English relative clauses in a cross-Germanic perspective

Julia Bacskai-Atkari
University of Konstanz

Abstract

The article examines the distribution of relativising strategies in English in a cross-Germanic perspective, arguing that English is special among Germanic languages both regarding the number of available options and their distribution. The differences from other Germanic languages (both West Germanic and Scandinavian) are primarily due to the historical changes affecting the case and gender system in English more generally. The loss of case and gender on the original singular neutral relative pronoun facilitated its reanalysis as a complementiser. The effect of the case system can also be observed in properties that are not evidently related to case. Specifically, the choice between the pronoun strategy and the complementiser strategy is known to show differences according to the Noun Phrase Accessibility Hierarchy. While English shows a subject vs. oblique distinction in this respect, matching its nominative/oblique case system, German dialects show a subject/direct object vs. oblique distinction, matching the nominative/accusative/oblique case setting in the language. The particular setting in English is thus not dependent on e.g. a single parameter but on various factors that are otherwise present in other Germanic languages as well, and it is ultimately the complex interplay of these factors that results in the particular setup.

1. Introduction

There are two major types of relative clauses in English, illustrated in (1) below:

(1) a. This is the linguist who has an interesting theory.
   b. This is the linguist that has an interesting theory.
   c. This is the book which presents an interesting theory.
   d. This is the book that presents an interesting theory.

The two strategies are the relative pronoun strategy, shown in (1a) and (1c), and the relative complementiser strategy, shown in (1b) and (1d). The form that is invariant, while relative pronouns (who/whom/which/whose) show distinctions with respect to syntactic function and also regarding whether they refer to a human or a non-human antecedent. The two relative clauses above demonstrate yet another difference, which is

---

1 would like to thank the three anonymous reviewers for their inspiring and constructive remarks. I also owe many thanks to the audience of CSGW 34 for their questions and suggestions, in particular to George Walkden, Ellen Brandner, Elly van Gelderen, Theresa Biberauer and Peter Svenonius.

2 Apart from these two, there is in fact a third strategy, called zero relatives, where neither the relative pronoun nor the relative complementiser is overt:

(i) This is the linguist I saw yesterday.

This option is limited; while Standard English allows zero relatives in object relatives like (i), zero subject relatives are prohibited, so the zero relative counterparts of (1) are ungrammatical:

(ii) *This is the linguist has an interesting theory.
   (iii) *This is the book presents an interesting theory.

Varieties of English differ regarding constraints here; I will briefly return to this issue in section 5. Zero relatives are attested from Old English onwards and they appear to be a genuinely native strategy (Traugott 1992:228), even though they are comparatively rare (see Ringe and Taylor 2014:467); English also differs from other Germanic languages in this respect. As the present paper is concerned with the morphosyntactic properties of overt markers, zero relatives will not be discussed in detail.

3 The categorisation, as will be discussed in section 2, is standard in generative theory. Traugott (1992:224–228) considers which an invariant form and groups it together with that; however, which is clearly a pronoun in Present-Day English as it shows sensitivity to the head noun and it can be combined with prepositions (e.g. from which).

---
that there are two possible sources of relative markers: demonstrative elements (the complementiser *that* stems from the surface-identical demonstrative) and *wh*-elements (the *wh*-element *who* stems from the surface-identical interrogative pronoun). 4

Considering the standard varieties of Germanic languages, English appears to be unique as far as the number and distribution of options is concerned: both strategies are represented. In other standard Germanic languages, there is a clear preference for either strategy: German and Dutch use relative pronouns, 5 while Scandinavian languages use relative complementisers. Non-standard varieties show more variation in all languages, including the option of doubling (that is, the co-occurrence of an overt relative pronoun and an overt relative complementiser).

The question arises what factors contributed to the emergence of the English pattern. In this article, I argue that the development is primarily related to the nominal inflection system (primarily case and gender) and to the feature properties of the source elements behind relative markers.

2. Operators versus complementisers

In Standard English, the relative pronoun and the complementiser are apparently in complementary distribution:

(2) *This is the linguist who that has an interesting theory.

The proposal of Chomsky and Lasnik (1977) assumes a COMP position: in varieties like Standard English, the co-presence of the two elements in (2) violates the Doubly Filled COMP Filter. This assumption does not imply that ungrammaticality stems from complementary distribution per se: in fact, the authors claim that the complementiser is there in the underlying structure but if a visible *wh*-element adjoins to it, it is not spelt out overtly.

However, the violation of the supposed filter is possible in non-standard varieties:

(3) %This is the town in which that I live.

For cases like (3), Chomsky and Lasnik (1977) assume that the filter in the relevant (non-standard and/or historical) dialects does not apply.

In more recent approaches (van Gelderen 2009:152, Brandner and Bräuning 2013, Bacskai-Atkari 2018a), doubling patterns are generally taken to involve an overt specifier and an overt head in the CP, as shown in (4) below. Non-doubling patterns realise only one of these positions overtly.

(4)

```
CP
PP
\[\text{in which} \]
\[\text{C} \]
\[\text{C}^\prime \]
\[\text{TP} \]
\[\text{that \ldots} \]
```

4Given this, the term “relative pronoun” in the present paper refers to both *wh*-based relative pronouns and demonstrative-based relative pronouns, as I will be concerned primarily with the syntactic status of these elements. I adopt a matching analysis for relative clauses; see the recent study of Salzmann (2017:55–179) on arguments in favour of this, as well as Lees (1960; 1961), Chomsky (1965), Sauerland (1998; 2003) for similar views.

5Relative pronouns in German are usually demonstrative-based pronouns (der/die/das); this is the most common pattern and it can be observed already in Old High German and in Old Saxon (Fleischer 2004a:232). However, it is also possible to use *wh*-pronouns (welcher/welche/welches) for the same functions: this option is less common and more formal (and very rare in dialects, as shown by Fleischer 2004a:220). In addition, adverbal relative clauses also show *wh*-pronouns, such as the locative adverbial *wo* ‘where’ and prepositional adverbiais (e.g. *woanich* ‘after what’). With certain matrix elements such as *etwas* ‘something’, not only the regular neuter relative pronoun *das* but also its *wh*-counterpart *was* can be used (Brandt and Füll 2014).
The representation in (4) highlights a crucial difference between canonical operators and canonical complementsisers, which will be a central issue in this paper. Canonical complementsisers are base-generated in C: they are functional heads merged directly with the TP. I will use the term “complementsiser” only for these kinds of elements.  
Prototypical operators, on the other hand, are base-generated in a lower position and undergo operator movement to the left periphery.

Doubling patterns like (3) are similar to doubling in interrogatives (Chomsky and Lasnik 1977 treat them similarly as well):

(5) %She asked me in which city that I lived.

However, doubling in interrogatives is apparently more frequent than in relative clauses.  
Essentially the same observation holds for German: while many dialects regularly use doubling in interrogatives, dialects prefer the complementsiser strategy and may additionally use a pronoun (cf. Bayer 1984, Salzmann 2006; 2009). This asymmetry between interrogatives and relative clauses is unexpected under the COMP-analysis of Chomsky and Lasnik (1977) that builds on a surface filter, as the surface string is parallel in the two constructions. The examples in (6) below illustrate the standard versus the dialectal pattern in German, respectively:

(6) a. . . . der Mann der seine Schuhe verloren hat
   the man that.M NOM his shoes lost has
   ‘the man who has lost his shoes’ (Brandner and Bräuning 2013:132)
   
   b. . . . dea Mo (dea) wo seine Schu verlora hot
   the man that.M NOM PRT his shoes lost has
   ‘the man who has lost his shoes’ (Alemannic/Bavarian; Brandner and Bräuning 2013:132)

As indicated, the standard pattern necessarily uses a relative pronoun, while the same pronoun is optional in the dialectal pattern, which regularly uses the complementsiser wo.  
Note that the relative pronoun is marked for case, number and gender in both Standard German and in South German varieties.

This is different from embedded interrogatives, where doubling (the additional insertion of the complementsiser dass ‘that’ following the wh-element) may be obligatory:

(7) I ask-REFL for what that-one a second TV needs
   I wonder what one needs a second TV for.’ (Bavarian; Bayer and Brandner 2008:88)

As reported by Bayer and Brandner (2008), the complementsiser dass is essentially obligatory for most Alemannic and Bavarian speakers if the wh-phrase is visibly phrase-sized (i.e. it contains other, lexical

---

6The definition thus excludes elements inserted directly into the specifier (such as polar operators, see Bianchi and Cruschina 2016) and elements moving from a clause-internal position to C (such as finite verbs in V2 clauses in most Germanic languages or, as will be discussed later, clause-typing operators moving to C).
7As van Gelderen (2013:59) notes, such examples are attested “in some varieties of English”, yet there is no evidence for this pattern to be the preferred strategy in any variety. The difference is difficult to measure precisely as the relevant constructions are substandard and stigmatised, so that both grammaticality judgements and corpora are problematic as sources: the acceptance may be influenced by prescriptive considerations, and doubling hardly occurs in written language. Some grammaticality judgements from speakers of relevant varieties reveal that speakers accepting doubling in relative clauses also accept doubling in embedded interrogatives, but the implication does not hold vice versa. This issue, however, should be investigated more systematically. Regarding English, a good collection of relevant data is the one by Beatrice Santorini (https://www.ling.upenn.edu/beatrice/examples/doublyFilledCompExamples.html): interestingly, all of the examples given here (as of 5 May 2020) are from embedded interrogatives and free relatives, but none from headed relative clauses.
8Bavarian dialects may also use the complementsiser wie, see Brandner and Bräuning (2013:132), citing Eroms (2005). Both elements go back to an equative marker.
9For this reason, the conclusions concerning nominal inflection and relative pronouns in German presented in this paper refer both to Standard German and to non-standard dialects, unless otherwise indicated.
As far as Dutch is concerned, very few doubling patterns are reported by Boef (2013:141) and they are restricted to southern areas: according to the SAND1 data, the doubling of a wh-based pronoun and the complementiser dat is attested in southern Dutch and the doubling of a d-pronoun with dat is attested in Waasland Dutch. It is questionable, however, whether dat in these cases can be considered a relative complementiser since dat-relatives are not reported from the same areas (these are reported from Vlaams-Brabant in SAND1). It is perfectly possible that dat in these cases is the regular finite subordinator.\footnote{The insertion of dass may be optional or even prohibited with head-sized wh-elements. Similar observations were made by Bayer (2013:24), citing Weiβ (2004) for Bavarian and Schönenberger (2006) for varieties of Swiss German, and by Vangsnes (2005) and Westergaard and Vangsnes (2005) for North Norwegian varieties. Bayer and Brandner (2008) and Bayer (2014; 2015) argue that this is because head-sized wh-elements may move to C instead of [Spec,C]. Apart from the observed absence of the complementiser dass in these cases, various phonological properties, such as ephenthesis and cliticisation, point to this conclusion. This indicates that the division line between complementisers and operators moving to the left periphery is not rigid: apart from canonical operators (moving to the specifier) and canonical complementisers (base-generated in C), there are other possibilities as well, including inserting operators directly into the specifier and moving operator elements to C (that is, merging these moved elements directly with TP). In fact, it is precisely this flexibility that allows the reanalysis of operators into complementisers proper. I will return to this issue in section 4.}

Apparently, there is a strong preference for lexicalising the C position in Germanic languages, which favours complementisers (as pointed out recently by Bacskai-Atkari 2018a;b, this preference can also be observed in main clauses, for instance in V2 patterns).\footnote{Under this view, lexicalising C is primarily a matter of syntactic features. In particular, the insertion of dass in constructions like (7) is in line with the lexicalisation of [fin] on C and it is not triggered by phonological factors. Bayer and Brandner (2008) and Bayer (2014; 2015) assume that doubly filled COMP patterns are maintained in the relevant dialects because these dialects have clitics, “and clitics can converge in the phonological interface only if the syntax provides the appropriate landing sites” (Bayer 2014:40–41). This view is problematic for two reasons. On the one hand, if the insertion of dass were primarily a phonological matter, one would expect either (i) dass to be absent in cases where no clitic is present altogether, or (ii) at least a significant improvement for dass-less clauses (with complex wh-phrases) in the absence of clitics. However, even for dialects like Bavarian and Alemannic, clitic-less clauses with dass are reported (also by the authors cited above) and no significant improvement has yet been identified. Regarding (i), Bayer (2014:41) remarks that a phonological motivation for a dialect retaining doubly filled COMP patterns “must be seen as affecting the grammar as a whole and not individual constructions”. This does not address (ii), though. Further, this explanation still leads to a second problem, which is that doubly filled COMP patterns in embedded interrogatives occur across West Germanic, also in dialects that are not known to use (subject) clitics in the Bavarian way. Specifically, the spoken (and dialectal, historical) English data mentioned above do not indicate that cliticisation would play any role.}

This speaks against a grammaticalisation scenario in the English way. Given this, the conclusions concerning nominal inflection and relative pronouns in Dutch presented in this paper, similarly to German, refer both to Standard Dutch and to non-standard dialects, unless otherwise indicated.

\footnote{This view is problematic for two reasons. On the one hand, if the insertion of dass were primarily a phonological matter, one would expect either (i) dass to be absent in cases where no clitic is present altogether, or (ii) at least a significant improvement for dass-less clauses (with complex wh-phrases) in the absence of clitics. However, even for dialects like Bavarian and Alemannic, clitic-less clauses with dass are reported (also by the authors cited above) and no significant improvement has yet been identified. Regarding (i), Bayer (2014:41) remarks that a phonological motivation for a dialect retaining doubly filled COMP patterns “must be seen as affecting the grammar as a whole and not individual constructions”. This does not address (ii), though. Further, this explanation still leads to a second problem, which is that doubly filled COMP patterns in embedded interrogatives occur across West Germanic, also in dialects that are not known to use (subject) clitics in the Bavarian way. Specifically, the spoken (and dialectal, historical) English data mentioned above do not indicate that cliticisation would play any role.}

In the classical scenario, the wh-part of a constituent question corresponds to a focused element in the answer (see Krifka 2008:250, citing Paul 1880). Krifka (2008:248) suggests that a focus property indicates the presence of alternatives (this idea in turn goes back to von Stechow 1981 and to Rooth 1985, and it was adopted by later analyses, see Büring 2013). Interrogative operators do not necessarily point to any antecedent in the discourse. (Note that while there is a strong correlation between discourse-new and stress, it implies no one-to-one correspondence, as discussed by Büring 2013:874–876.) It follows that the wh-element in interrogatives cannot be left unpronounced. Adding this factor to the preference for an overt element in C, doubling patterns predictably arise. The question remains why Standard English does not allow this construction; in fact, all the three West-Germanic languages examined here impose a ban on doubling patterns in their standard varieties. It appears that this ban can be drawn back to prescriptive pressure.

\footnote{Whether constructions with zero operators are actually available depends on various factors; the absence of an overt relative pronoun leads either to zero relatives or to relative clauses introduced by an overt complementiser. Zero relatives constitute a restricted option (see section 1 and section 5). The complementiser strategy is clearly not an option in varieties that do not permit relative complementisers in the first place (e.g., Standard German and Standard Dutch). But restrictions arise even in varieties that have both the pronoun and the complementiser strategy. As will be discussed in section 5, the syntactic function of the gap makes an important difference here: the complementiser strategy (as a single option) tends to be more available in certain functions (such as the subject and the direct object) than in others (such as indirect objects or objects of prepositions). The case of the head noun may also be decisive. Bayer (1984:215–217, 221–222) reports for Bavarian that the presence or absence of the relative pronoun along the relative complementiser wo depends on the morphological case of the relative operator and the head noun: if the cases match, the definite
This leads to a predictable preference for complementisers in relative clauses not only in dialects of German but also in English. As pointed out by van Gelderen (2009:163), wh-pronouns are promoted by prescriptive rules (especially in the case of human referents) but English speakers in fact prefer that over a wh-pronoun (by at least a 4:1 ratio; see also Romaine 1982:129, citing Sweet 1900; Montgomery and Bailey 1991, van Gelderen 2004, Tagliamonte et al. 2005). Note that while several studies point towards this conclusion, the exact formulation may differ considerably. For instance, Romaine (1982:212) originally made a very strong claim in suggesting that while wh-pronouns became established in the written language, they barely had an effect on the spoken language. As shown by Ball (1996), such a strong claim is not tenable since the wh-strategy has affected spoken language as well, though with considerable dialectal and sociolinguistic differences. The prescriptive preference for wh-pronouns is especially strong in the case of human referents. The asymmetry between written and spoken language goes back to at least the 17th century: the preference for that over wh-pronouns is stronger in trials (Ball 1996) and in drama texts (Dekeyser 1984) than in prose, poetry and letters (see Johansson 2012:778). In addition, a change in the preference during the 17th century can be detected in various genres, as shown by Lezcano (1996). As mentioned above, this phenomenon in written language is largely due to prescriptive pressure. Austen (1985:21, 24) points out that this was already the case with 18th-century grammarians (including, for instance, Addison). Ball (1996:247) shows that this trend was present in the 17th century as well, with writers like Dryden preferring who over that, but the sharp decline of that took place in the 18th century.

A clear preference for the complementiser strategy can also be observed in Old English (Ringe and Taylor 2014:467) and in present-day dialects (cf. Herrmann 2005, Kortmann and Wagner 2007, Beal 2008). Regarding dialects, the preference holds not only for the canonical demonstrative-based complementiser that, but also for the wh-based complementisers what and where, and traditional patterns with as and at, with considerable regional differences.

Such a preference is pointed out by the corpus study of Bacskaï-Atkari (2020) regarding Early Modern English. This study was carried out on two versions of the King James Bible, the original version (1611/1769) and the modernised version (1989), and it examined the distribution of relative markers across the two texts, looking at parallel loci.\(^{15}\) The sample results from the original version for Genesis are given in Table 1:

<table>
<thead>
<tr>
<th>TOTAL</th>
<th>who</th>
<th>whom</th>
<th>which</th>
<th>that</th>
<th>as</th>
</tr>
</thead>
<tbody>
<tr>
<td>374</td>
<td>14</td>
<td>25</td>
<td>123</td>
<td>210</td>
<td>2</td>
</tr>
<tr>
<td>(3.74%)</td>
<td>(6.68%)</td>
<td>(32.89%)</td>
<td>(56.15%)</td>
<td>(0.53%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The distribution of relative markers in the KJB (Genesis)

The parallel sample results from the new version (Genesis) are given in Table 2.

---

\(^{15}\)One may wonder to what extent the two versions reflect the distribution of relative markers in the respective periods generally. Regarding the modern version, Bacskaï-Atkari (2020:96) mentions that it is very conservative (for instance, in terms of consistently using whom for all non-nominative forms) and the text adheres to the original in many respects. What matters is not so much the distribution of the individual markers in the new version in itself but rather the difference between the original and the modernised version, which reflects conscious deviations from the previous pattern in line with of prescriptive rules and language change. As pointed out by the cited study, the distribution of relative markers is comparable to modern dialect data. The original version may in turn be more archaic in general than other texts from the period (as, for instance, in using -th instead of -s for 3Sg on verbs, see van Gelderen 2014:173); what matters for us is rather the fact that it can be dated back to a period when the prescriptive pressure disfavouring that was not yet active. The preference for that over the wh-strategy in the 17th century can be detected in other texts as well (to varying degrees), as pointed out by Ball (1996), Dekeyser (1984) and Johansson (2012), see the discussion above. In other words, the asymmetries discussed here cannot be regarded as mere archaisms.
ENGLISH RELATIVE CLAUSES IN A CROSS-GERMANIC PERSPECTIVE

<table>
<thead>
<tr>
<th>TOTAL</th>
<th>who</th>
<th>whom</th>
<th>which</th>
<th>that</th>
</tr>
</thead>
<tbody>
<tr>
<td>374</td>
<td>106</td>
<td>41</td>
<td>118</td>
<td>109</td>
</tr>
<tr>
<td>(28.34%)</td>
<td>(10.96%)</td>
<td>(31.55%)</td>
<td>(29.14%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: The distribution of relative markers in the new version (Genesis)

The data indicate a clear overall preference for the complementiser strategy (cf. Grainger 1907:28), where more than half of the examples use a complementiser (mostly that, in two cases as), which is not reflected in the modernised version, where less than a third of the examples features a complementiser. I will return to the distribution in more detail later; at this point, what matters for us is that the high proportion of wh-elements in Standard English is not simply the natural result of their general spread following their appearance in Middle English, but this high proportion is also largely due to requirements of norm-oriented language use. The detailed analysis (Bacskai-Atkari 2020:97–108) shows that several prescriptive rules interact here. Apart from a general preference for wh-relatives, who is especially promoted over that in relative clauses with human referents (the above-mentioned prescriptive rules, dating back to the 17th and 18th centuries, consider that impolite in such cases), and the original examples of which with human referents were naturally changed to who/whom as well, since which-relatives with human referents are no longer possible in Standard English (this is less strictly but largely adhered to in present-day British dialects as well, see Herrmann 2005).16

A similar preference towards complementisers can be observed across Germanic (cf. van Gelderen 2009:164). Many German dialects use wo or was (see Brandner and Brüning 2013 on Bodensee Alemannic; Salzmann 2017 on Zurich German; Fleischer 2004a; 2017 on Hessian; Weiß 2013 on Bavarian; see also Kaufmann 2018 on Mennonite Low German). Mainland Scandinavian languages use som and Icelandic uses sem. In Flemish, dat is possible as a relative complementiser, but it is mostly attested only in combination with a pronoun (Bennis and Haegeman 1984, Boef 2013).

This indicates that the English pattern is not unique across Germanic in favouring a complementiser strategy. However, English seems to be peculiar among West-Germanic languages in employing a demonstrative-based complementiser across varieties and also regarding the extent to which the pronoun strategy competes with the complementiser strategy. As will be shown in the next section, this pattern can be explained by examining the historical development that led to it.

3. The relative cycle

The historical development underlying the grammaticalisation of that as a relative complementiser also indicates that the distinction between relative pronouns and relative complementisers is not as rigid as it may seem. This section briefly summarises the reanalysis of that in English relative clauses, which is described by van Gelderen (2004:77–99) and van Gelderen (2009:161–168) as an instance of the so-called relative cycle.17

16 While the ban on which with human referents is categorical in the standard variety and apparently very strong in regional dialects of Britain (to the extent these dialects use wh-relatives at all), whom may be used with non-human referents even in the standard variety in the so-called ‘sanctioned borderline cases’, such as when a dog is considered to be a family member, see Herrmann (2005:41), citing Quirk et al. (1985); regional dialects do not show any difference in this respect.

17 The reanalysis of that from an original demonstrative element to a relative pronoun and subsequently a complementiser contains two reanalysis steps that are both attested cross-linguistically: see van Gelderen (2009). Relative pronouns very often go back to demonstrative elements (see also Roberts and Roussou 2003). It is worth highlighting that while the reanalysis process to be discussed in this section eventually led to the disappearance of the relative pronoun, the same does not hold for the demonstrative pronoun: English has both a demonstrative that (D) and a complementiser that (C). In other words, reanalysis involves the featural reinterpretation of a certain element in a given construction (here: relative clauses) but not in others (here: demonstrative uses such as that man there etc.). Similar lexical splits can be observed with complementisers in general. Apart from demonstratives and interrogative/relative operators, complementisers often stem from prepositions, whereby the original preposition can be retained, as in the case of English for (P and non-finite C) and before and after (P and finite C, see van Gelderen 2013:49). Complementisers can have various origins; this also applies to English (see the list of van Gelderen 2013:48–49).
In stage 1 (the initial stage), *that* was a relative pronoun inflected for gender, number and case. The form *þæt* was the one used with a singular neuter antecedent (van Gelderen 2004:85, citing Grossmann 1906:38). In this stage, other relative pronouns could occur as well, as demonstrated by the examples in (8): (8a)\(^{18}\) shows the pronoun *ba* and (8b)\(^{19}\) shows the pronoun *þæt* in a headless (free) relative clause. The original relative complementiser is the invariant *þe* (Traugott 1992:226); this relative marker is the most frequent one in Old English (Ringe and Taylor 2014:467) and regularly introduces relative clauses on its own in Old English.\(^{20}\)

(8) a. *ac gif we asmeagaþ þa eadmoldican dæda þa þe he worht, bonne ne þineþ us þæt nan wundor seems us that no wonder 

‘But if we consider the humble deeds that he wrought, then that will not appear marvellous to us.’ *(Blickling Homilies* p. 33)

b. *þurh þæt þe he on þam gebede gehyrð ‘through that. N.ACC REL he in this. DAT prayer hears ‘through what he hears in this prayer’ *(Aethelwold, The Benedictine Rule*, 884–885)

The structure is given in (9) below.

(9) \[ CP \]
\[ \{þæt/þa/…\} \]
\[ \{C \}
\[ \} \]
\[ \{C TP \}
\[ \} \]
\[ \{þe \} \]

In stage 2, *that* is reanalysed as a complementiser. Consider the following example:\(^{21}\)

(10) *and suggeð feole þinges; bi Arðure þan kinge. // þat næure nes i-wurðen and say many things about Arthur the king that never not-was happened ‘and say many things about Arthur the king that never happened.’ *(Layamon, Brut, Caligula version, 11473–11474)*

In this stage, *that* occurs on its own and it can be used with essentially any kind of lexical head. The structure is given in (11) below.

---

\(^{18}\)The example is taken from the following edition: *The Blickling Homilies of the Tenth Century*, ed. R. Morris, Early English Text Society, os. 58, 63, and 73. London: N. Trübner & Co., 1880. The page number refers to the one in this edition. The translation of the example is taken from this edition as well. Allen (1977:88) uses the same example and gives the following translation (taken over by Watanabe 2009:363–364): ‘But if we consider the humble deeds which he wrought, that will seem no wonder to us.’

\(^{19}\)The example is from the Helsinki Corpus. The translation is adopted from Jacob Riyeff (*The Old English Rule of Saint Benedict: With Related Old English Texts*. Liturgical Press, 2017).

\(^{20}\)As pointed out by Traugott (1992:221), citing Wiegand (1987), *þe* is also a demonstrative-based element: it appears that this element underwent the relative cycle as well. Traugott (1992:224, 226), citing Mitchell (1985), remarks that the doubling pattern, illustrated here in (8), is rare in poetry; she suggests that it is favoured when the head noun had no determiner or quantifier. See Traugott (1992:224–225) for further examples. Moreover, Traugott (1992:226) suggests that *þe* most frequently occurs when the gap in the relative clause is a subject or a direct object; I will return to this issue in section 5.

\(^{21}\)The example is taken from the Michigan Corpus of Middle English Prose and Verse. In a shortened form, it is used by van Gelderen (2004:85), citing Mätzner (1864:142).
Finally, in stage 3, *wh*-elements can appear in the specifier as well. This is exemplified below:

\[
\text{(12) the est Orisonte, which } \text{that is cleped communly the assendent}
\]

\[
\text{‘the East horizon, which is commonly known as the ascendent’}
\]

(Chaucer *Treatise on the Astrolabe*, folio 10)

This is again a doubling pattern. The structure is shown in (13) below.

\[
\text{(13) CP}
\]

\[
\text{\hspace{1cm} wh C’}
\]

\[
\text{\hspace{1cm} that}
\]

\[
\text{\hspace{1cm} C TP}
\]

In stage 3, then, new relative pronouns can move to the position left “empty” by the reanalysis of the original pronoun, leading to new doubling patterns. Naturally, the appearance of *wh*-elements in relative clauses also involves a change in the feature properties of these elements. Watanabe (2009) argues that these elements lost their original quantificational features and were thus no longer associated with complete propositions: consequently, they could appear in relative clauses.

Apart from typing the clause as relative (see e.g. Rizzi 1997 on relative as a clause type, distinct from interrogative; see also Traugott 1992:221), relative markers fulfil other functions as well. The complementiser encodes finiteness, while the relative pronoun carries phi-features in argument relative clauses. This implies that relative pronouns have to lose features that are not available on complementisers if they are to be reanalysed as complementisers; such features include case and gender, as well as number (but see below).

The reanalysis step itself (resulting in stage 2) is motivated by economy (van Gelderen 2004:77, 87, van Gelderen 2009:151–153, 161) and can be described as an example for the general preference of Merge over Move. While this proposal is undoubtedly attractive for several reasons, it raises the question how strongly such economy considerations operate, that is, whether reanalysis automatically happens once the relevant morphological conditions are met.

---

22 The introduction of *wh*-pronouns in relative clauses deviates from the original West Germanic pattern involving *d*-pronouns. This may well be due to Latin and French influence (see van Gelderen 2004:82–88, citing Mustanoja 1960 and Rydén 1983), though stylistic factors may also have played a role: see Romaine (1982:53–80). As noted by Gisborne and Truswell (2017:35–37), there does not seem to have been any internal pressure for the introduction and the spread of *wh*-relatives. Analogy with embedded interrogatives, however, may have been an additional factor, as suggested by Bacskai-Atkari (to appear).

23 The example is taken from the Michigan Corpus of Middle English Prose and Verse; van Gelderen 2004:87 uses the same example (with a different orthography) in her argumentation.
4. Relative elements and nominal inflection

As mentioned above, the loss of case and gender features on *that* is a central issue in terms of its reanalysis. The assumption underlying this statement is that these features (case and gender) are not compatible with a complementiser (in Germanic). It is worth mentioning at this point that inflection is not excluded per se: indeed, some dialects of German such as Bavarian allow complementisers like *ob* ‘if’ to be inflected for person and number (see, for instance, Fuß 2004 for discussion). However, these inflectional features are essentially ones that can be marked on verbs as well (as contextual inflection); the contiguity between complementisers and verbs is especially evident in Germanic, where verbal elements regularly undergo movement to C. However, the presence of genuine nominal inflection (including case and gender) is not attested on grammaticalised complementisers.24

In principle, the loss of case and gender is not exceptional in English, as it appears to be in line with the general loss of morphological inflection in Middle English, often subsumed under the general “reduction of forms and endings” (van Gelderen 2014:128). As noted by Allen (1997:63), similar reductions in inflections took place in other Germanic languages (Mainland Scandinavian languages and Dutch, the latter to be discussed in this section, cf. also Thomason and Kaufman 1988:315–321), but the changes in English appear to have been very rapid. She argues that this impression “is to a large extent due to changes in scribal practice” in Middle English and also to differences in the dialects: the earliest Middle English documents (that is, ones that are not merely copies of earlier compositions)” “come from an area in which it is likely that syncretism was already well advanced” in late Old English, while the late Old English documents that can be considered “come from more morphologically conservative areas” (Allen 1997:85).

This suggests that feature loss in the case of the relative marker *that* is not exceptional but it is in fact in line with the changes attested in the nominal inflectional system. Importantly, this does not mean that the changes in relative clauses were caused by the general loss of morphological case. In fact, it appears that the loss of morphological features on *that* is an early development when compared to changes in the general system: in addition to these morphological changes (which made *that* into an invariant form showing no sensitivity to the head noun), it was reanalysed as a head, permitting new operators in its specifier from the 12th century onwards (van Gelderen 2004:87, citing Dekeyser 1986:100–101 and Allen 1977:197–199). The roots of the changes in the morphological properties date back considerably earlier: the occurrence of *that* with masculine, feminine or plural antecedents is attested from at least the 9th century on (van Gelderen 2004:84, citing Kock 1897:30, Wülfing 1901:407–408, Dekeyser 1986:99; see also Traugott 1992:227 and Allen 2016:45), which is indicative of this element losing its gender features.25 Importantly, as Allen (2016:54–56) notes, while the system of demonstratives is still active in Early Middle English (with considerable dialectal differences), this diversity is not mirrored by relative clauses even in morphologically rich dialects: the old (demonstrative-based) relative pronouns are infrequent.26

24 As discussed in section 3, *that* not only lost its original case and gender features (nominal features) but it also acquired a feature related to clause typing, finiteness, which is essentially a verbal feature (the finite verb can also check off the [fin] feature by moving to the left of TP, see Fanselow 2004). In this sense, one could argue that the grammaticalisation of *that* from a relative pronoun into a relative complementiser involves the transition from nominal inflection to verbal inflection.

25 This contrasts with the view of Mitchell (1985:101–108), who proposed that the relevant examples can be explained otherwise and are not indicative of *fort* not being part of the regular demonstrative/relative declension system. While the use of *fort* with non-neutral antecedents is clearly a minority pattern in Old English, this fact simply reinforces the assumption that reanalysis was not completed. However, there is no reason to exclude the possibility of the relevant process starting already in Old English (as also suggested by Traugott 1992:227), given that such processes are expected to take place gradually: in other words, nothing speaks for explaining away examples merely because they do not fit neatly in the general system of the given period.

26 This asymmetry supports the idea that while relative pronouns can have their origins in demonstrative pronouns, the two are not one and the same: in particular, it is problematic to treat demonstrative-based relative pronouns as demonstrative pronouns “used as” relative pronouns (as is done by Allen 2016:47). Allen (2016:57) notes that the pronoun in relative clauses “does not have a simple anaphoric function, but refers to a hypothetical entity introduced by the relative clause”. This indicates an important difference in referential properties of the two kinds of elements. In addition, as mentioned already, relative pronouns also have clause-typing features. Traugott (1992:225–226) also distinguishes between the two functions and argues that there are naturally examples where the distinction is not clear, as punctuation is not indicative; it is precisely the similarity of the two kinds of constructions, though, that explains why the reanalysis of demonstrative pronouns into relative pronouns is attested in many languages of the world.
Regarding case, the fact that the original relative pronoun is ahead of the rest of the system is not unexpected given that the relative pronoun is removed from its case position. Such an asymmetry can be observed in main clause interrogatives with *whom* as the complement of a preposition in formal English (taken from Radford 1997:142; cf. also van Gelderen 2004:80):

(14) a. *Whom* were you talking to?
   b. To *whom* were you talking?

The sentence in (14a) demonstrates preposition stranding, which is otherwise a grammatical option (although it goes against certain prescriptive rules), and the sentence would be fine with the form *who* not showing morphological case. In (14b), the preposition is pied-piped and the form *whom* is adjacent to its case licenser.

As pointed out above, *that* in Old English was used with singular neutral referents; this changes during Middle English, as evidenced by examples like (10), repeated below as (15), where *that* occurs with a plural antecedent:27

(15) and suggeð feole pinges; bi Arðure þan kinge. // þat næure nes i-wurðen
   and say many things about Arthur the king that never not-was happened
   ‘and say many things about Arthur the king that never happened.’
   (Layamon, *Brut*, Caligula version, 11473–11474)

This raises at least two questions, namely how other (Germanic) languages behave in this respect and whether the loss of overt case marking has any other traceable effect. It is important to note at this point that the reanalysis from (relative) operators is not the only source of relative complementisers: many relative complementisers derive from equative complementisers. This is the case for Mainland Scandinavian *som*, Icelandic *sem*, dialectal and historical *as* in English (see Bacskai-Atkari 2020), as well as German *so* historically and present-day German *wo* (Brandner and Bräuning 2013). Nominal inflection (case and gender in particular) is evidently irrelevant for these items as case was never part of their feature array. Moreover, they are taken over from constructions in which they were complementisers already, which means that no reanalysis in the sense of the relative cycle is necessary (or possible).

In German, no reanalysis of the pronouns *der/die/das* is attested: these elements are evidently marked for case, number, gender, in the same way their surface-identical demonstrative counterparts are. The fact that there is no reanalysis in relative clauses is thus in line with there being overt case and gender marking in the language. This suggests that economy considerations are definitely constrained by morphology.

Dutch shows no reanalysis of demonstrative-based relative pronouns either. In Dutch, however, unlike in German, overt case marking has been lost, though it took place somewhat later than in English, where it was completed by the end of the Middle English period (end of 15th century). As noted by Nobels and Rutten (2014:34), for both Old and Middle Dutch, four cases (nominative, accusative, dative, genitive – just as in German) are distinguished (citing van der Horst 2008a:573–581 and van der Wal and van Bree 2008:132–135; see also Scott 2014:110). The system was weakened during the Middle Dutch period and

---

27In (15), the antecedent of the gap in the relative clause (the head noun) is plural; nevertheless, the verb in the relative clause shows singular agreement. According to van Gelderen (2004:85–86), citing further examples (from Chaucer, the Paston Letters, as well as Shakespeare), this agreement is triggered by the element *that*, which as a complementiser has lost its gender and case features but is still often specified as 3rd person singular. Further, van Gelderen (2004:86–87) cites some examples (in part from Chaucer and Shakespeare again) where this specification is not given. However, I consider it unlikely that the complementiser is specified as 3Sg, especially as it is unclear whether such examples constitute a statistically significant pattern. For one thing, van Gelderen (2004:86) mentions that similar examples showing an agreement mismatch occur even in present-day spoken English; in other words, examples like (15) cannot even be reliably identified as more archaic in this respect. What appears to be more likely is that (15) shows default agreement on the verb, whereby the zero relative operator simply does not match the head noun in number. Heck and Cuartero (2012) discuss cases of agreement mismatches in English (and German) relative clauses that point to a similar direction: third person is evidently the default value, while the picture is less clear regarding number inasmuch as the default singular is not always possible. However, Heck and Cuartero (2012) examine *wh*-relatives (in particular, relatives with *who*), so their data set does not cover the spoken language data of van Gelderen (2004) for *that*-relatives. The point is that mismatches in verb agreement arise without having to postulate person and number features on the complementiser.
by the 17th century it seems plausible that overt case marking was lost in the spoken language (Nobels and Rutten 2014:34, citing van der Horst 2008b:1074–1075; see also Scott 2014:110), while it continued to be part of the written language due to prescriptive pressure from the 16th century till the early 20th century (Scott 2014:110, Nobels and Rutten 2014:34). Case marking is well-attested on the relative pronouns die and dat in Old and Middle Dutch (Krogull et al. 2017:161).

By looking at the issue of morphological case only, the difference between English and Dutch is in fact minimal and does not in itself account for the asymmetry attested in relative clauses. However, there is another aspect to be considered, which is that gender marking has not been lost in Dutch and this is reflected in the difference between the relative pronouns die (common gender) and dat (neuter). The difference is illustrated below:

(16) a. het boek dat ik heb gelezen
   \textit{the.N book that.N I have read.PTCP}
   ‘the book I have read’

b. de man die daar staat
   \textit{the.M man that.M there stands}
   ‘the man who is standing there’

Old and Middle Dutch had a regular West Germanic three-way distinction between masculine, feminine and neuter, but the distinction between the masculine and the feminine started to weaken and eventually came to be lost in Modern (northern) Dutch (from the 16th century onwards, see also Scott 2014:110, citing van der Horst 2008a:803). At any rate, the gender specification on the relative pronoun is a feature incompatible with a complementiser. It seems that the ultimate difference between English and Dutch lies not so much in the marking of case but rather in the marking of gender: whereas morphological case and gender disappeared more or less simultaneously from English, the two are not tied together in the history of Dutch.

At this point, we can conclude that overt gender marking also constrains reanalysis: hence, economy considerations are constrained by various aspects of an inflectional paradigm, such as case, gender, and the difference between human vs. non-human referents. Case is evidently important in German, where there is a distinction between nominative (\textit{der/die/das}), accusative (\textit{den/die/das}) and dative (\textit{dem/der/dem}), as well as genitive (\textit{dessen/deren/dessen}). The German system also shows gender distinctions (the triplets above are masculine/feminine/neuter), which is also relevant for Dutch (\textit{die/dat}). Finally, the distinction between human vs. non-human referents is a relevant factor for English (\textit{who(m)/which}).

There is yet another issue related to gender which is worth mentioning here. As pointed out above, there were various relative pronouns in Old English, out of which \textit{that} came to be the “winner”: it started to appear more frequently in the environments where the other potential candidates also appeared. This is expected not to be possible for different members of the same inflectional paradigm, as in a paradigm all items show sensitivity towards the referent: the elements inserted into the syntax from such paradigms (in the sense of Wunderlich and Fabri 1995) carry the relevant paradigmatic inflectional information. After English \textit{that} lost its case and gender specification, i.e. it was extended to other slots in the paradigm, leading to the ultimate loss of that paradigm, it could be reanalysed as a complementiser. By contrast, German \textit{das} carries both case and gender specification (it is nominative/accusative and neuter), and Dutch \textit{dat} carries gender specification (neuter).

In order to illustrate how this works, let us have a look at the system in Standard Dutch. In this variety, the paradigm contains die (common gender) and dat (neuter gender). Out of the two, neuter is

---

28 Under this view, paradigms are subcomponents of the inflectional component of grammar and they operate at the interface between the lexicon and the syntax (this is in line with the concept of the Numeration of the Minimalist Program, see Chomsky 1995). According to this ‘early insertion’ approach, the output forms of the paradigms are atomic entities which project into the syntax, where they can be involved in procedures such as agreement checking.

29 As should be evident, only elements that have the same distribution (that is, they can appear in the same syntactic environment) are members of a paradigm; in this case, only relative pronouns are included but not other pronouns (such as demonstratives).
less marked (cf. Wunderlich 2004:377–378). Along the lines of this markedness hierarchy, the two genders can be described using the single feature [non-neuter] such that only the positive value is actually added (this is referred to as “a radical version of the theory of underspecification” by Wunderlich and Fabri 1995:242), namely to the more marked value: die is [+non-neuter]. The individual members of a paradigm can be summarised in an inheritance tree (see Wunderlich and Fabri 1995:255–259 and Wunderlich 1996 for verb inflection). This is given in (17) below.

(17) REL_PRON[+D]
die[+non-neuter] dat

Under this view, the more marked form die is inserted only if the feature [+non-neuter] is needed; in all other cases, the unmarked form is inserted. The same kind of argumentation can be applied to case, where the unmarked form is the nominative (as Wunderlich 2004:376, based on Wunderlich 1997, puts it, non-nominative cases can be described “by the feature [+hr] ‘there is a higher role’, whereas nominative bears no such feature”).

The underspecified member dat receives its [–non-neuter] interpretation by virtue of being a member of the paradigm, all the members of which are specified as belonging to the category [+D]. As it constitutes the unmarked member of the paradigm, this element has fewer features to lose as the other: it is more susceptible to grammaticalisation than die (as discussed in section 2, this step is attested only in Vlaams-Brabant, see Boef 2013:141). Strictly in this sense, one may go further and say that neuter relative pronouns do not have to actually lose a gender feature but they have to be interpreted as elements independent of the paradigm, for which the loss of [+D] is necessary.

A possible prediction here is that pronouns not appearing in contrastive paradigms become complementisers, while the same step is more difficult if the contrastive paradigm is still present. This is true in the sense that pronouns appearing in contrastive paradigms are apparently never reanalysed: the reanalysis of English that was completed after the original masculine and feminine relative pronouns were lost.

However, the above statement is false in the sense that pronouns not appearing in contrastive paradigms mentioned in connection with Old English, the relevant processes are gradual and overlap in time so that the reanalysis of the neuter pronoun clearly does not start only after the masculine and feminine (and plural) forms have disappeared. In other words, for a considerable period of time, we find both examples where that behaves as a regular neuter pronoun and also ones where it behaves as an element located in C, and many examples are ambiguous; van Gelderen (2004:84–87) assumes that as a pronoun it is sometimes in [Spec,CP] and sometimes in C. Reanalysis is completed once that cannot be treated as a specifier any more and once it is specified for finiteness, which is not marked by relative pronouns. Under this view, one important aspect of flexibility concerns the possibility of inserting the relative pronoun directly into C, that is, to merge it directly with the TP. This configuration also accounts for why prepositions are not pied-piped with but (unlike with other relative pronouns, see Traugott 1992:230–231, with the exception of R-pronouns, see Ringe and Taylor 2014:444–445) but are rather stranded (though van Gelderen 2004:84–85 provides some examples with pied-piping), contrary to Allen (1977), Traugott (1992:227, 230–231) and Mitchell (1985). In addition, this kind of variation in the syntactic position also addresses the observed variation in verb movement in relative clauses with but (van Gelderen 2004:85). Apart from relative clauses, the option of merging operators directly with the TP can be detected in polar and alternative questions introduced by whether as well, which also shows variation in word order with respect to verb movement (Bacskai-Atkari 2019). This option is not unique to Middle English but it can be attested in other Germanic languages as well, as shown by Bayer and Brandner (2008) and Bayer (2014) for embedded interrogatives in South German varieties. Just like in relative clauses, this is possible because the operators carry the relevant clause-typing features; however, there is no reason to assume that these elements were “proper” grammaticalised complementisers (see Bacskai-Atkari 2019).
are not necessarily reanalysed. As mentioned earlier, relative pronouns are regularly demonstrative-based in German and in Dutch; however, some dialects show wh-based relative pronouns as well. For instance, was in Hessian is such a pronoun: unlike the complementiser wo, it is restricted to neuter antecedents (Fleischer 2017) but it has no masculine/feminine wh-based counterparts, leaving this element without a proper contrastive paradigm. This resembles the case of English that historically, but it is at the same time evident from the distribution of this element that reanalysis into a complementiser has not taken place. A similar phenomenon can be observed in Dutch for the element wat dialectally, which is specifically used with neuter antecedents: this element usually occurs without a contrastive paradigm as the masculine/feminine counterpart wie is significantly rarer (see Krogull et al. 2017:161 citing van der Horst 2008b:1683). Again, reanalysis has not taken place even though a contrastive paradigm is not (yet) given. It is interesting to note that both processes, that is, the reanalysis of a pronoun into a complementiser (English) and the introduction of a new relative pronoun (German and Dutch), involve the neuter pronoun, which is evidently the most unmarked of all the forms (cf. Wunderlich 2004).

This strongly suggests that while reanalysis is apparently a strong tendency, it cannot be taken to be a strict economy constraint.

Table 3 shows the inventory of Germanic complementisers in argument relative clauses (without the aim of completeness):

<table>
<thead>
<tr>
<th>Pronoun-based Operator</th>
<th>d-based</th>
<th>wh-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>der/die/das (Ger.)</td>
<td>who/whom/which (Eng.)</td>
</tr>
<tr>
<td></td>
<td>die/dat (Dutch)</td>
<td>welcher/welche/welches (Ger.)</td>
</tr>
<tr>
<td></td>
<td><em>that</em> (Eng.)</td>
<td>was (Hessian)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wat (Dutch dial.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>what</em> (Eng. dial.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>was (Bavarian)</td>
</tr>
</tbody>
</table>

Table 3: The inventory of relative markers

As indicated, all the potential combinations (regarding the source of the elements) are attested, suggesting that there is no one-to-one correspondence between any of these factors.

5. Case and the Noun Phrase Accessibility Hierarchy

An additional question that arises in connection with the English pattern is why a relative pronoun strategy is maintained/renewed even if a complementiser is available. One relevant point is that relative pronouns identify the gap in the relative clause and are highly relevant in terms of processing. Note that this factor is not deterministic: while it is expected to contribute to maintaining the pronoun strategy and to foster its spread, what it cannot account for is the appearance of that strategy in a system relying on complementisers. In other words, if relative pronouns were a necessity for processing, we would not expect Stage 2 of the relative cycle to be possible at all (with single that as a complementiser in Middle English). \(^{33}\)

\(^{33}\) Apart from Middle English, this would rule out Scandinavian relative clauses: as mentioned in section 1, Scandinavian languages rely on the complementiser strategy (som in Mainland Scandinavian, sem in Icelandic). This provides an argument in favour of the view discussed in section 3, according to which the introduction of the new, wh-based relative pronouns was due to analogy with embedded interrogatives and to Latin/French influence.
Regarding the gap in the relative clause and processing, Keenan and Comrie (1977:66) observed relevant differences along what they dubbed the Noun Phrase Accessibility Hierarchy. The hierarchy is given below (oblique object referring to complements of prepositions, genitive referring to possessor phrases):

\[(18) \text{subject} > \text{direct object} > \text{indirect object} > \text{oblique object} > \text{genitive} > \text{object of comparison}\]

This is an implicational hierarchy in which subjects are the least marked and objects of comparison are the most marked functions. The original observation of Keenan and Comrie (1977) pertained to the occurrence of resumptive pronouns: these are more likely to appear lower in the hierarchy, such that if resumptive pronouns are obligatory at a given point, then they will be obligatory for all lower functions (as far as they are available in the given language), but they may be optional or even prohibited in higher functions. Conversely, if resumptive pronouns are prohibited at a given point, then they will be prohibited in all higher functions as well, but they may be optional or obligatory in lower functions. Resumptive pronouns are rare in the subject function, which is the highest-ranked function. A further implication concerns the occurrence of relative clauses in a given language: the subject function can always be relativised, while lower functions can only be relativised if all the functions ranked higher are. To provide a simple example: if a language relativises obliques, we can be sure that it also relativises subjects, direct objects and indirect objects.

The Accessibility Hierarchy makes good predictions in terms of relative frequencies as well, though the picture is less clear. Various studies on relative clauses confirm that the subject is the most frequently relativised function, followed by direct objects, but indirect objects are less frequent than obliques: this is discussed for English relative clauses in dialects of Britain by Herrmann (2005:51–52) in detail (see also the observation of Fleischer 2004a:61 concerning the infrequency of indirect object relative clauses in corpora). Herrmann (2005:53–59) also examined the frequencies of the individual strategies: for each strategy, it is true that they show higher frequencies in the functions that are higher in the Accessibility Hierarchy. In particular, the pronoun strategy is attested in all functions and that-relatives occur in all functions except the genitive; the comparatively novel strategy involving the complementiser what seems to be spreading from the subject function.

Another aspect of the relation between frequencies and the Accessibility Hierarchy concerns the relative distribution of the individual strategies with respect to each other. As relative pronouns lexicalise the gap, they may be similar to resumptive pronouns in that they can ease processing for less accessible gaps (see also Romaine 1984:440, Fleischer 2004a:230). This was formulated by Hawkins (1999:252–258) as the Filler-Gap-Complexity Hypothesis: according to this, [−case] elements are expected to occur in functions that are higher in the Accessibility Hierarchy, while [+case] elements are expected especially in lower functions. Under this view, we expect to find cut-off points analogous to the ones with resumptive pronouns.

Indeed, there are some remarkable similarities that arise, while there are obvious differences as well. For one thing, the occurrence of resumptive pronouns is compared to the non-occurrence of the same element (pronoun vs. zero); clause typing is independently carried out by a complementiser in the left periphery (so that no choice in the form “pronoun vs. complementiser” arises). Relative pronouns, however, primarily compete with overt complementisers (that is, the relative clause is either introduced by an overt relative pronoun or by an overt relative complementiser), so that the question “pronoun vs. complementiser” is more sensible to ask.

---

34Keenan and Hawkins (1987), Hawkins (1995), and Kirby (1996) ground the Hierarchy in processing factors (subject relative clauses are the easiest to process).

35This implication is a strong tendency, yet there are languages that contradict it, as discussed by Keenan and Comrie (1979:652); see Fleischer (2004a:213–214) for discussion.

36This does not exclude the possibility of the combination of the two (resulting in doubly filled COMP patterns) or the zero relative strategy but both of these are considerably rarer and restricted in English as well as in other Germanic varieties, to the extent they are possible at all. The doubling pattern is not examined as an independent strategy by Herrmann (2005). Regarding zero relatives, the asymmetries between the various relativised functions do not follow from the Accessibility Hierarchy in a way analogous to...
Regarding the emergence of the *wh*-pronoun strategy in English (dating back to the 12th century), Romaine (1982:61) and Gisborne and Truswell (2017:31–32) observe that the *wh*-strategy first appeared very low on the scale, namely with genitives and with objects of prepositions (see also van Gelderen 2004:87, citing Dekeyser 1986:100–101 and Allen 1977:197–199). Gisborne and Truswell (2017:31–32) consider the *wh*-strategy to be a secondary option in this period (as Romaine 1982:152 puts it, this option appears to have entered the system through the “back door”). In other words, the complementiser strategy with *that* was more typical in the higher functions; a similar observation is made by Traugott (1992:226) regarding the distribution of the original Old English complementiser *he*. The relevant preferences were present in the English language throughout its history. In her study on British dialects, Herrmann (2005) also observes that the complementiser strategy is more likely to occur higher on the scale. In this respect, *wh*-based relative pronouns indeed show similar behaviour to resumptive pronouns in other languages (Kirby 1996; but see the comments above and the critical evaluation by Gisborne and Truswell 2017:32–35). For Middle Scots, Romaine (1980:228–229) and Romaine (1982:144–157) show that the proportion of *wh*-relatives was altogether higher in the lower functions. However, as the proportion of *wh*-relatives reaches 14% in restrictive relative clauses, this data set can hardly be considered as representative of the initial stage.\(^\text{37}\)

In sum, while there are clear asymmetries between the individual strategies with respect to the function of the gap, there appear to be no clear-cut lines in the way it is attested for resumption. This holds not only for Middle English and present-day dialects, but also for Early Modern English data. Bacskai-Atkari (2020) describes relevant asymmetries in the King James Bible. Table 4 shows the refined results for the data in Table 1 (taken from Bacskai-Atkari 2020:104).

<table>
<thead>
<tr>
<th>Role</th>
<th>who</th>
<th>whom</th>
<th>which</th>
<th>that</th>
<th>as</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>14</td>
<td>–</td>
<td>41</td>
<td>169</td>
<td>2</td>
</tr>
<tr>
<td>(226)</td>
<td>(6.19%)</td>
<td></td>
<td>(18.14%)</td>
<td>(74.78%)</td>
<td>(2.21%)</td>
</tr>
<tr>
<td>direct object</td>
<td>–</td>
<td>–</td>
<td>14</td>
<td>78</td>
<td>35</td>
</tr>
<tr>
<td>(127)</td>
<td></td>
<td></td>
<td>(11.02%)</td>
<td>(61.42%)</td>
<td></td>
</tr>
<tr>
<td>PP complement</td>
<td>–</td>
<td>–</td>
<td>11</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>(21)</td>
<td></td>
<td></td>
<td>(52.38%)</td>
<td>(19.05%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: The distribution of relative markers in the KJB (Genesis)

Table 5 shows the refined results for the data in Table 2 (taken from Bacskai-Atkari 2020:104).

---

\(^\text{37}\)This actually applies to Middle English texts as well. The combination *which that* represents Stage 2 of the relative cycle; for this combination, the Corpus of Middle English Prose and Verse gives altogether 6 hits in Chaucer’s *Troilus and Criseyde* (from the 1380s; available in the Campsall, the Harleian and the Cambridge University Library manuscripts). Out of these, 4 occurrences are in subject relative clauses, one is in a direct object relative clause and one in an oblique relative clause (where *which* is part of a PP).
As can be seen, the strongly norm-oriented new version uses that in about 30% of the cases in all functions, while the original version shows clear asymmetries in that the complementiser strategy (using that and to a lesser extent as) is predominant in the subject function but not in the object functions. The original version thus shows the distribution that was discussed for Middle English and modern regional dialects above and it is congruent with the hypothesis that the pronoun strategy should be more substantial in the lower functions of the Accessibility Hierarchy. This difference is levelled out in the newer version.

Regarding the changes in subject relatives, Bacskai-Atkari (2020:105) points out that there are two major changes to be observed. First, the proportion of that-relatives decreases in favour of wh-relatives; the distribution of the two strategies is ultimately very similar across the three functions examined here. Second, original which-relatives with a human referent were changed to who-relatives, but due to the general decrease in the use of that-relatives, the proportion of which-relatives in subject relatives is actually higher than in the original version.

As far as English is concerned, the main difference in the data concerns the subject vs. lower functions. This is not necessarily so in other languages, though. In his study on German dialects, Fleischer (2004a:230) observes that some of the examined varieties show a particle strategy for the subject and direct object relative clauses (such as the Low Alemannic dialect of Basel), while other varieties display a particle strategy for subject, direct object, and indirect object relative clauses (such as the Low Alemannic dialect of Oberrotweil). In other words, subjects and direct objects pattern together, while indirect objects vary concerning whether they pattern with these two or whether they pattern with obliques.

Table 5: The distribution of relative markers in the new version (Genesis)

<table>
<thead>
<tr>
<th>Role</th>
<th>who</th>
<th>whom</th>
<th>which</th>
<th>that</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>106</td>
<td>–</td>
<td>52</td>
<td>68</td>
</tr>
<tr>
<td>(226)</td>
<td>(46.90%)</td>
<td>–</td>
<td>(23.01%)</td>
<td>(30.09%)</td>
</tr>
<tr>
<td>direct object</td>
<td>–</td>
<td>30</td>
<td>62</td>
<td>35</td>
</tr>
<tr>
<td>(127)</td>
<td>(23.62%)</td>
<td>–</td>
<td>(48.82%)</td>
<td>(27.56%)</td>
</tr>
<tr>
<td>PP complement</td>
<td>–</td>
<td>11</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>(21)</td>
<td>(52.38%)</td>
<td>–</td>
<td>(19.05%)</td>
<td>(28.57%)</td>
</tr>
</tbody>
</table>

The data presented by Fleischer (2004a:227) show that the predominant strategies across dialects of German are either the pronoun

---

38 In fact, this has its historical origins as well: at least from the 17th century onwards, that for human referents has often been considered as impolite or inappropriate by prescriptive grammarians (see Ball 1996:249–250). The same is not attested in written texts (much more than in spoken language) started rather in the 17th century, in line with the fact that prescriptive grammarians started to advocate the wh-strategy. In other words, there is no reason to assume that the decrease in the frequency of that should be attributed to some language-internal pressure; rather, as far as written and/or norm-oriented language use are concerned, the observed changes reflect conscious considerations. Note also that the new version is indeed very norm-oriented in using exclusively whom for the object functions (which is otherwise not the case, cf. van Gelderen 2004:80); as discussed in section 2, this also means that the new version cannot be taken to reflect Present-Day English reliably. What matters for the present discussion is rather where the deviations from the original version occur.

39 The difference between the two versions emerges perhaps even more clearly from two other data sets presented in the cited study. There are altogether 3569 subject relatives using who in the Old Testament in the newer version: out of these, 74.64% had that in the original version (Bacskai-Atkari 2020:100). This contrasts sharply with what is reported for Lower functions. There are altogether 398 direct object relatives with whom in the entire newer version: out of these, only 2.51% were originally that-relatives (Bacskai-Atkari 2020:101). Obliques are similar (altogether 265 occurrences with whom, with 0.75% being originally that-relatives); the 2 occurrences of indirect objects with whom both used whom in the original version as well (Bacskai-Atkari 2020:101).

40 As Herrmann (2005:41) points out, which was possible with human referents in Middle English (cf. Mossé 1991) and the grammaticalisation of which as [–human] started in the 16th century (cf. Nevalainen and Raumolin-Brunberg 2002). According to Austen (1985:18), the use of which with human antecedents gradually came to be marginalised during the 18th century and several grammarians of this period already saw it as an archaic feature; its availability also correlates with social status (the higher the social status, the less likely it is to appear with human referents). Ball (1996:246–247) makes a similar point. Naturally, there was also considerable variation concerning this change. Austen (1985:18) reports that while which survived longer in restrictive relative clauses than in non-restrictive relative clauses in general, this is not reflected in the Clift letters examined by her, where the distribution is exactly the opposite.

41 The data presented by Fleischer (2004a:227) show that the predominant strategies across dialects of German are either the pronoun
Somewhat simplistically, the observed asymmetries in relative clauses can be summarised as follows: English favours a subject/oblique pattern (oblique referring to all functions lower than the subject), while German favours a subject/direct object/oblique pattern (oblique referring to all functions lower than the direct object). Interestingly, these differences pattern well with the observed differences in the relevant case systems. English has a nominative–oblique system with few morphological distinctions (essentially restricted to pronouns) across varieties. German distinguishes between nominative, accusative and dative: this can be clearly observed in the standard variety (where it is in fact complemented by the genitive), while dialects differ in whether they morphologically distinguish between accusative and dative. Again, just as was observed with respect to the historical changes, the general case system is not fully mirrored by what can be observed in relative clauses, where grammaticalisation may lead to different patterns. The point is that there seems to be a correspondence between morphological (case) distinctions in the general system and in morphological distinctions in relative markers (case, as well as pronoun vs. complementiser). If so, the differences in the case system also have an effect on the distribution of relative markers (including complementisers); in this sense, the case system is not only relevant diachronically (in terms of specifier–head reanalysis) but also synchronically. This issue should be investigated by future research.

6. Conclusion

In this article, I examined relative markers in English and their distribution compared to the patterns attested in other Germanic languages. In several respects, such as the general preference for the complementiser strategy and the restricted use of doubling patterns (compared to embedded interrogatives), English is similar to other Germanic languages. On the other hand, it was shown that the English pattern is special among Germanic languages, which is due to the interplay of various factors. First, there are purely syntactic factors to be considered, which concern the difference between operators moving to [Spec,CP] versus complementisers base-generated in C: in English, both options are possible. Second, there are morphological factors as well, which primarily concern the availability of overt lexical elements. Third, the feature content of potential complementisers also has to be taken into account: relative pronouns can be reanalysed as complementisers only if they lose their case and gender features. In this respect, it was also shown that the effect of the case system plays an important role, not just in terms of reanalysis but also in terms of various asymmetries along the lines of the Noun Phrase Accessibility Hierarchy.

It seems that the particular setting in English is not dependent on a single parameter but on various factors that are otherwise present in other Germanic languages as well. It is ultimately the complex interplay of these factors that results in the particular setup found in English.

References


Brandt, Patrick and Eric Fuß, 2014. Most questionable pronouns: Variation between das- vs. was-relatives in German. Linguistische Berichte 239: 297–329.


