nikolas. 2000. "Relative clauses in Hong Kong English." *World Englishes* 19: 357-371.

Givón, Talmy. 1993. *English Grammar. A Function-based Introduction*. Amsterdam and Philadelphia: John Benjamins.

Gut, Ulrike. 2009. "Stylistic variation in the New Englishes—a correlate of their 'evolutionary stage'?" Paper presented at *The Third International Conference on the Linguistics of Contemporary English* (ICLCE3), London 14-17 July 2009.

Guy, Gregory R. and Robert Bayley. 1995. "On the Choice of Relative Pronouns in English." *American Speech: A Quarterly of Linguistic Usage* 70: 148-162.

Herrmann, Tanja. 2005. "Relative clauses in English dialects of the British Isles." In Kortmann, Bernd, Tanja Herrmann, Lukas Pietsch and Susanne Wagner. *A Comparative Grammar of British English Dialects*. Berlin: Mouton de Gruyter: 21-123.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Ihalainen, Ossi. 1980. "Relative clauses in the dialect of Somerset." *Neuphilologische Mitteilungen* 81: 187-196.

Jack, George. 1975. "Relative pronouns in language AB." *English Studies* 56: 100-107.

Jack, George. 1988. "Relative pronouns in Layamon's Brut." *Leeds Studies in English* 19: 31-66.

Jacobsson, Bengt. 1994. "Nonrestrictive relative that-clauses revisited." *Studia Neophilologica* 66: 181-195.

Keenan, Edward. 1985. "Relative clauses." In Shopen, Timothy (ed.). *Language Typology and Syntactic Description*. Cambridge: Cambridge University Press: 141-170.

Keenan, Edward and Bernard Comrie. 1977. "NP accessibility and universal grammar." *Linguistic Inquiry* 8: 63-99.

Kivimaa, Kirsti. 1966. "*Þe* and *þat* as clause connectives in early Middle English with special consideration of the emergence of the pleonastic *þat*." *Commentationes Humanarum Litterarum* 39: 1-271.

McIntosh, Angus. 1947-1948. "The relative pronouns *þe* and *þat* in early Middle English." *English and Germanic Studies* 1: 73-87.

Meier, Heinrich H. 1967. "The lag of relative *who* in the nominative." *Neophilologus* 51: 277-288.

Milroy, James. 1992. "Middle English dialectology." In Blake, Norman (ed.) *The Cambridge History of the English Language*, *vol II, 1066-1476*. Cambridge: Cambridge University Press: 156-206.

Mitchell, Bruce. 1985. *Old English Syntax*. Oxford: Clarendon Press.

Mustanoja, Tauno. F. 1966. *A Middle English Syntax*. (*Memóires de la Société Néophilologique de Helsinki* 23). Helsinki: Société Neophilologique.

Nevalainen, Terttu and Helena Raumolin-Brunberg. 2002. "The rise of relative *who* in Early Modern English." In Poussa, Patricia (ed.). *Relativization on the North Sea Littoral*. Munich: Lincom Europa: 109-121.

Peitsara, Kirsti. 2002. "Relativizers in the Suffolk dialect." In Poussa, Patricia (ed.). *Relativization on the North Sea Littoral*. Munich: Lincom Europa: 167-180

Quirk, Randolph. 1957. "Relative clauses in educated spoken English." *English Studies* 38: 97-109.

Rissanen, Matti, Merja Kytö, Leena Kahlas-Tarkka, Matti Kilpiö, Saara Nevalinna, Irma Taavitsainen, Terttu Nevalainen and Helena Raumolin-Brungerg (comp.). 1991. *The Helsinki Corpus of English Texts*. Helsinki: Department of English, University of Helsinki.

Rissanen, Matti. 1991. "On the history of *that*/zero as object clause in links in English." In Aijmer, Karin and Bengt Altenberg (eds.). *English Corpus Linguistics*. New York: Longman: 272-89.

Rissanen, Matti. 1997. "Optional THAT with subordinators in Middle English." In Hickey, Raymond and Stanisław Puppel (eds.). *Language History and Language Modelling. A Festschrift for Jaček Fisiak in this 60th Birthday*. Berlin: Mouton de Gruyter: 373-383.

Romaine, Suzanne. 1982. *Socio-historical Linguistics. Its Status and Methodology*. Cambridge: Cambridge University Press.

Rydén, Mats. 1963. "The emergence of *who* as relativizer." *Studia Linguistica* 37: 126-134.

Seppänen, Aimo and Göran Kjellmer. 1995. "*The dog that's leg was run over*: On the genitive of the relative pronoun." *English Studies* 76: 389-400.

Sigley, Robert. 1997. "The influence of formality and channel on relative pronoun choice in New Zealand English." *English Language and Linguistics* 1: 207-232.

Stein, Dieter. 1998. "Relative Sentences in Late Middle English: The Paston and Cely Letters." In Dieter Stein and Rosanna Sornicola (eds.). *The Virtues of Language. History in Language, Linguistics and Texts. Papers in Memory of Thomas Frank.* Amsterdam: John Benjamins: 67-77.

Suárez, Cristina. 2004. *Relativization in Early English (with Special Reference to the Distribution of Relativizers and the Position of Relative Clauses).* Universidade de Santiago de Compostela: Servizo de Publicacións [CD-ROM].

Suárez, Cristina. 2008a. "Strategies in competition: pronominal relativizers in the history of English." *English Studies* 89: 339-350.

Suárez, Cristina. 2008b. "Syntactic dialectal variation in Middle English: Relativizers and relative clauses." In Gotti, Maurizio, Marina Dossena and Richard Dury (eds.). *English Historical Linguistics 2006. Vol. I: Syntax and Morphology.* Amsterdam and Philadelphia: John Benjamins: 141-156.

Tagliamonte, Sali, Jennifer Smith and Helen Lawrence. 2005. "No taming the vernacular! Insights from the relatives in northern Britain." *Language Variation and Change* 17: 75-112.

Taylor, Ann 2008. "Contacts effects of translation: distinguishing two kinds of influence in Old English." *Language Variation and Change* 20: 341-365.

Tottie, Gunnel and Michel Rey. 1997. "Relativization strategies in Earlier African American Vernacular English." *Language Variation and Change* 9: 219-247.

Tottie, Gunnel and Harvie Dawn. 2000. "It's all relative: Relativization strategies in early African American English." In Poplack, Shana (ed.). *The English history of African American English.* Oxford: Blackwell: 198-230.

Traugott, Elizabeth C. 1992. "Syntax." In Hogg, Richard (ed.). *The Cambridge History of the English Language, vol I.* Cambridge: Cambridge University Press: 168-229.

Van den Eynden, Nadine. 1993. *Syntactic Variation and Unconscious Linguistic Change: A study of Adjectival Relative Clauses in the Dialect of Dorset.* Frankfurt am Main: Peter Lang.

Yamashita, Junko. 1994. "An analysis of relative clauses in the Lancaster/IBM spoken English corpus." *English Studies* 75: 73-84.