Structural and Sociolinguistic Factors Conditioning the Choice of Relativizers in Late Modern English: A Diachronic Study Based on the *Old Bailey Corpus*

Magnus Huber, University of Giessen

Abstract

The present study aims at broadening our understanding of the development of relativizer choice (pronouns, *that* and zero) in Late Modern English restrictive relative clauses. It investigates the effect of the linguistic factor groups ANIMACY OF THE HEAD and ROLE OF THE RELATIVIZER IN THE RELATIVE CLAUSE and combines these with the social factors GENDER and SOCIAL CLASS.

The data used in this study come from the *Old Bailey Corpus* (Huber et al. 2012). As a corpus of trial proceedings, it affords a glimpse into spoken Late Modern English. Because of its size, the time-span covered and the detail of sociolinguistic utterance-level annotation, the *Old Bailey Corpus* is an ideal database for a fine-tuned, quantitative-variationist study of relativizers in the 18th and 19th centuries.

It is commonly assumed that at the beginning of Late Modern English the relativizers *that* and zero were felt to be rather colloquial, particularly as far as the written mode was concerned. However, even in the very formal setting of trials as represented by the *Old Bailey Corpus*, *that* and zero account for as much as 76.1% of all relativizers in the 1720–1789 period. The frequency of *that* declined considerably during the two centuries investigated here (from 52.7% down to 27.8%) and it developed into a relativizer that predominantly marked non-human antecedents. Zero, on the other hand, increased in frequency (from 24.2% to 38.3%, thus becoming the most common relativizer) but remained a marker that mainly followed non-human heads. During the 18th and 19th centuries, human antecedents became increasingly marked by *who*, but there was little change with regard to *which*, marking about one-fifth of non-human heads in all subperiods. As to sociolinguistic factors, the use of the relative pronouns was promoted by male speakers, while female speakers led in the adoption of the zero relativizer.

1. Introduction: Previous research

Relativization is one of the better studied areas of English syntax and there is no shortage of synchronic studies. The history of English relative clauses (RCs) has also attracted considerable attention, especially with regard to Old and Middle English. There are fewer (though by now means few) descriptions of RC formation in the recent history of English.

Variationist studies are less frequent: corpus-linguistic studies on Early Modern English (EModE, 1500–1700) relativization include Rydén (1966, 1983), Dekeyser (1984), Rissanen (1984), Nevalainen and Raumolin-Brunberg (2002, 2003) and Johansson (2012).

Late Modern English (LModE, 1700–1900) relativization is not quite as well documented. Short selective overviews can be found in e.g. Strang (1970: 141–144), Görlach (1999: 86–87, 2001: 126–127), van Gelderen (2006: 217–219), Adamson (2007) and Aarts, López-Couso and Méndez-Naya (2012). Quantitative-variationist studies for LModE are even rarer than for the previous period: Ball's (1996) perceptive study investigates restrictive subject RCs from the 16th to the 20th century, considering, among others, the independent variables ANIMACY OF THE ANTECEDENT, SYNTACTIC ROLE and GENRE. As Ball's data also include British trial proceedings, it is possible to compare some of her findings to those in this article. The very detailed analysis by Johansson (2006) focusses on RCs in the 19th century. Her independent variables include ANIMACY OF THE ANTECEDENT, SYNTACTIC ROLE, GENRE as well as the GENDER and SOCIAL CLASS of the speaker or writer. However, Johansson does not systematically distinguish between restrictive and non-restrictive RCs, includes clefts and existentials, and does not consider the zero relativizer because it is difficult to retrieve. Her results are therefore not directly comparable to those in this article.

2. Relativization at the transition from EModE to LModE

By the early 1700s, the beginning of the period analyzed in this article, the inventory of relativizers had come to be the same as in Present Day English (PDE). It included the older relativizers *that* and zero, but also the more recent addition of the relative pronouns. The introduction of pronouns as relativizers has often been described as a change from above,[1] modelled on Latin or French patterns (e.g. Romaine 1982: 61; Dekeyser 1984: 76; Rissanen 1984: 420, 423) and motivated by the greater explicitness of the pronouns as to case and animacy of the head of the RC (Strang 1970: 142f). The fact that the pronouns first appeared in complex, formal registers and only later advanced to less complex,

---

[1] Changes from above "take place at a relatively high level of social consciousness, show a higher rate of occurrence in formal styles [and] are often subject to hypercorrection" (Labov 1990: 213).

informal ones (Dekeyser 1984: 77) is a strong argument for a change from above, as is the fact that they entered the system through the less accessible, lower positions of the Noun Phrase Accessibility Hierarchy (Keenan and Comrie 1977), i.e. as genitives or as objects (Romaine 1982: 62) and only later appeared as subjects.

Romaine (1982: 69, 71) maintains that by the end of the 17th century, the PDE division of work between *who, which* and *that* had been reached, but this is an overgeneralization. Even though the relativizer inventory around 1700 was the same as today, this does not mean that no changes have occurred in the function and distribution of relativizers over the last three centuries: "The crucial difference between [Middle English] and [PDE] is not the number of relatives, but the system that governs their distribution" (Dekeyser 1984: 61).

By the end of EModE individual relativizers had completely (or almost completely) attained some of today's distributional characteristics: zero was virtually blocked from non-restrictive RCs (Dekeyser 1984: 79–80; Johansson 2012: 779) and the particle *that* became increasingly confined to restrictive RCs as the pronouns became more frequent (Romaine 1982: 61; Dekeyser 1984: 69; Rissanen 1984: 429).

Who was a human relativizer from the time it entered the system (for *who* in EModE cf. Johansson 2012: 783–784). *Which*, on the other hand, was initially animacy-neutral but in EModE developed into a relativizer for non-human heads (Dekeyser 1984: 71, 79; Ball 1996: 246ff; Görlach 2001: 126–127; Johansson 2012: 784–786). This development was "virtually completed by 1700" (Dekeyser 1984: 71), especially in polite and formal English (Rissanen 1984: 429).

One change that started in EModE and continued in LModE was that the zero relativizer was increasingly disfavoured in subject position (Romaine 1982: 76ff; Dekeyser 1984: 78; Rissanen 1984: 430; Nevalainen 2006: 85; Johansson 2012: 782). One reason for this was the inherent ambiguity of subject contact clauses (Strang 1970: 143; see discussion in Section 4), contributing to their increasing stigmatization. There is conflicting evidence regarding the frequency of zero in spoken and written EModE. Strang (1970: 142) observes that at the end of EModE the zero relativizer came under attack—and retreated from—written English but showed no signs of decline in spoken English. However, she does not specifically look at restrictive relativization and

the example she provides is an existential sentence. Strang's claim is substantiated by Johansson (2012: 778), who finds a low frequency of zero in written EModE (1520–1560: 2%, 1600–1649: 7%) and a higher percentage in the spoken medium (1560–1599: 11%, 1600–1649: 24%, 1680–1719: 21%), but again these figures include non-restrictive RCs as well as clefts and existentials. Rissanen (1984: 430), on the other hand, does not find evidence for zero to be more frequent in spoken language.

As the frequency of the recent relativizer *who* increased in EModE, it gradually displaced *that* in clauses with human antecedents. Ball (1996: 246–247) observes that in 1650–1700, *who* came to be used more and more with human heads in restrictive subject relativization and caused *that* to be associated with non-human antecedents. The latter process was still in progress at the end of EModE (Dekeyser 1984: 79–80). At the same time, and because of the prescriptive pressure in favour of relative pronouns, *that* was felt to be more colloquial both in EModE (Rissanen 1984: 420) and LModE (Görlach 2001: 127). Van Gelderen (2006: 217–218) maintains that from 1700 on, *that* became more and more frequent at the expense of the pronouns, but that this development was curbed by prescriptive forces.

3. The Old Bailey Corpus

The database for this study is the *Old Bailey Corpus*, version 1.0 (OBC, Huber et al. 2012). It consists of 415 selected *Proceedings* from the Old Bailey, London's central criminal court (cf. Hitchcock et al. 2015), from 1720 to 1913. The OBC contains a total of 13.9 million words, with about 750,000 speech-related words per decade. The only exceptions are the first and last decades: in the 1720s, the *Proceedings* only sporadically contain verbatim reports and consequently not more than 72,500 words of direct speech are available for this decade, which were all included in the corpus. Also, since the publication of the *Proceedings* stopped in 1913, this last "decade" contains just four years and accordingly only 350,000 words were included here.

The *Proceedings* were taken down in shorthand and are thus a reasonably close representation of what was said in the courtroom, even though scribes, printers, publishers and the constraints of the printed medium would have acted as linguistic filters. The OBC thus offers the rare opportunity of analyzing speech-related texts in a period that has

been neglected both with regard to the compilation of primary linguistic data and the description of the structure, variability, and change of English.

A particular strength of the OBC is that it includes a high number of speakers and thus is a fairly representative sample of spoken, rather formal LModE in the courtroom setting. Moreover, every speaker turn has detailed annotation for sociobiographical (GENDER, SOCIAL CLASS, AGE), pragmatic (role in the trial) and textual variables (the shorthand scribe, printer and publisher of individual *Proceedings*). The OBC is therefore particularly suited for studies that correlate linguistic change and structural variability in LModE with the social context.

4. Data extraction and coding

RCs are comparatively frequent structures, so to obtain a manageable number of tokens, utterances amounting to 20,000 words per decade (10,000 by females and 10,000 by males) were extracted randomly from the OBC, resulting in a subcorpus of 400,000 words (henceforth OBC^S). This was searched for the relativizers *who*, *whose*, *whom*, *which*, *that* and zero. Other relativizers, like question words (*the place <u>where</u>*, *the time <u>when</u>*, *the reason <u>why</u>*) are not considered in this study, nor are compound forms such as *whereof* and *whereby*, which are rare in the OBC and mostly found in formulaic legal language (e.g. *How say you, James Annesley, are you guilty of this Felony <u>whereof</u> you stand indicted or not guilty?* OBC-t17420715-1).[2] Also, there are no instances of the non-standard relativizers *as* and *what* in the OBC.

Zero relativizers in non-subject position were retrieved by extracting from the POS-tagged version of OBC^S all sentences that contained two adjacent noun phrases plus a verb, such as *every thing they could* or *the books the prisoner brought*. This procedure ensured a 100% recall but precision was low, necessitating extensive manual and semi-automatic post-processing. In the end, fewer than 10% of the NP NP Vb sequences turned out to be part of RCs. Every care was taken in identifying zero relatives and I am confident that the results present an accurate picture. However, the fact that post-processing was partly manual means that

---

[2] Quotes from the OBC are referred to by the trial identifier, which is composed of a t (for trial), followed by the date of the publication of the Proceedings (yyyymmdd) and the consecutive trial number of the sessions (-xx).

some zero RCs may have been overlooked and the figures below therefore represent minimum numbers.

Zero in restrictive subject relativization is even more difficult to extract automatically because on the surface such structures look like subject-predicate sequences. In addition, unless they occur in cleft or existential sentences, these constructions are highly ambiguous (cf. Romaine 1982: 78). While

(1) I saw a woman beat a man with a basket (OBC-t17531024-40)

is a possible candidate for a subject zero relativizer (equivalent to *I saw a woman **who** beat a man with a basket*), it can also be read as an infinitive construction (roughly equal to *I saw a woman **beating** a man with a basket*) or as containing a zero complementizer (corresponding to *I saw **that** a woman beat a man with a basket*). Since the disambiguation of such structures proved impossible, they had to be disregarded for the present study.

Some general remarks regarding the RCs retrieved from OBC^S are in order before I go on to describe the data and coding. In a study based on the *Corpus of Early English Correspondence* Nevalainen and Raumolin-Brunberg (2003: 73ff) found that by 1600 the earlier relativizer *the which* had virtually disappeared from letters. Similarly, *the which* is last found in the 1570–1640 period of the *Helsinki Corpus* (Rissanen et al. 1991). Nevertheless, there is one token of this in OBC^S, from 1755, in non-restrictive relativization (not further analyzed in this paper):

(2) my master sent me for a constable, [the which he got a knowledge of and ran away] (OBC-t17550515-23)

The which had become marginal by the beginning of the 18th century: the entire *Old Bailey Proceedings* (125 million words; Hitchcock et al. 2015) yield only 92 tokens of *the which*, 82 of these occurring before 1700, and by far the most (88) in non-spoken, rather formulaic passages. The last attestation of *the which* in the *Proceedings* is from 1794.

As mentioned in Section 2, with the spread of the relative pronouns, *that* began to be relegated to restrictive RCs. Still, examples of non-restrictive *that* can be found in OBC^S, e.g.

(3) She was permitted to fetch her uncle, [that she said she had] (OBC-t17950114-13)

But with just six tokens (one of them dubious) non-restrictive *that* is again marginal. The last attestation in OBC^S is from 1836.

The *of which* genitive is found in the OBC^S data, e.g.

(4) 37 was an empty house, [of which Cook was caretaker] (OBC-t18910406-330)

but there are no clear cases where this is used in restrictive relativization. Similarly, *whose* only occurs four times and is found exclusively with human antecedents in restrictive RCs, as in

(5) I frequently buy of persons [whose names I do not know] (OBC-t18350105-376)

The subcorpus yielded a total of just under 2,500 RCs in the following categories:

Restrictive RCs: 1,421 restrictive RCs were identified in OBC^S, including the following example:

(6) he had taken a dose [that might have proved fatal] (OBC- t18720108-117)

Distinguishing between restrictive and non-restrictive RCs was not always straightforward, as illustrated in (7):

(7) the prisoner and a young man, [which she called her brother], came unto my shop (OBC-t17520408-13)

Disregarding the presence of the comma for the moment,[3] (7) can be read as containing either a restrictive or a non-restrictive RC, depending on

---

[3] The punctuation rule of demarcating non-restrictive RCs by commas was (and is) often disregarded and cannot therefore be used as a foolproof criterion for the classification of RCs in the OBC. Compare the following example of a non-restrictive RC not separated from the matrix clause by a comma: *the Money in*

whether *a young man* represents the whole or only part of the referent in the focus of the speaker. If the referent is one particular young man only, then the RC merely provides additional information and is non-restrictive. If, however, the speaker's intention is to single out one specific individual from the entire class of young men in London / Britain / the world*,* then the RC gets a restrictive reading. Where the trial context was of no help in disambiguation, the structure was conservatively classified as a non-restrictive RC (including (7)).

There is a considerable decrease in the relative frequency of restrictive RCs over the two centuries considered in this paper, especially from the mid-19th century onwards, from an overall 388 clauses per 100,000 words in Subperiod 1 (1720–1789) down to 363 in Subperiod 2 (1790–1849) to as low as 285 in Subperiod 3 (1850–1913). As can be seen in Table 1, this overall decline is caused entirely by RCs in which the relativizer is in subject position: compare the 209-153-105 drop to the relatively stable 179-209-180 figures for non-subject RCs:

Table 1. Relative frequency of restrictive RCs, 1720–1913

Subperiod	Total words	SBJ RCs per 100T words	Non-SBJ RCs per 100T words	Total per 100T words
1720–1789	139455	209 (N=291)	179 (N=250)	388 (N=541)
1790–1849	124652	153 (N=191)	209 (N=261)	363 (N=452)
1850–1913	150198	105 (N=157)	180 (N=271)	285 (N=428)

It is possible that the fall in relative frequency of restrictive RCs was a result of the rise of non-finite *-ing* and *-ed* participle structures (*I was sitting by the woman **playing** the rogue*, OBC-t17600709-21; *a ginger moustache **waxed** at the ends*, OBC-t19120109-42), since these structures are alternatives to RCs only when the relativizer is in subject position.

Non-restrictive RCs: OBC[S] contains 975 non-restrictive RCs, illustrated by (8):

(8) I saw Cuthbert with a weapon in his hand, [which I took to be a chopper] (OBC-t18970914-563)

---

this Paper [which is 30 l. 1 s. 6 d.] were found in his Pocket (OBC-t17350226-14).

In Present Day Standard English, non-restrictive RCs usually only allow relative pronouns, while zero and *that* are blocked. All in all, the OBC[S] data show that these restrictions were already in place in LModE: of the 914 non-restrictive RCs, none is relativized by zero and only seven (0.8%) have *that* as a relativizer, as in the following example:

(9)   We were talking about Lutwicke [that was taken some time
      ago] (OBC-t17621020-11)

Since the constraints on relativizer choice in restrictive and non-restrictive RCs are clearly different, the latter are not further considered in this paper.

Cleft sentences: There are 56 cleft sentences, including the following:

(10)   it was his mother-in-law [who did it] (OBC-t18840421-458)

Although the constructions found in clefts formally resemble RCs, their status as such is controversial as they do not modify a head (cf. e.g. Collins 1991). As with non-restrictive RCs, relativizer choice in cleft sentences differs from that in restrictive RCs. For example, in PDE, *which* is marginal in clefts. In line with this, Huber (2009) found that in the OBC, *which* does not occur in clefts at all and that in contrast to restrictive and non-restrictive RCs clefts allow zero in subject relatives:

(11)   Bill Newton said it was me [took the pigeons]
       (OBC-t18451215-291)

As clefts behave differently from other RCs in the OBC, they are also disregarded in this study.

Existentials: 40 existential sentences were extracted from OBC[S], among them the following:

(12)   there is a closet in the front room [which holds a bed]
       (OBC-t18000528-1)

Whether the dependent construction in existentials actually is a RC is debated (cf. Ball 1996: 236). As with clefts, the OBC data suggests that

the rules for relativizer choice are different for existentials (and they were accordingly excluded from the present study). For example, like clefts they can select the zero relativizer in subject position (Huber 2012):

(13)   there is a bell [rings at eight o'clock] (OBC-t18261026-34)

The 1,421 restrictive RCs analyzed in this study were coded for the dependent variable RELATIVIZER and the independent variables time PERIOD, ANIMACY of the head, ROLE of relativizer in the RC, speaker GENDER and SOCIAL CLASS of speaker. These variables will be described in the following.

The dependent variable RELATIVIZER has three main variants:
- *that* (607 tokens)
- pronoun = PRN (398 tokens, comprising 202 *which*, 181 *who*, 11 *whom*, 4 *whose*)
- zero (416 tokens)

PERIOD: the corpus was divided into three subperiods (P1–P3), spanning 60 to 70 years each:
- P1: 1720–1789 (541 tokens)
- P2: 1790–1849 (452 tokens)
- P3: 1850–1913 (428 tokens)

The variants of ANIMACY of the head are:
- human (468 tokens, including 454 nominal and 14 pronominal heads with human referents)
- non-human (952 tokens, comprising 17 animate and 935 inanimate heads). Human groups and organizations like *class of people*, *regiment* or *company* were classified as non-human

The syntactic ROLE of relativizer in the RC has two variants:
- subject position (639 tokens)
- non-subject position (782 tokens, comprising: object 472, adverbial 304, genitive 4, complement 2)

The variable speaker GENDER shows the following distribution of tokens:
- female (668 tokens)
- male (753 tokens)

SOCIAL CLASS: the OBC uses the HISCLASS scheme (van Leeuwen and Maas 2011) to indicate class membership of the speakers. HISCLASS comprises 13 social classes, which for the present study were conflated into two classes:
- higher (non-manual professions, HISCLASS 1–5, 332 tokens)
- lower (manual professions, HISCLASS 6–13, 243 tokens)

The social class membership of the speaker is known for 575 RCs, i.e. 40.5% of the tokens. Using occupations to assign LModE speakers to social classes is of course rather crude and mechanical. It probably distorts 18th and 19th century sociolinguistic reality by imposing our present-day conception of social classes on the past. It also does not consider social networks or communities of practice. Nevertheless, the justification for the HISCLASS approach is that occupations are often the only social indicator found in the OBC and they are easily operationalized even with a large number of tokens. Future work will have to devise ways of teasing out more sociolinguistically relevant information from the OBC.

5. Analysis: The development of relativizers in Late Modern English

Before having a closer look at the effect of individual variables and their combinations on the choice of relativizers (Section 5.1) and their semantics (Section 5.2), the present section summarizes the results of binomial logistic regressions of the data using Rbrul (Johnson 2009). The models themselves can be found in the Appendix. The regressions were performed for each of the relativizers as the application value (PRN = models *1*, *that* = models *2*, zero = models *3*). Each regression was performed twice: models *a* are based on 1,420 tokens (this excludes the factor group SOCIAL CLASS as well as one token for which ANIMACY is

unknown); models *b* are based on the 575 tokens for which both GENDER and SOCIAL CLASS are specified.[4]

In Model *1a* (application value: PRN; N=1,420) the factor groups ANIMACY, ROLE, GENDER and PERIOD were selected as significant: PRN is somewhat more likely to be selected with human heads and by males, while subject position has the strongest positive effect. The interaction between ANIMACY and ROLE is not significant. Further, PRN becomes more and more likely as we progress from Subperiod P1 to P3.

The results of a logistic regression performed on only those tokens where the SOCIAL CLASS of the speaker is specified (Model *1b*; application value: PRN; N=575) does not substantially alter the picture: the factor weights change a little, but human heads, subjects and male speakers still promote the use of PRN. Note also that SOCIAL CLASS is not significant and no interaction was found in the ANIMACY-ROLE and GENDER-SOCIAL CLASS pairs. The only real difference is found in the factor group PERIOD, where P1 and P2 get reversed.

In Model *2a* (application value: *that*; N=1,420) ANIMACY, ROLE, GENDER and PERIOD turn out to be significant in the selection of *that*. This time non-human heads and females show a slight preference for *that*. Subject positions again have the strongest effect on the selection of *that*. This is because in subject positions *that* is the only alternative to PRN, while in non-SBJ positions *that* competes with zero as an alternative. *That* becomes less and less likely from P1 to P3. As in Model *1a*, there is no significant ANIMACY-ROLE interaction.

Adding the factor group SOCIAL CLASS to the regression (Model *2b*; application value: *that*; N=575), confirms that the particle is preferred with non-human heads and in subject position. Apart from the fact that P1 and P2 swap their places in comparison to Model *2a* (but their factor weights are rather similar in Models *2a* and *2b*), the major difference to Model *2a* is that GENDER is dropped. As with PRN, SOCIAL CLASS is not significant. The regression found no significant ANIMACY-ROLE or GENDER-SOCIAL CLASS interaction.

As zero never occurs in subject relativization, ROLE was not fed into the regression for Model *3a* (application value: zero; N=1,420). ANIMACY, GENDER and PERIOD are significant, with non-human heads

---

showing a strong and females a moderate likelihood to select zero. Zero also becomes slightly more likely in the later subperiods.

Including SOCIAL CLASS in the regression for Model *3b* (application value: zero; N=575) does not substantially change the results. The effects of ANIMACY and GENDER are almost the same as in regression 3a, but PERIOD got dropped from the model. Again, SOCIAL CLASS is not significant. There is no significant interaction in GENDER-SOCIAL CLASS.

All in all, the regressions yield rather consistent results, which are generalized in Table 2:

Table 2. Overview of factors favouring PRN, *that* and zero.
Round brackets indicate that in Regressions b the factor group was dropped or, in the case of PRN, that the order of factors was reversed

Factor group	PRN	*that*	zero
ANIMACY	human	non-human	non-human
ROLE	subject	subject	✕
GENDER	male	(female)	female
SOCIAL CLASS	-	-	-
PERIOD	(P1 < P2) < P3	P1/P2 > P3	(P1 < P2/P3)

The following sections present a detailed and in-depth investigation of these trends. A general overview and discussion of the main results can be found in Section 6.

5.1 Relativizer choice
Table 3 provides an overview of the relativizer forms in restrictive RCs over three subperiods:

Table 3. Restrictive RCs, forms of relativizer, 1720–1913

Period	*that*	zero	PRN				Sum
			which	*who*	*whom*	*whose*	
1720–1789	285	131	75	44	5	1	541
1790–1849	203	121	60	64	1	3	452
1850–1913	119	164	67	73	5		428
Sum	607	416	202	181	11	4	1,421

Figure 1 summarizes these figures by subsuming the relativizers *which, who, whom* and *whose* under the category 'relative pronoun' (PRN):
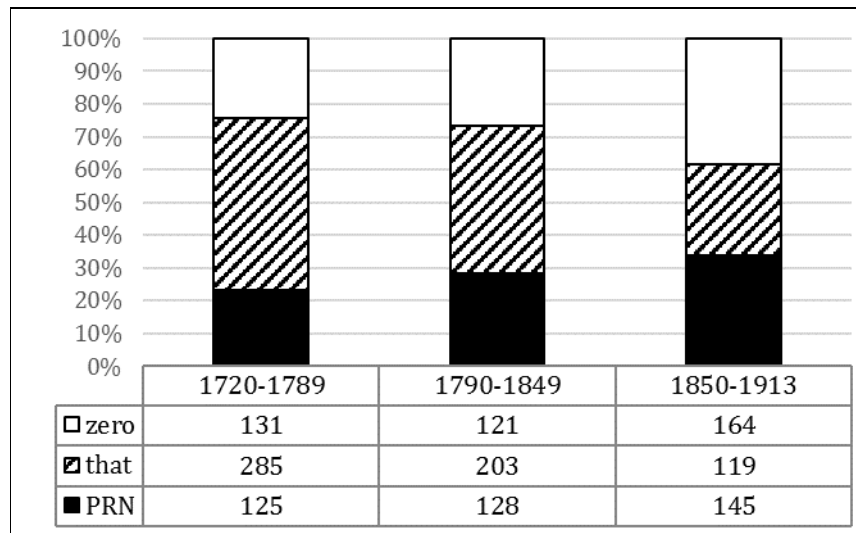
Figure 1. Relativizers, overall development[5]

PRN vs rest: P1–P2 p=0.06(*), P1–P3 p<0.001***, P2–P3 p=0.075(*)
Zero vs. rest: P1–P2 p=0.357ns, P1–P3 p<0.001***, P2–P3 p<0.001***
That vs. rest: P1–P2 p=0.015*, P1–P3 p<0.001***, P2–P3 p<0.001***

From P1 to P3, Figure 1 shows a moderate increase in PRN (23.1% to 33.9%), a somewhat higher rise of the zero relativizer (24.2% to 38.3%), accompanied by a considerable fall of the particle *that* (52.7% to 27.8%). Dekeyser (1984: 66) found 28.7% pronouns, 48.1% *that* and 23.3% zero in restrictive RCs in the first half of the 17th century. Comparing these figures with the picture in P1, there seems to have been surprisingly little change over the next 100 or so years. The somewhat lower percentage in P1 of PRN (23.1%) as compared to Dekeyser's data is probably due to

---

[5] The significance levels in this article are as follows:

Symbols	Description	p Value
ns	not significant	p > 0.1
(*)	marginally significant	p ≤ 0.1
*	significant	p ≤ 0.05
**	highly significant	p ≤ 0.01
***	extremely significant	p ≤ 0.001

the fact that 40% of Dekeyser's corpus was made up of prose and another 10% of poetry (where the rate of pronouns was considerably higher than in his written-to-be-spoken texts; Dekeyser 1984: 77), whereas the OBC contains real utterances reproduced in the written medium and thus is closer to spoken language.

Strang's (1970: 142) observation that, in contrast to written English, zero showed no signs of decline in the spoken medium at the end of EModE is confirmed by the findings from the OBC. On the other hand, van Gelderen's (2006: 217–218) claim that in LModE the particle *that* becomes more and more frequent at the expense of the pronouns is not borne out.

The retreat of *that* in the 19th century was also mentioned by Johansson (2006: 136), but her results differ somewhat from the picture in the OBC data: Johansson (2006: 138) found 67% pronouns and 33% *that* in 19th century trials.[6] This diverges considerably from the OBC data, where in the 1800–1899 period PRN is used 45% and *that* 55% of the times (473 tokens, disregarding zero, which is not considered in Johansson's study). One reason for this difference is that Johansson's dataset of 578 RCs in trials includes 116 non-restrictive RCs, which block *that*. If we add a proportionate figure of pronouns to the 19th century OBC data to make them comparable to Johansson's data, the PRN : *that* ratio (56% : 44%) becomes more similar, but by no means very close to, Johansson's results.

Figure 2 shows that there is a decided difference in the distribution and change of relativizers in RCs where the relativizer is in subject position (SBJ) on the one hand and RCs where it is in non-subject position (n-SBJ) on the other:

---

[6] With 11%, *that* is even more marginal in Johansson's (2006: 137) aggregate of science texts, trials and letters.
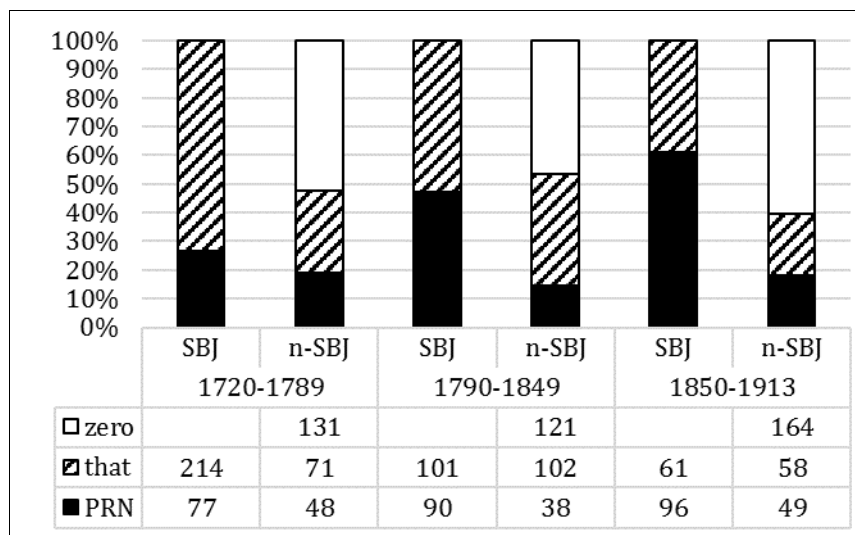
	SBJ	n-SBJ	SBJ	n-SBJ	SBJ	n-SBJ
	1720-1789		1790-1849		1850-1913	
□ zero		131		121		164
▨ that	214	71	101	102	61	58
■ PRN	77	48	90	38	96	49

Figure 2. Relativizers by ROLE of relativizer in the RC

SBJ: P1–P2 p<0.001***, P1–P3 p<0.001***, P2–P3 p=0.009**
n-SBJ, PRN-rest: P1–P2 p=0.161ns, P1–P3 p=0.744ns, P2–P3 p=0.272ns
n-SBJ, *that*-rest: P1–P2 p=0.011*, P1–P3 p=0.064(*), P2–P3 p<0.001***
n-SBJ, zero-rest: P1–P2 p=0.172ns, P1–P3 p=0.062(*), P2–P3 p=0.001***

As can be seen, PRN advances only in subject relativization, and considerably so, from 26.5% in P1 to 61.1% in P3. By contrast, the ratio of relativizers in non-subject positions changes much less drastically from P1 to P3, with the proportion of PRN remaining stable at about 18–19% and zero encroaching on the territory of *that* by a moderate 8.1 percentage points.[7]

---

[7] The advance of PRN in subject relativization is even stronger when only human antecedents are considered: in this case, *who* in subject position increases from 24.3% in P1 to 76.1% in P3 (p<0.001) while there is no significant change in non-subject positions (P1 25.8%, P3 38.1%, p=0.346ns). On the other hand, there is no significant change regarding *which*.

As mentioned in Section 2, Romaine (1982: 62) and Dekeyser (1984: 76–77) found that in EModE pronouns first occurred in, and then remained more frequent in, *non–subject* relativization, and we would have expected this development to continue in LModE. Yet, for the 18th and 19th centuries Figure 2 shows a significantly higher and growing percentage of PRN in *subject* position. It is unlikely that there was a complete reversal of the pattern by which the pronouns made their inroads into the system, especially since the prescriptive pressure in favour of pronouns continued in LModE. The explanation must therefore lie in the different composition of the corpora. The OBC is a mono-genre corpus of real utterances reproduced in writing (roughly "written-as-spoken"), whereas Dekeyser's corpus included several text types: prose 40%, drama 40%, poetry 10%, letters 10% (Dekeyser 1984: 62). Of these, only letters can be said to be similar to trial proceedings in that they reproduce real language events (utterances in the case of trials; mental formulation in the case of letters). Drama is also speech-related (and possibly poetry, but the language here is much more artificial). However, the directional relationship between writing and speech is reversed as drama generates rather than reproduces real-life utterances ("written-to-be-spoken"). The restrictions imposed by rhyme and metre add another level of artificiality to the language of EModE drama. In comparison to the OBC, therefore, Dekeyser's corpus represents more formal and complex written language, where the prescriptive influence advocating the use of pronouns would have been much more felt than in spoken language (see Section 2).
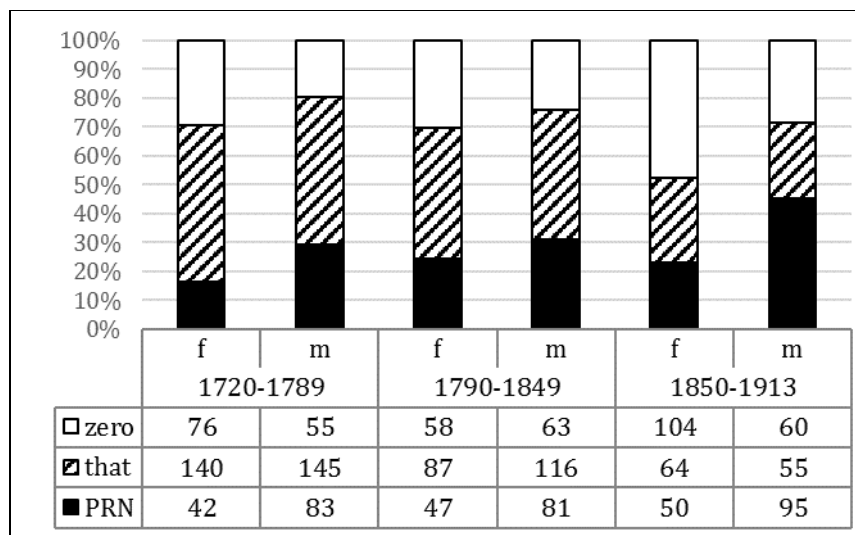
Figure 3. Relativizers by speaker GENDER

F-m PRN-rest: overall p<0.001***, P1 p<0.001***, P2=0.119ns, P3 p<0.001***
F-m *that*-rest: overall p=0.544ns, P1 p=0.481ns, P2 p=0.882ns, P3 p=0.465ns
F-m zero-rest: overall p<0.001***, P1 p=0.007**, P2 p=0.156ns, P3 p=0.001***

Figure 3 shows the use of relativizers by speaker gender. Within each of the three subperiods, the gender differences in the use of *that* (as opposed to the other relativizers) are not significant. At first glance, this conflicts with Johansson's (2006: 140, 173) finding that 19th-century women letter writers used a higher rate of *that* (17.5%, averaged over the 19th century) than their male counterparts (12.5%). Recall, however, that Johansson does not consider zero relativization. If we ignore zero in the OBC data, the 19th-century rates of *that* also show that women are ahead of men: women 61.3%, men 50.7%. Note, however, that the percentages of *that* in the OBC are much higher than Johansson's, which is probably due to a combination of factors: Johansson's data include non-restrictive RCs, where *that* is blocked, which lowers its overall percentage. Add to that the genre difference between trial proceedings (speech-related) and letters (written) as well as the fact that her letter writers were highly literate (Johansson 2006: 139) and would thus have been more susceptible to the prescriptive pressure to use pronouns.

For the entire 1720–1913 period ("overall" in the significance report accompanying Figure 3), the gender differences in the use of PRN and

zero are extremely significant. When broken down into subperiods, these differences remain significant in P1 and P3, and it is mainly females who are responsible for the rise of the zero relativizer (from P1 29.5% to P3 47.7%), while males promote PRN (P1 29.3%, P3 45.2%).
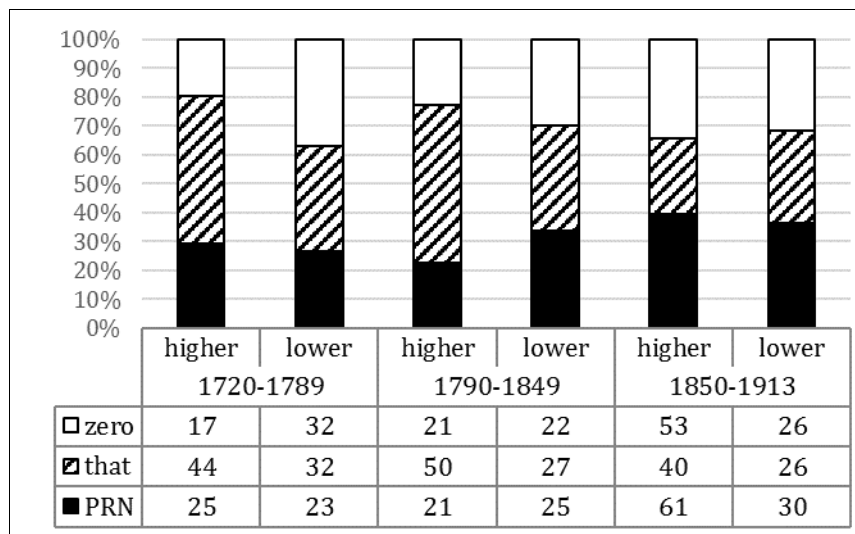

	higher	lower	higher	lower	higher	lower
		1720-1789		1790-1849		1850-1913
☐ zero	17	32	21	22	53	26
▨ that	44	32	50	27	40	26
■ PRN	25	23	21	25	61	30

Figure 4. Relativizers by SOCIAL CLASS of speaker

Higher-lower PRN-rest: overall p=0.975ns, P1 p=0.699ns, P2=0.117ns, P3 p=0.649ns
Higher-lower *that*-rest: overall p=0.189ns, P1 p=0.057(*), P2 p=0.022*, P3 p=0.350ns
Higher-lower zero-rest: overall p=0.153ns, P1 p=0.013*, P2 p=0.313ns, P3 p=0.675ns

The chi-square tests for Figure 4 show that there are no class differences in the selection of relativizers for the entire 1720–1913 period ("overall"). However, there are (marginally) significant differences with regard to *that* and zero in individual subperiods: the higher social classes preferred *that* by an average of 16.2 percentage points in P1 and P2, and in P1 the lower classes favoured zero by 17 percentage points. However, these differences vanished by P3. Given that the non-pronominal relativizers were stigmatized and felt to be more colloquial (see Section 2) it is perhaps not surprising that zero was more common in the lower social classes, but the wider use of *that* in the higher strata is certainly noteworthy. In this connection it is also interesting that there are no class differences regarding the use of PRN—we would have expected them to

be more frequent in the higher social classes, if indeed this was a change from above.

The genders are rather unevenly distributed across the social classes in the sample analyzed here (as in the OBC in general), one reason being that judges and lawyers, who belonged to the higher social classes and had a prominent place in trials,[8] were always male. Table 4 demonstrates that of the females considered in this study, only about one-third (35.0%) belonged to the higher social classes, as compared to two-thirds (66.7%) of the males:

Table 4. Restrictive RCs by GENDER and SOCIAL CLASS, 1720–1913

	GENDER		
SOCIAL CLASS	Female	Male	Sum
higher	57 35.0%	275 66.7%	332
lower	106 65.0%	137 33.3%	243
(unknown)	(505)	(341)	(846)
Sum	668	753	1,421

It is therefore imaginable that what appears to be a gender difference is in fact a masked class difference, or vice versa. To check whether there is an interaction between GENDER and SOCIAL CLASS, the following figures cross-tabulate the use of PRN, *that* and zero. Since the number of tokens for which both the GENDER and SOCIAL CLASS of the speaker is known (575) is too small to obtain significant results for individual subperiods, the cross-tabulation will be performed for the entire 1720–1913 period:

---

[8] In the OBC, 17.6% of the words uttered by higher class speakers are from judges and lawyers.

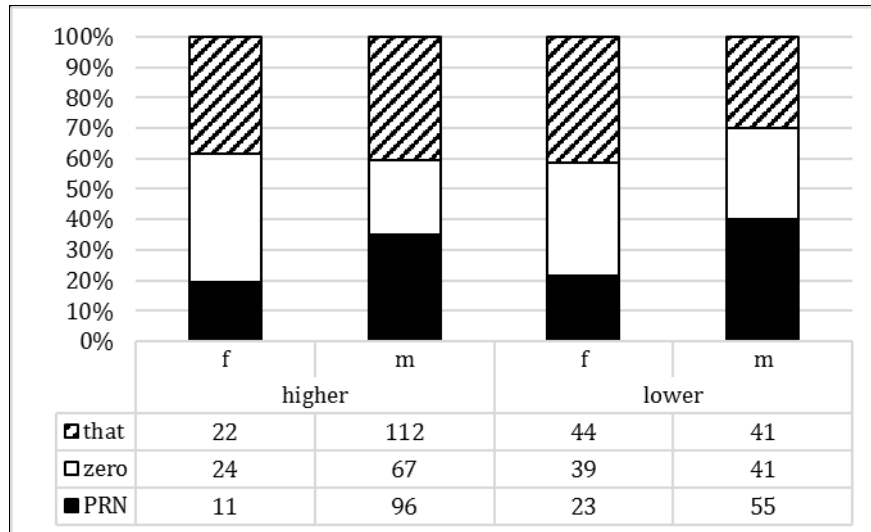|	f	m	f	m
	higher		lower	
◪ that	22	112	44	41
▫ zero	24	67	39	41
■ PRN	11	96	23	55

Figure 5. Relativizers by SOCIAL CLASS and GENDER, 1720–1913[9]

PRN vs. other relativizers
Higher f-m p=0.022*, lower f-m p=0.002**
Females h-l p=0.719ns, males h-l p=0.299ns

That vs. other relativizers
Higher f-m p=0.765ns, lower f-m p=0.060(*)
Females h-l p=0.717ns, males h-l p=0.033*

Regarding PRN, the cross-tabulation in Figure 5 confirms that we are dealing with GENDER rather than class differences here: the gender differences are significant within the higher and lower social classes, but there are no significant differences between females or males across the classes. Splitting these figures up by ROLE does not much alter this result, except that in subject relativization a marginally significant (p=0.084) class difference arises between higher class men (48.5% PRN) and lower class men (61.5% PRN).

Regarding the use of *that*, Figures 3 and 4 suggest that there are class differences but no gender differences. Figure 5 refines this picture: there are indeed no significant differences between the genders in the higher social classes, and no class differences for females. However, the lower class males' use of *that* is 11.6 percentage points lower than that of the lower class females, and 10.8 percentage points lower than that of the

---

[9] Note that for reasons of readabilty, *that* has been moved to the top of the bars in this figure.

higher class males. Note that this difference is only apparent in subject relativization, where higher class males have a higher percentage of *that* (51.5%) than lower class males (38.5%; p=0.084).[10]

The low rate of *that* in lower class males is a direct consequence of this group's higher use of PRN, but it is still surprising. As *that* was felt to be colloquial in LModE (see Section 2) we would have expected it to be more, rather than less, frequent in the speech of lower class males.
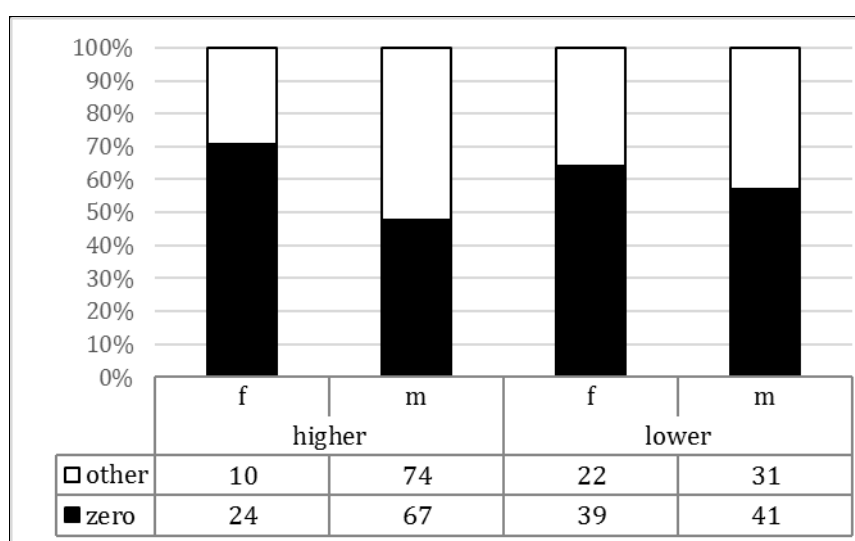

	higher		lower	
	f	m	f	m
□ other	10	74	22	31
■ zero	24	67	39	41

Figure 6. Zero vs. other relativizers by SOCIAL CLASS and GENDER, 1720–1913 (non-subject relativization)

Higher f-m p=0.016*, lower f-m p=0.412ns
Females h-l p=0.511ns, males h-l p=0.193ns

Figure 6 shows the results of the SOCIAL CLASS by GENDER cross-tabulation for the zero relativizer. As zero does not occur in subject position in the data analyzed here, the figure is based on non-subject relativization only. The only significant difference here is in the higher classes, where females prefer zero by 23.1 percentage points over males. This result is again surprising: Labov (1990: 213) postulates that in change from above, "women lead in […] the elimination of stigmatized

---

[10] None of the differences in non-subject RCs are significant.

forms". If zero was indeed as stigmatized as previous studies suggest (see Section 2), the expectation would have been that women used zero less often than men.

5.2 The semantic associations of the relativizers

The following figures focus on the change in the associations of the relativizers with the ANIMACY of the head, operationalized in this article as human vs. non-human.

There was no major change in the zero relativizer, which predominantly co-occurred with non-human heads in all three subperiods (90.6% on average).[11] Similarly, *who*, *whom* and *whose* exclusively relativize human heads throughout the two centuries analyzed here, while *which* is the non-human relativizer. Regarding the latter, there are only three tokens in P1 (4.0%) where *which* co-occurs with a human head, as in

(14)   a little child [which he had lost] (OBC-t17860719-1)

But apart from these exceptions, the OBC data confirm that today's division of work between *who* and *which* was already in place at the beginning of the 18th century.

The one relativizer that shows a statistically significant development with regard to the ANIMACY of the head is the particle *that*:

---

[11] There do not seem to have been major changes in the 20th century either: Quirk's (1957) investigation of spoken British English showed that 92.6% of the zero relativizers occurred with non-human heads. For the end of the 20th century, Tottie (1997) has a similar figure (93.4%) for the spoken part of the *British National Corpus*. The main reason for this stability is that zero is only allowed in non-subject relativization and that non-subjects are predominantly non-humans (see Table 5).

Figure 7. *That* by ANIMACY of the head

Human vs. non-human: P1–P2 p<0.001***, P1–P3 p<0.001***, P2–P3 p=0.378ns

In P1, *that* showed a slight preference to co-occur with human heads (53.2%), but it had become clearly associated with non-human heads by the first half of the 19th century (P2: 73.9%; there is no statistically significant development after that). Note also that this change takes place while the overall frequency of *that* decreases both in absolute numbers and relative to the other relativizers, see Table 3 and Figure 1). This finding is in accordance with Dekeyser's (1984: 79–80) observation that the particle developed into a non-human relativizer from EModE to PDE. ROLE has a strong effect on the semantic association of *that*:
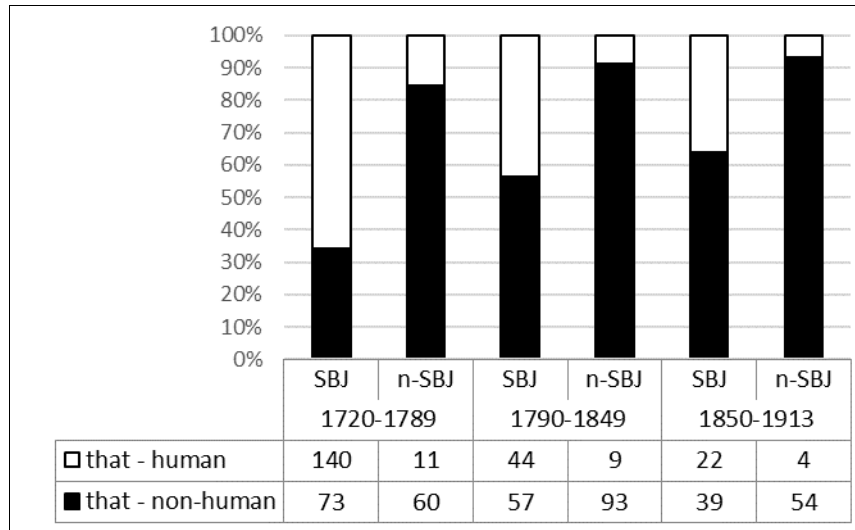
	SBJ	n-SBJ	SBJ	n-SBJ	SBJ	n-SBJ
	1720-1789		1790-1849		1850-1913	
□ that - human	140	11	44	9	22	4
■ that - non-human	73	60	57	93	39	54

Figure 8. *That* by ANIMACY of the head and ROLE of the relativizer

Subject relativization: P1–P2 p<0.001***, P1–P3 p<0.001***, P2–P3 p=0.347ns
Non-subject relativization: P1–P2 p=0.177ns, P1–P3 p=0.130ns, P2–P3 p=0.668ns

In subject relativization, *that* occurred predominantly with human heads in P1 (65.7%) but by P3 had developed into a primarily non-human relativizer (63.9%). Overall, the rate of the co-occurrence of *that* with non-human heads is 44.5 percentage points lower in subject relativization. This is because relative *that* in non-subject positions has from the early 18th century been overwhelmingly associated with non-human heads and there is no significant change over time.

The reason for the very different picture regarding the semantic associations of *that* in subject and non-subject positions is that human heads tend to be followed by subject RCs (82.3% in the present dataset) while non-human heads are usually accompanied by non-subject relativization (73.4%; see Table 5).

In sum, over the two centuries investigated here, *that* develops more and more into a relativizer for non-human heads, making room for the human relativizer *who* (cf. also Ball 1996: 250). This development takes place in subject relativization only, non-subject *that* having already since P1 been strongly associated with non-human heads.
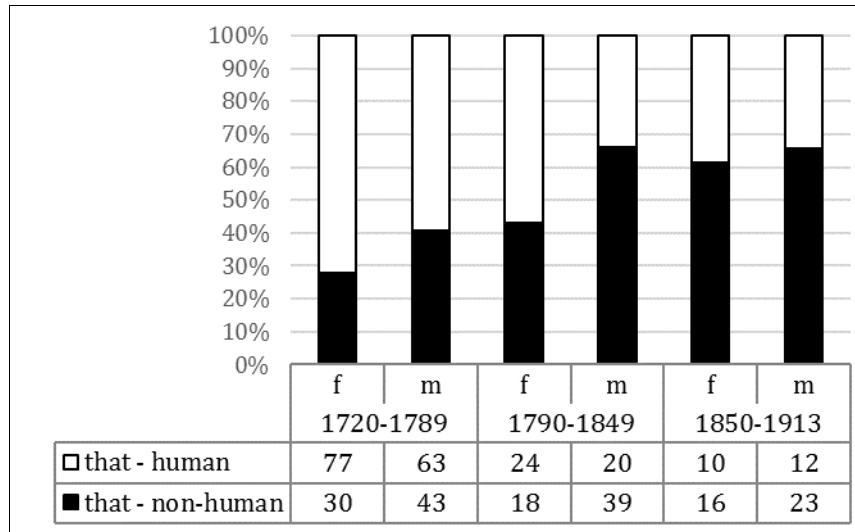
Figure 9. *That* by ANIMACY of the head and GENDER, subject relativization

F-m: overall p=0.002***; P1 p=0.054(*); P2 p=0.020*; P3 p=0.737ns

The gender differences in the semantic associations of *that* in subject RCs are shown in Figure 9. It emerges that overall, men were ahead of women in turning *that* into a non-human relativizer in subject position (there is no overall gender difference for non-subject relativization). The gender differences are significant in P1 and P2, but women caught up with men in the second half of the 19th century.[12]

From the perspective of relativizer selection motivated by the ANIMACY of the head, human antecedents show a distinct development:

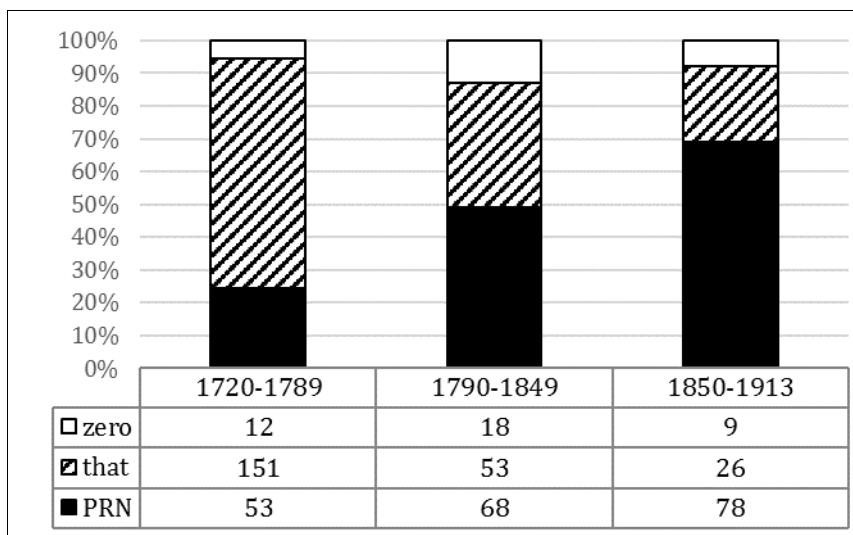[12] There are no significant class differences for the overall 1720–1913 period.

Figure 10. Human heads, relativizer choice

PRN vs rest: P1–P2 p<0.001***, P1–P3 p<0.001***, P2–P3 p=0.001***
That vs. rest: P1–P2 p<0.001***, P1–P3 p<0.001***, P2–P3 p=0.010**
Zero vs. rest: P1–P2 p=0.014*, P1–P3 p=0.396ns, P2–P3 p=0.203ns

During the 18th and the 19th centuries, human heads were increasingly relativized by PRN, rising steeply from 24.5% in P1 to 69.0% in P3, at the expense of *that*. The zero relativizer remained marginal, at 8.3% on average. The reason for the peripheral status of zero is that Figure 10 considers only RCs with human heads. Humans are prototypical agents, and agents typically occur in subject position, as can be seen by the ANIMACY-ROLE cross-tabulation:

Table 5. Human and non-human heads by ROLE of the relativizer

	Subject	Non-subject
Human	385 (82.3%)	83 (17.7%)
Non-human	253 (26.6%)	699 (73.4%)

In the sample analyzed here, the vast majority (82.3%) of relativizers with human antecedents are in subject position, while this is the case for only just above a quarter (26.6%) for non-human antecedents. Since zero

does not occur in subject relativization in the RCs extracted from the OBC (see Section 4), this explains its low percentage in Figure 10.

The advance of PRN was particularly strong with subject RCs modifying human heads, as Figure 11 demonstrates:

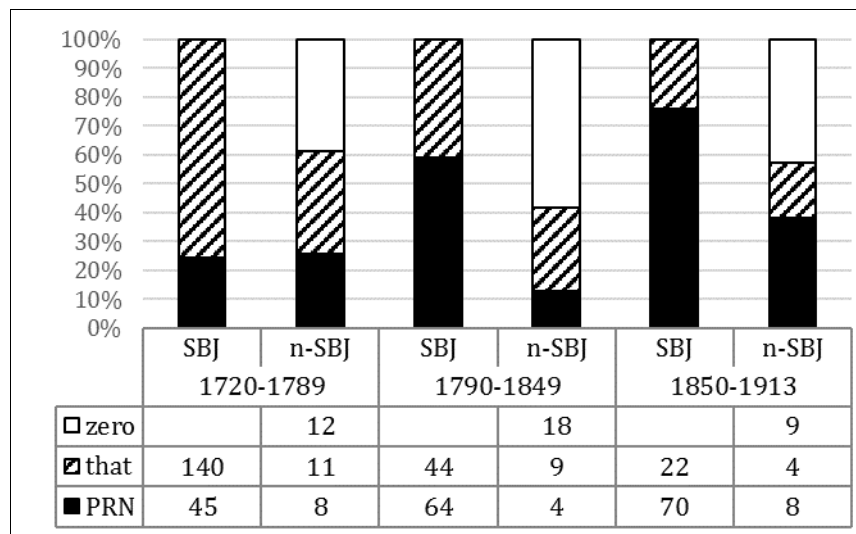|	SBJ	n-SBJ	SBJ	n-SBJ	SBJ	n-SBJ
	1720-1789		1790-1849		1850-1913	
☐ zero		12		18		9
▨ that	140	11	44	9	22	4
■ PRN	45	8	64	4	70	8

Figure 11. Human heads, relativizer choice by ROLE of the relativizer

Subject relativization
P1–P2 p <0.001***, P1–P3 p<0.001***, P2–P3 p=0.012*

Non-subject relativization
PRN-rest: P1–P2 p=0.199ns, P1–P3 p=0.346ns, P2–P3 p<0.034*
That-rest: P1–P2 p=0.587ns, P1–P3 p=0.199ns, P2–P3 p=0.414ns
Zero-rest: P1–P2 p=0.127ns, P1–P3 p=0.765ns, P2–P3 p=0.281ns

In her analysis of subject relativization, Ball (1996: 246–247) found that in the second half of the 17th century *that* started to be associated with non-human heads and *who* with human heads. This specialization was more advanced in written British English (human: 42% *who*, 57% *that*; non-human: 26% *which*, 74% *that*) than in spoken British English (human: 13% *who*, 84% *that*; non-human: 9% *which*, 89% *that*).[13] The

---

[13] Note that Ball's (1996) data include an unspecified number of existential sentences (excluded in my study), 15 of which had a zero relativizer in subject

OBC data show that this development continues in spoken LModE: in subject RCs with human heads, PRN shows a sharp increase, from 24.3% in P1 to 76.1% in P3, while *that* recedes proportionately (and gets more and more associated with non-human heads, see Figure 8). Ball (1996: 249) does not have spoken data for the LModE period, but her figures for written English in the 18th (92%) and 19th centuries (97%) show that the adoption of relative pronouns in subject relativization was almost completed. The OBC data shows that at the end of EModE, spoken language lagged behind considerably in this change (compare similar results for 17th-century British and American trials in Ball 1996: 247–248), only catching up at around 1900.

In contrast to subject relativization, relativizer choice in non-subject RCs with human heads remained relatively stable. There was no significant change in the frequency of pronominal relativizers, *that* or zero from P1 to P3.

There are no substantial class differences in the selection of relativizers with regard to the ANIMACY of the antecedent, but GENDER proves to be significant:

---

position, so the figures given here do not add up to 100%. The percentages for spoken British English indicated above are averaged across Ball's British *State Trials* categories ST 1 and ST 2.
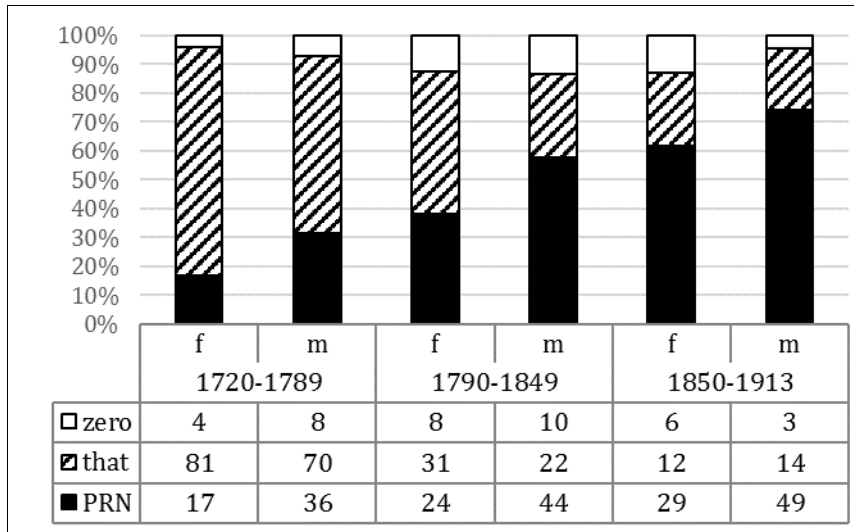
	f	m	f	m	f	m
	1720-1789		1790-1849		1850-1913	
□ zero	4	8	8	10	6	3
▨ that	81	70	31	22	12	14
■ PRN	17	36	24	44	29	49

Figure 12. Human heads, relativizer choice by GENDER

F-m PRN-rest: overall p<0.001***, P1 p=0.011*, P2 p=0.020*, P3 p=0.155ns
F-m *that*-rest: overall p<0.001***, P1 p=0.004**, P2 p=0.014*, P3 p=0.621ns
F-m zero-rest: overall p=0.909ns, P1 p=0.321ns, P2 p=0.938ns, P3 p=0.112ns

In relativization with human antecedents, males show a significantly higher rate than women in the use of *who* in the first two subperiods (14.9 percentage points higher in P1 and 19.8 percentage points in P2), and a roughly proportionately lower rate of *that*. These differences disappear in the second half of the 19th century.

Again, this is particularly apparent in subject relativization, where males lead in the promotion of PRN by 21.1 percentage points on average:
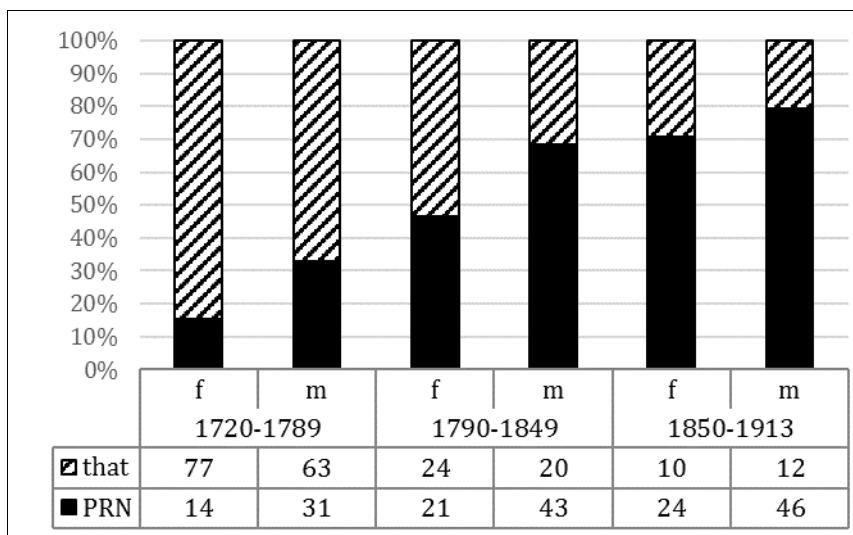
	1720-1789		1790-1849		1850-1913	
	f	m	f	m	f	m
▨ that	77	63	24	20	10	12
■ PRN	14	31	21	43	24	46

Figure 13. Human heads, relativizer choice by GENDER, subject relativization
F-m: Overall p<0.001***, P1 p=0.005**, P2 p=0.024*, P3 p=0.344ns

As before, these gender differences disappeared after 1850. In non-subject relativization, (marginally) significant gender differences are only found with regard to zero (f 46.2%, m 52.5%, p=0.1).

In contrast to Ball (1996: 247–248), who found that lawyers and aristocrats led in the adoption of the pronouns in the late 17th century, no significant class differences for subject relativization with human heads were found in the OBC data.

Figure 10 showed a marked rise of PRN with human heads. The development of relativizer selection was different and less drastic for non-human heads:
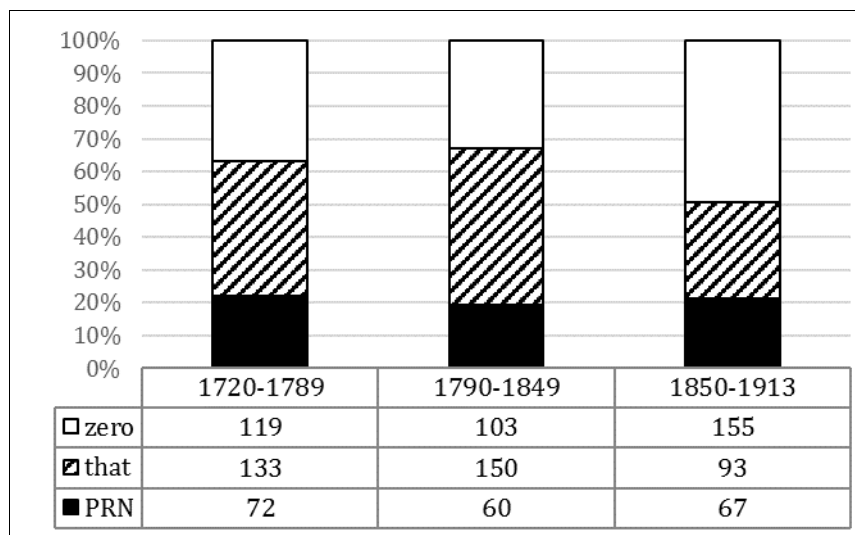
	1720-1789	1790-1849	1850-1913
☐ zero	119	103	155
◩ that	133	150	93
■ PRN	72	60	67

Figure 14. Non-human heads, relativizer choice

PRN vs rest: P1–P2 p=0.342ns, P1–P3 p=0.771ns, P2–P3 p=0.512ns
That vs. rest: P1–P2 p=0.081(*), P1–P3 p=0.002**, P2–P3 p<0.001***
Zero vs. rest: P1–P2 p=0.312ns, P1–P3 p<0.001***, P2–P3 p<0.001***

Relativization of non-human heads with pronouns (i.e. *which*) remained stable at round about 20.9%, but there was a moderate increase of zero from 36.7% in P1 to 49.2% in P3, with *that* decreasing proportionally. The percentage of zero is higher than in Figure 10 because non-human heads usually co-occur with non-subject RCs (see Table 5), where zero is permitted.

Factoring in ROLE shows no significant change in non-human subject relativization, with the pronoun *which* chosen in 33.2% of the cases on average.[14] The picture is more nuanced in non-subject relativization, where significant differences emerge for *that* and zero:

---

[14] For the 18th and 19th centuries, Ball's (1996: 249) figures indicate a higher percentage of *which* in written English, 74% and 75%, respectively. This again is evidence that the adoption of pronouns progressed more slowly in spoken English.
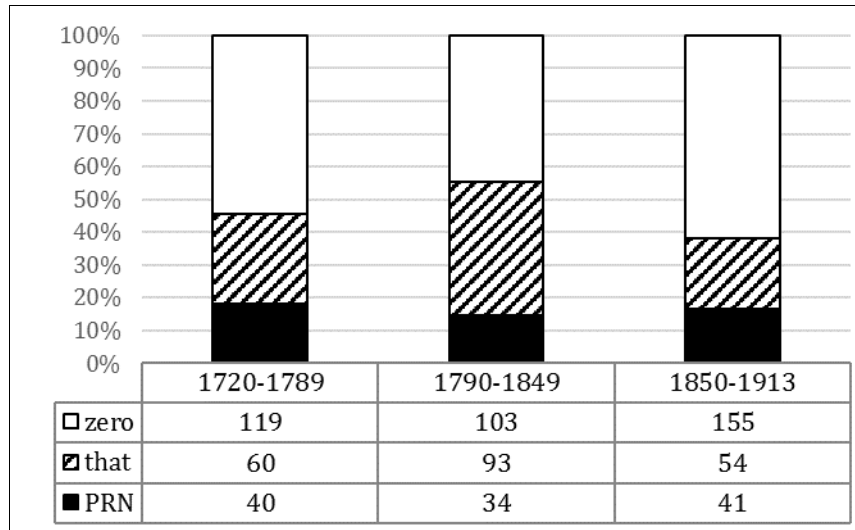
Figure 15. Non-human heads, non-subject relativization

PRN-rest: P1–P2 p=0.320ns, P1–P3 p=0.594ns, P2–P3 p=0.626ns
That-rest: P1–P2 p=0.004**, P1–P3 p=0.144ns, P2–P3 p<0.001***
Zero-rest: P1–P2 p=0.043*, P1–P3 p=0.093(*), P2–P3 p<0.001***

Disregarding the somewhat irregular development in P2, the general overall trend was that zero increased slightly from P1 (54.3%) to P3 (62.0%) at the expense of *that* (P1 27.4% > P3 21.6%), while the frequency of PRN remained stable.

Adding SOCIAL CLASS as a factor to Figure 14 yields no significant differences, but adding GENDER does:
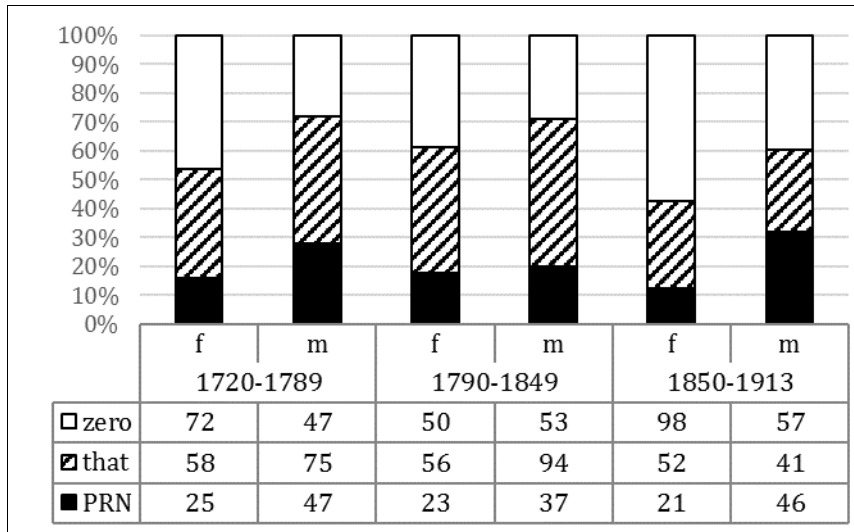
	f	m	f	m	f	m
		1720-1789		1790-1849		1850-1913
☐ zero	72	47	50	53	98	57
◨ that	58	75	56	94	52	41
■ PRN	25	47	23	37	21	46

Figure 16. Non-human heads, relativizer choice by GENDER

F-m PRN-rest: overall p<0.001***, P1 p=0.012*, P2 p=0.614ns, P3 p<0.001***
F-m *that*-rest: overall p=0.069ns, P1 p=0.203ns, P2 p=0.181ns, P3 p=0.707ns
F-m zero-rest: overall p<0.001***, P1 p<0.001***, P2 p=0.065(*), P3 p=0.002**

Calculated over the entire 1720–1913 period, females preferred zero in the relativization of non-human heads by 16.8 percentage points. As with human heads, PRN is used more by males than by females, on average by 11.0 percentage points. There is no significant gender difference with regard to *that*.

The following two figures split Figure 16 up into subject and non-subject relativization:
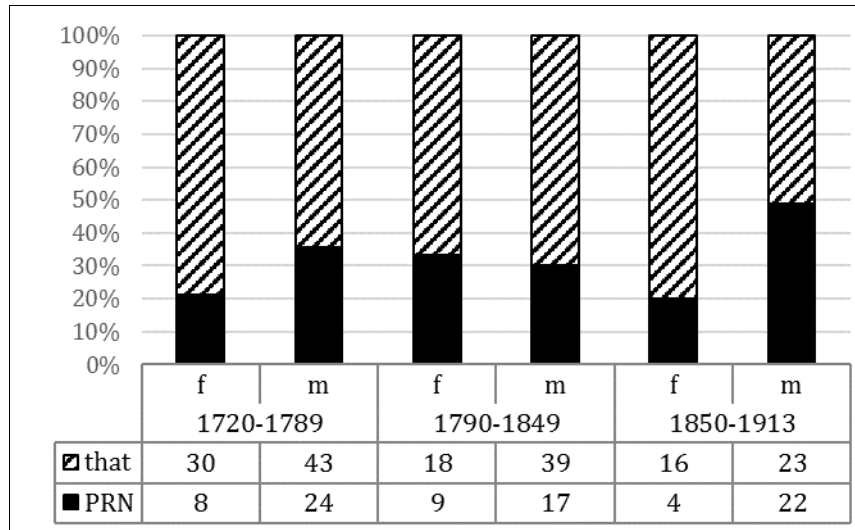
	f	m	f	m	f	m
	1720-1789		1790-1849		1850-1913	
▨ that	30	43	18	39	16	23
■ PRN	8	24	9	17	4	22

Figure 17. Non-human heads, relativizer choice by GENDER, subject relativization

F-m: overall p=0.041*, P1 p=0.114ns, P2 p=0.784ns, P3 p=0.028*

Taking the entire 1720–1913 period as a basis, men lead in the promotion of PRN by 18.4 percentage points. Within the individual periods, the difference is only significant in P3, where men are 28.9 percentage points ahead.

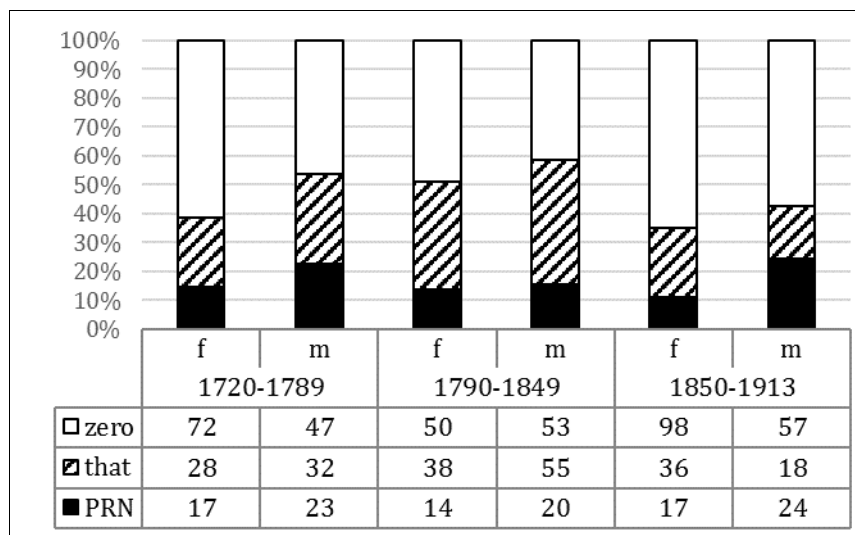Figure 18 displays the figures for non-subject relativization:

Figure 18. Non-human heads, relativizer choice by GENDER, non-subject relativization

PRN-rest: overall p=0.009**, P1 p=0.126ns, P2 p=0.686ns, P3 p=0.007**
That-rest: overall p=0.209ns, P1 p=0.218ns, P2 p=0.381ns, P3 p=0.288ns
Zero-rest: overall p=0.002**, P1 p=0.022*, P2 p=0.249ns, P3 p=0.243ns

There are significant, if moderate, overall gender differences in the use of PRN (males slightly ahead of females by 7.4 percentage points in the selection of *which*) and zero (females ahead of males by 11.8 percentage points). However, when zooming in to the three subperiods, the gender differences are significant only in P3 for PRN and in P1 for zero.

6. Summary, discussion and outlook

By way of conclusion, the main findings of the previous section will be summarized and discussed.

LModE prescriptivism supported the use of pronouns, which were felt to be most explicit with regard to animacy and case. Figure 1 accordingly showed a moderate rise of PRN and a steep drop of *that* over the 18th and 19th centuries. The GENDER-SOCIAL CLASS cross-tabulation showed that lower class males used *that* about 10 percentage points less than lower class females or higher class speakers (Figure 5).

The increase of zero in the period investigated in this article is unexpected in view of its maximally unspecified nature. The explanation for this may be that the effects of prescriptive pressure came to be felt much later in spoken English, at least as far as the zero relativizer was concerned. Females led in the adoption of zero but the gender difference is only significant in the higher social classes (Figure 6).

It is interesting that the rise of PRN was restricted to subject relativization (Figure 2), i.e. the most accessible position of the Noun Phrase Accessibility Hierarchy. If it is true that the pronouns made their inroads into the English relativizer system through the less accessible syntactic positions (Romaine 1982: 62; see Section 2), then we would have expected the pronouns to be more frequent in LModE non-subject positions. Future research will have to find an explanation for the unanticipated advance of pronouns in 18th and 19th century in subject relativization.

Another unexpected result with regard to PRN was that male speakers led in their introduction (Figure 3; there was no interaction with SOCIAL CLASS, see Figure 5). If we accept that the addition of pronouns to the relativizer system was a change from above (see Section 2), then the results in Figures 3 and 5 violate Labov's (1990: 213–214) Principle Ia, which states that "in change from above, women favor the incoming prestige form more than men" and according to which women should have used more pronouns than men. However, Labov makes the important qualification "that for women to use standard norms that differ from everyday speech, they must have access to those norms" (Labov 1990: 213). It is possible that unequal access to the prescriptively prescribed pronouns is the explanation why in the OBC it is men rather than women who lead in the adoption of the prestigious pronominal relativizers. The EModE and LModE arguments in favour of relative pronouns were based on rules modelled on Latin grammar. Pronominal relativizers were consequently most frequently used in complex, formal registers (see Section 2), to which the highest social classes had the greatest exposure. As mentioned in Section 4, the "higher" social class in this article is an aggregate of the five non-manual occupation groups in the HISCLASS scheme:

1. Higher managers
2. Higher professionals
3. Lower managers
4. Lower professionals, clerical and sales personnel
5. Lower clerical and sales personnel

We can assume that the highest two (higher managers and professionals) would have been most regularly exposed to classical languages as well as to stylistically and grammatically elaborate texts. This is especially true for the judges and lawyers, who in the OBC make up 63.6% of the male speakers in HISCLASS 1 and 2. Table 4 showed that only about one-third of all women in this study belonged to the higher social class, as compared to two-thirds of the men. Within the higher class, the proportion of women is even more unevenly distributed: a mere 5.7% of the higher class females belong to HISCLASS 1 and 2, as opposed to 31.4% of the higher class males. That is, only a minute fraction of women represented in the OBC would have had direct and frequent access to texts in which the prestige norm advocating relative pronouns would have been felt the most.

However, a closer inspection of the data reveals that this line of argument does not explain the results: among male speakers it is actually the higher managers and professionals who show a particularly low percentage of PRN (13/69=18.8%), in contrast to male members of HISCLASS 3–5, whose percentage (83/206=40.3%) is very much like that of lower class males. The situation is even more extreme among the judges and lawyers, who use only one PRN in 27 RCs. That is, the group with the best access to the prescriptive norm actually used the lowest rate of PRN.

The question thus remains of how to account for the unexpected patterning of the relativizers in the cross-tabulation of GENDER and SOCIAL CLASS. Labov (1990: 210, 220) observes that "[e]vidence for Principle I is uniform and voluminous" and that "[t]here are no significant exceptions for I". The distribution of the PRN, *that* and zero in the OBC data possibly constitutes one of the rare exceptions where women are not ahead in a change from above. Alternatively, there may be an explanation which is a yet not apparent, e.g. that the social value of pronouns was different in spoken and written language (note that most

studies commenting on the prestige of relativizers in EModE and LModE look at written language or text types that are closer to this mode).

With respect to the semantics of the relativizers, *that* in non-subject position was almost exclusively used with non-human heads right from the beginning of the 18th century. In the course of LModE, subject *that* developed in the same direction, from more or less animacy-neutral in P1 to predominantly non-human in P3 (Figure 8). This change was led by men (Figure 9) and is closely linked to male speakers' promotion of *who* as a relativizer for human heads, a change that is limited to subject RCs (Figures 10–13).

Regarding relativizer choice for non-human heads, PRN remained stable and marginal at ca. 21% throughout the period analyzed here, while there was a slight increase of zero at the expense of *that* (Figure 14). This time, the development took place in non-subject relativization (Figure 15). Once again, females preferred zero and males preferred PRN (Figures 16–18).

The present study has shown that—although the PDE relativizer inventory was already in place at the beginning of LModE—the 18th and 19th centuries witnessed an extensive reorganisation of the distribution of relativizers, correlating with an interplay of syntactic, semantic and also social variables. Some of the findings question earlier assumptions about the development of English RCs and about the social mechanisms of language change in general. These will have to be followed up in future studies.

References

Aarts, Bas, María José López-Couso and Belén Méndez-Naya. 2012. "Late Modern English: Syntax." *English Historical Linguistics. An International Handbook*. Vol. I. Eds. Alexander Bergs and Laurel J. Brinton. Berlin: De Gruyter Mouton. 869–887.

Adamson, Sylvia. 2007. "Prescribed Reading: Pronouns and Gender in the Eighteenth Century." *Historical Sociolinguistics and Sociohistorical Linguistics*. http://www.let.leidenuniv.nl/hsl_shl/Adamson.htm.

Ball, Catherine. 1996. "A Diachronic Study of Relative Markers in Spoken and Written English." *Language Variation and Change* 8: 227–258.

Collins, Peter. 1991. *Cleft and Pseudo-Cleft Constructions in English*. London: Routledge.

Dekeyser, Xavier. 1984. "Relativicers [sic] in Early Modern English: A Dynamic Quantitative Study." *Historical Syntax*. Ed. Jacek Fisiak. Berlin: Mouton. 61–87.

Görlach, Manfred. 2001. *Eighteenth-Century English*. Heidelberg: Universitätsverlag C. Winter.

Görlach, Manfred. 1999. *English in Nineteenth-Century England*. Cambridge: Cambridge University Press.

Helsinki Corpus = The Helsinki Corpus of English Texts. 1991. Compiled by Matti Rissanen (Project leader), Merja Kytö (Project secretary); Leena Kahlas-Tarkka, Matti Kilpiö (Old English); Saara Nevanlinna, Irma Taavitsainen (Middle English); Terttu Nevalainen, Helena Raumolin-Brunberg (Early Modern English). Department of Modern Languages, University of Helsinki.

Hitchcock, Tim, Robert Shoemaker, Clive Emsley, Sharon Howard, Jamie McLaughlin et al. 2015. *The Old Bailey Proceedings Online, 1674–1913*. www.oldbaileyonline.org, version 7.2, March 2015.

Huber, Magnus. 2012. "Existential Sentences with RCs in 18th and 19th Century English. A Diachronic Study Based on the Old Bailey Corpus". Paper presented at the 33rd Conference of the International Computer Archive for Modern and Medieval English (ICAME), Leuven, Belgium, 30 May–3 June 2012.

Huber, Magnus. 2009. "More News from Old Bailey. The Development of Cleft Constructions in the 18th and 19th Centuries." Paper presented at the 30th Conference of the International Computer Archive for Modern and Medieval English (ICAME), Lancaster, UK, 27–31 May 2009.

Huber, Magnus, Magnus Nissel, Patrick Maiwald and Bianca Widlitzki. 2012. *The Old Bailey Corpus. Spoken English in the 18th and 19th Centuries*. www.uni-giessen.de/oldbaileycorpus.

Johansson, Christine. 2012. "Relativization in Early Modern English. Written Versus Speech-Related Genres." *English Historical Linguistics. An International Handbook*. Vol. I. Eds. Alexander Bergs and Laurel J. Brinton. Berlin: De Gruyter Mouton. 776–790.

Johansson, Christine. 2006. "Relativizers in Nineteenth-Century English." *Nineteenth-Century English. Stability and Change*. Eds.

Merja Kytö, Mats Rydén and Erik Smitterberg. Cambridge: Cambridge University Press, 136–182.

Johnson, Daniel Ezra. 2009. "Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis." *Language and Linguistics Compass* 3: 359–383.

Keenan, Edward L. and Bernard Comrie. 1977. "Noun Phrase Accessibility and Universal Grammar." *Linguistic Inquiry* 8: 63–99.

Labov, William. 1990. "The Intersection of Sex and Social Class in the Course of Linguistic Change." *Language Variation and Change* 2: 205–254.

Nevalainen, Terttu. 2006. *An Introduction to Early Modern English*. Edinburgh: Edinburgh University Press.

Nevalainen, Terttu and Helena Raumolin-Brunberg. 2003. *Historical Sociolinguistics. Language Change in Tudor and Stuart England*. Harlow: Pearson Education Limited.

Nevalainen, Terttu and Helena Raumolin-Brunberg. 2002. "The Rise of the Relative *who* in Early Modern English." *Relativisation on the North Sea Littoral*. Ed. Patricia Poussa. Munich: Lincom Europa. 109–121.

Quirk, Randolph. 1957. "Relative Clauses in Educated Spoken English." *English Studies* 38: 97–109.

Rissanen, Matti. 1984. "The Choice of Relative Pronouns in 17th Century American English." *Historical Syntax*. Ed. Jacek Fisiak. Berlin: Mouton. 417–435.

Rissanen, Matti et al. 1991. See *Helsinki Corpus*, above.

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

Rydén. Mats. 1983. "The Emergence of *who* as Relativizer." *Studia Linguistica* 37: 126–134.

Rydén, Mats. 1966. *Relative Constructions in Early Sixteenth Century English*. Uppsala: Almqvist & Wiksell.

Strang, Barbara. 1970. *A History of English*. London: Methuen & Co.

Tottie, Gunnel. 1997. "Relatively Speaking. Relative Marker Usage in the British National Corpus." *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*. Eds. Terttu Nevalainen and Leena Kahlas-Tarkka. Helsinki: Société Néophilologique. 465–481.

van Gelderen, Elly. 2006. *A History of the English Language*. Amsterdam: Benjamins.

van Leeuwen, Marco H.D. and Ineke Maas. 2011. *HISCLASS: A Historical International Social Class Scheme*. Leuven: Leuven University Press.

Appendix: Binomial regressions
Performed using Rbrul (Johnson 2009)


```
1a. Step up/down, N 1420, excluding class and 1 empty token from
animacy: likelihood of REL

BEST STEP-DOWN MODEL OF RESPONSE relativizer IS WITH PREDICTOR(S):
pos_in_RC (6.1e-11) + period (1.17e-07) + gender (8.86e-07) +
animacy (8.74e-06)
[p-values dropping from full model]

$animacy
    factor logodds tokens PRN/PRN+that+zero centered factor weight
     human   0.328    468             0.425                   0.581
 non-human  -0.328    952             0.209                   0.419

$role
 factor logodds tokens PRN/PRN+that+zero centered factor weight
    SBJ   0.481    638             0.412                   0.618
  n-SBJ  -0.481    782             0.173                   0.382

$gender
 factor logodds tokens PRN/PRN+that+zero centered factor weight
      m   0.315    753             0.344                   0.578
      f  -0.315    667             0.208                   0.422

$period
    factor logodds tokens PRN/PRN+that+zero centered factor weight
 1850-1913   0.434    428             0.339                   0.607
 1790-1849   0.010    452             0.283                   0.503
 1720-1789  -0.445    540             0.231                   0.391

$misc.1
    n df intercept overall proportion centered input prob
 1420  6      -0.9                0.28                 0.289

$misc.2
 deviance      AIC     AICc   Dxy    R2
 1514.134 1526.134 1526.193 0.409 0.161


1b.  Step  up/down,  N  575,  only  tokens  specified  for  class:
likelihood of REL

BEST STEP-DOWN MODEL OF RESPONSE relativizer IS WITH PREDICTOR(S):
pos_in_RC (1.27e-08) + gender (0.000332) + animacy (0.00094) +
period (0.00746)
[p-values dropping from full model]
```

```
$animacy
    factor logodds tokens PRN/PRN+that+zero centered factor weight
     human  0.373    169               0.527                   0.592
 non-human -0.373    406               0.236                   0.408

$role
 factor logodds tokens PRN/PRN+that+zero centered factor weight
    SBJ   0.62     266              0.492                   0.65
  n-SBJ  -0.62     309              0.175                   0.35

$gender
 factor logodds tokens PRN/PRN+that+zero centered factor weight
     m   0.412    412              0.367                   0.602
     f  -0.412    163              0.209                   0.398

$period
    factor logodds tokens PRN/PRN+that+zero centered factor weight
 1850-1913  0.410    236              0.386                   0.601
 1720-1789 -0.151    173              0.277                   0.462
 1790-1849 -0.259    166              0.277                   0.436

$misc.1
  n df intercept overall proportion centered input prob
 575  6  -0.922             0.322                 0.284

$misc.2
 deviance     AIC     AICc   Dxy   R2
  622.179 634.179 634.326 0.489 0.213
```

2a. Step up/down, N 1420, excluding class and 1 empty token from animacy: likelihood of *that*

BEST STEP-DOWN MODEL OF RESPONSE relativizer IS WITH PREDICTOR(S):
pos_in_RC (4.23e-27) + period (1.91e-11) + animacy (0.000697) + gender (0.0152)
[p-values dropping from full model]

```
$animacy
    factor logodds tokens that/that+PRN+zero centered factor weight
 non-human  0.244    952              0.395                   0.561
     human -0.244    468              0.491                   0.439

$role
 factor logodds tokens that/that+PRN+zero centered factor weight
    SBJ   0.727    638              0.588                   0.674
  n-SBJ  -0.727    782              0.295                   0.326


$gender
 factor logodds tokens that/that+PRN+zero centered factor weight
     f   0.142    667              0.435                   0.535
     m  -0.142    753              0.420                   0.465
```

```
$period
    factor logodds tokens that/that+PRN+zero centered factor weight
 1720-1789  0.387    540                0.526                    0.596
 1790-1849  0.192    452                0.449                    0.548
 1850-1913 -0.579    428                0.278                    0.359

$misc.1
    n df intercept overall proportion centered input prob
 1420  6    -0.37              0.427                  0.409

$misc.2
 deviance      AIC     AICc    Dxy    R2
 1750.985 1762.985 1763.044 0.395 0.148
```

2b. Step up/down, N 575, only tokens specified for class: likelihood of *that*

BEST STEP-DOWN MODEL OF RESPONSE relativizer IS WITH PREDICTOR(S): pos_in_RC (1.26e-10) + period (0.000213) + animacy (0.0042) [p-values dropping from full model]

```
$animacy
     factor logodds tokens that/that+PRN+zero centered factor weight
 non-human  0.322    406                0.377                     0.58
     human -0.322    169                0.391                     0.42

$role
 factor logodds tokens that/that+PRN+zero centered factor weight
    SBJ  0.659    266            0.508                    0.659
  n-SBJ -0.659    309            0.272                    0.341

$period
     factor logodds tokens that/that+PRN+zero centered factor weight
 1790-1849  0.333    166                0.464                    0.582
 1720-1789  0.168    173                0.439                    0.542
 1850-1913 -0.501    236                0.280                    0.377

$misc.1
    n df intercept overall proportion centered input prob
 575  5    -0.561             0.381                  0.363

$misc.2
 deviance      AIC    AICc    Dxy    R2
  704.776 714.776 714.881 0.387 0.12
```

3a. Step up/down, N 1420, excluding class, animacy and 1 empty token from animacy: likelihood of zero

BEST STEP-DOWN MODEL OF RESPONSE relativizer IS WITH PREDICTOR(S): animacy (6.75e-37) + gender (2.46e-06) + period (0.0015) [p-values dropping from full model]

```
$animacy
    factor logodds tokens zero/zero+PRN+that centered factor weight
 non-human  0.977    952                0.396                  0.726
     human -0.977    468                0.083                  0.274

$gender
 factor logodds tokens zero/zero+PRN+that centered factor weight
     f  0.295    667               0.357                  0.573
     m -0.295    753               0.236                  0.427

$period
    factor logodds tokens zero/zero+PRN+that centered factor weight
 1850-1913  0.317    428                0.383                  0.579
 1790-1849 -0.143    452                0.268                  0.464
 1720-1789 -0.174    540                0.243                  0.457

$misc.1
    n df intercept overall proportion centered input prob
 1420  5   -1.399               0.293                 0.198

$misc.2
 deviance      AIC     AICc   Dxy    R2
 1509.154 1519.154 1519.197 0.454 0.228
```

3b. Step up/down, N 575, only tokens specified for class: likelihood of zero

```
$animacy
    factor logodds tokens zero/zero+PRN+that centered factor weight
 non-human  0.983    406                0.387                  0.728
     human -0.983    169                0.083                  0.272

$gender
 factor logodds tokens zero/zero+PRN+that centered factor weight
     f  0.312    163               0.387                  0.577
     m -0.312    412               0.262                  0.423

$misc.1
   n df intercept overall proportion centered input prob
 575  3   -1.318              0.297                 0.211

$misc.2
 deviance      AIC    AICc   Dxy    R2
  629.403 635.403 635.445 0.385 0.201
```