Processing and prescriptivism as constraints on language variation and change: Relative clauses in British and Australian English parliamentary debates.

Sofie Labat (Ghent University)

Haidee Kotze (Utrecht University / North-West University)

Benedikt Szmrecsanyi (KU Leuven)

Abstract

We investigate the choice between the relative markers *which* and *that* in 8283 restrictive relative clauses on subject position, with inanimate antecedents, in a written corpus consisting of British and Australian Hansard materials over five sampling years (1901, 1935, 1965, 1995, 2015). Our aim is to determine how processing-related factors and prescriptivism-related factors influence processes of language variation and change across two varieties of English. We analyse how the language-external variables of period and variety (British, Australian) interact with two groups of language-internal variables, namely predictors related to language processing and linguistic predictors associated with prescriptivism. The analysis shows that the relativiser *that* has been on the rise over the past century. The increase is particularly pronounced in the British Hansard, as *which* was comparatively frequent in early twentieth-century British material. As to the relative

importance of predictors, we find that language-external predictors are the most important in conditioning the variation in relative markers, followed by processing-related constraints. Prescriptivism-related variables tend to generally be less important in this type of variation.

Keywords: restrictive relative clause, Australian English, British English, real-time study, language processing, prescriptivism

1. Introduction

The alternation (or, in Labovian parlance, 'variable') we are investigating in this paper is that between *which* and *that* as relativisers introducing subject restrictive relative clauses after inanimate antecedents, as in Example (1).

(1) a. The mail$_{\text{antecedent}}$ [that was sent yesterday]$_{\text{relative clause}}$ has arrived.

b. The mail$_{\text{antecedent}}$ [which was sent yesterday]$_{\text{relative clause}}$ has arrived.

We stress at the outset that the way we define the variable context (which follows the circumscription in e.g. Szmrecsanyi 2019) means that zero relativers (of the type *the book ___ I bought*), reduced relative clauses (of the

type *the book written*), and *wh*-relativisers other than *which* (e.g. *the man who won*) do not come under the remit of this study. In Standard English, zero and reduced relatives only occur in *object* relative clauses, while for example *who* we find only after *animate* antecedents. In short, these are variants that are part of different alternations or variables in Labovian terminology.

Although in Standard English restrictive relative clauses after inanimate antecedents both *that* and *which* are grammatically correct (Biber et al. 2007: 615–616), their distribution differs drastically across registers. This register variation reflects the difference in formality between the two relative markers, as *that* is more frequently used in informal contexts such as conversations and fiction, while *which* is the preferred relativiser in formal, usually written, contexts such as newspaper articles and academic prose (Biber et al. 2007: 610; Sigley 1997). Moreover, Tagliamonte, Smith and Lawrence (2005) explain that the *wh*-relativisers probably entered the English language through formal written registers. According to their data, the *wh*-forms have surprisingly low frequencies in three vernacular varieties of English in Britain, which seems to be a product of "resistance to or general 'distance' from changes taking place in mainstream varieties" (Tagliamonte, Smith & Lawrence 2005: 105).

Variation in relativisation is a well-researched topic in linguistics (see e.g. Collins 2014; D'Arcy & Tagliamonte 2010; Hinrichs, Szmrecsanyi & Bohmann 2015; Leech & Smith 2006, 2009; Leech et al. 2009; Sigley 1997; Tagliamonte, Smith & Lawrence 2005). Most research has focused on American and British

English, with some exceptions such as Collins (2014) who investigated Australian English, or Sigley (1997) who used corpora of New Zealand English to determine the role of formality in the choice of relative marker. Some important outputs from this large body of literature are (a) the general increase in *that* at the cost of *which* in written English, (b) the fact that this shift is spearheaded by American English, while British English seems to be more conservative, and (c) arguments that propose an explanation for the shift from *which* to *that*, such as colloquialisation and prescriptivism.

The current study extends previous work on relativisation by drawing on a unique hybrid spoken-written register, the Hansard, the edited written record of parliamentary debates in Commonwealth countries. We sampled British and Australian Hansard material over five sampling years: 1901, 1935, 1965, 1995 and 2015. One of our aims is to examine how the choice between *that* and *which* is probabilistically conditioned over time and across varieties of English. We view the Hansard as produced by the linguistic processing of two sets of users, overlaid on each other: the spoken debates and speeches of parliamentarians, and the transformation of this spoken material into a written record by reporters and editors (Kruger, Van Rooy & Smith 2019; see also this volume, Chapter 2). In this respect, it is important to underscore that writing is distinct from speech in several ways. It is not subject to processing pressures under time constraints, and can be carefully crafted and edited to adhere to particular normative and prescriptive constraints (see e.g. Harder 2012). As the distribution of restrictive relative markers changes across registers, channels, and according to the degree

of formality involved, our analysis renders some interesting insights on this distribution in the hybrid spoken-written register of the Hansard.

In addition to these language-external factors, we also investigate language-internal predictors and their role in conditioning the alternation between *that* and *which*. One group of language-internal factors is linked to processing effects and constraints. Existing research shows that as the degree of sentence complexity or general processing load increases (which is linked to a similar rise in formality), the relative marker *which* becomes increasingly preferred (see e.g. Sigley 1997). For example, the length of the antecedent and the length of the relative clause are significant predictors in the alternation, and as the length of these constituents increases, the likelihood of *which*, rather than *that*, also rises. The second group of language-internal factors is related to prescriptivism. Hinrichs, Szmrecsanyi and Bohmann (2015) devise an indirect way of determining whether prescription has an influence on the choice of relativiser, by defining a set of predictors related to other prescriptivist advice given in the same style guides (see e.g. Strunk & White 2000) that advocate the so-called *that*-rule. This rule prescribes the use of *that* in restrictive relative clauses, while it aims to restrict the use of *which* to non-restrictive relative clauses. The assumption is that writers are likely to follow prescriptive advice consistently. For example, writers who disregard advice to avoid passive constructions will probably also not comply with the *that*-rule. We investigate the interplay between the language-external (period and variety) and language-

internal predictors by means of a multivariate analysis of 8283 annotated restrictive relative clauses.

Our findings are presented in two parts. Firstly, we consider a number of contingency tables that model the distribution of restrictive relative markers *that* and *which* across time and geography. Our analysis shows that while the relativiser *that* increases in both British and Australian Hansard material, the increase is especially pronounced in the British Hansard between 1965 and 1995, while the shift seems to have happened more gradually in Australian Hansard material. Secondly, a number of conditional random forest models (Breiman 2001) are calculated to compare the relative importance of constraints. In this respect, we find that language-external predictors are the most important constraints on variation in relative markers, followed by processing-related constraints. Prescriptivism-related variables generally tend to be less important, with the exception of the predictor passive-active ratio which is the third most important predictor for both Australian and British English.

The next section provides a concise overview of relevant literature on (a) the relative marker alternation, (b) research on Australian and British English, and (c) prescriptivism and its influence on language variation and change. Section 3 reflects on the methodology that was used to investigate the relative marker alternation in this study. Our hybrid spoken-written corpus is presented in some more detail in Section 3.1, whereafter we give a technical description of the variable context of interest in Section 3.2. Section 3.3 introduces the different predictors used in the analysis, and Section 3.4 outlines the analysis methods. In

Section 4, we present the findings of the analysis, focusing first on the distribution of the relativisers over time and across varieties, and then on the importance of the different predictors in conditioning the variation. Finally, Section 5 offers some concluding remarks and suggestions for future research.

2. Background

2.1 The relative marker alternation

The literature distinguishes between two types of relative clauses in English, namely restrictive relative clauses (RRCs) and non-restrictive relative clauses (NRRCs), as in Example (2a) and (2b) respectively.

(2) a. In China no consideration is shown to witnesses who are brought up and detained as long as may be thought necessary in the interests of justice, and are tortured if they do not give the evidence [that is expected of them]$_{RRC}$. (Australian House of Representatives Hansard, 30 August 1901)

b. This harmless-looking Amendment, [which is almost like a drafting Amendment and which has been presented partly as an improvement in drafting]$_{NRRC}$, has considerable

implications. (British House of Commons Hansard, 22 January 1965)

According to Biber et al. (2007: 195), "[r]estrictive relative clauses are used to establish the reference of the antecedent, while non-restrictive relatives give additional information which is not required for identification". The restrictive relative clause in Example (2a) identifies the antecedent *evidence*: the clause distinguishes the antecedent *evidence* from other pieces of evidence by adding that this is the evidence *that is expected of them.* In the relative clause in Example (2b), some additional information about the *harmless-looking Amendment* is given, but this information is "not required for identification" (Biber et al. 2007: 195). In general, relative clauses can be introduced by a range of relativisers: *that, ø,*[1] *what, when, where, which, who, whom* and *whose*. The relativiser *that* can only introduce restrictive relative clauses in Standard English, while *which* is used for both restrictive and non-restrictive relative clauses.

In this paper, we focus on restrictive relative clauses in subject position; in other words, restrictive relative clauses of which the relativiser introducing the clause is also the subject of the clause. In standard English, such restrictive clauses can be introduced by *that, which* or *who*. Biber et al. (2007) explain that the "relative pronoun *which* rarely occurs with an animate head. Although

[1] In general, *ø* and *zero* are used interchangeably to refer to the 'non-existing' relative marker of relative clauses that start without a relativiser.

which is attested in conversation as a relative pronoun with animate heads, this occurs so rarely that it might be considered a speech error" (Biber et al. 2007: 613). In contrast, the relative marker *who* is exclusively used with animate heads (Biber et al. 2007: 612). The relative marker *that*, however, introduces both animate and inanimate antecedents. This is why we restrict our attention to relative clauses with inanimate antecedents, thus excluding the relative marker *who* from the analysis.

As mentioned in Section 1, the distribution of *that* and *which* differs among registers. Whereas *that* is more frequently used in conversations and fiction, *which* is the most popular relativiser in British and American English newspapers and academic prose (Biber et al. 2007: 610). In the same vein, Leech et al. (2009) emphasise that "there is a division between the *wh*-relatives, which are more literate in their affinity, and *that* and *zero* relatives, which are more oral" (Leech et al. 2009: 227). Sigley (1997) finds that register variation illustrates the underlying difference in formality between the two relativisers. The formal character of *which* can be explained taking a diachronic perspective. In contrast to *that* and *zero* relativisers, the first *wh*-relativisers (including *which*) only entered the English language around the twelfth century (Tagliamonte, Smith & Lawrence 2005: 77). Mustanoja (1960) suggested that the *wh*-forms were borrowed from Romance (Tagliamonte, Smith & Lawrence 2005: 93). Nowadays, linguists generally agree on the theory that the *wh*-relativisers entered the English language through formal registers, in other words, complex and prestigious styles.

Hence, the entrance of *wh*-relativisers into English can probably be qualified as a Labovian 'change from above', since language users were aware of the implicit prestigious status of these relativisers (Tagliamonte, Smith & Lawrence 2005: 93).

Even though there is a difference in formality between *that* and *which*, thus causing variation among these two relativisers, researchers have also demonstrated that the distribution of *that* and *which* is in flux both diachronically and geographically. During the twentieth century, the relativiser *that* has spread as *wh*-relativisers have been declining in frequency (Collins 2014; Hinrichs Szmrecsanyi & Bohmann 2015; Leech & Smith 2006, 2009; Leech et al. 2009). The shift is much more pronounced in American English than in British English and is "entirely due to one relative pronoun, *which*" (Leech et al. 2009: 230). Collins (2014) claims that in Australian English, the relativiser *that* "overtakes Br[itish] E[nglish] strongly in the second half of the nineteenth century and maintains a lead over [its] British English parent thereafter" (Collins 2014: 367). Collins's research indicates that in the second half of the twentieth century, there was no significant difference between American and Australian English in the *that* vs. *which* alternation. There is, however, a significant difference between Australian and British English for the same period (Collins 2014: 367). Whereas British English is regarded as the more conservative variety with regard to the relative marker alternation, American English seems to be more progressive in that respect (Biber et al. 2007: 616). The steep increase in the

frequency of *that* at the expense of *which* in American English is possibly the result of colloquialisation (Leech & Smith 2006: 195–196; Leech & Smith 2009: 179–181). Moreover, Leech et al. (2009), and Leech and Smith (2009) argue that the sharp decline of *which* in American English might also be the result of a strong prescriptive tradition in the US. The role of prescriptivism and its influence on the alternation is discussed in Section 2.3.

2.2 Australian English

The relationship between Australian and British English is especially interesting from a diachronic perspective, as Australian English has gradually diverged from British English over time. Schneider (2007) describes the linguistic evolution of postcolonial English varieties as a path consisting of five stages: (a) foundation, (b) exonormative stabilisation, (c) nativisation, (d) endonormative stabilisation, and (e) differentiation. These five stages reflect the gradual loosening of bonds between new varieties of English and British English. Following this framework, the period covered in our Australian Hansard material ranges from the third stage (i.e. nativisation) to the final stage (i.e. differentiation). The exact dates that Schneider (2007) provides for the different phases in Australian English are, however, much debated, suggesting that our corpus potentially also covers the exonormative stabilisation stage. In this respect, Peters (2014) emphasises that the shift from exonormative to endonormative stabilisation is far from self-evident.

Although the endonormative codification of new linguistic norms "helps to stabilize the regional variety and contributes to its endonormativity, even settler varieties of English like Canadian, New Zealand and Australian English are affected by internal variation in the changed language habitat" (Leitner 2004, paraphrased in Peters 2014: 582).

Scholarly interest in Australian English developed in the 1960s (see e.g. Mitchell & Delbridge 1965; Ramson 1966) and gained momentum in the 1980s and 1990s (see e.g. Allan 1984; Bryant 1997; Collins & Blair 1989). While there has been considerable research on this variety since, diachronic studies are limited, largely as a result of the unavailability of suitable corpora. This issue has been resolved in recent years with the compilation of more diachronic corpora of Australian English (see e.g. Collins & Yao 2018; Kruger & Smith 2018; Kruger, Van Rooy & Smith 2019; Yao & Collins 2019), also to match available diachronic corpora of British and American English to facilitate cross-varietal studies.

2.3 Prescriptivism, varieties of English and language change

Prescriptive grammarians aim to formulate rules that distinguish 'correct' use of language from other, incorrect forms. In an attempt to define prescriptivism, Curzan (2014: 24) introduces four strands of prescriptivism that together cover all of 'prescriptivism's umbrella': (a) standardising prescriptivism, (b) stylistic prescriptivism, (c) restorative prescriptivism, and

(d) politically responsive prescriptivism. Prescriptive advice on the *that*/*which*-alternation is better known as the *that*-rule (see e.g. Strunk & White 2000): usage guides recommend to only use *that* in restrictive relative clauses and reserve *which* for use in non-restrictive relative clauses. The rule falls under the second strand of stylistic prescriptivism, as it is not historically or politically motivated and both *that* and *which* are grammatically correct options to introduce restrictive relative clauses. The only motivation for this rule is a quest for more grammatical symmetry: if non-restrictive relative clauses can only be introduced by *which*, then restrictive relative clauses should only be introduced by *that*.

The severity with which this *that*-rule and other prescriptive recommendations are enforced or adhered to is variable over time and across varieties of English. In their survey of post-war American, Australian and British usage books, Peters and Young (1997) investigate eleven variables of grammar and collocational syntax. They find that American usage grammars are in general more likely to accept proscribed forms than their British counterparts which are "more homogenous, cautious and conservative" (Peters & Young 1997: 322). Moreover, the range of advice given in British guides is more restricted than the advice provided in American publications. The *that*-rule is somewhat unusual in this respect, as it has consistently been implemented by American editors (see e.g. Strunk & White 2000), whereas "British and Australian usage guides take a more nuanced approach to the

distribution of *that* and *which*" (Peters & Young 1997, paraphrased in Peters 2014: 584).

The influence of prescriptive or usage grammars on language is a controversial topic in modern linguistics. The descriptive linguistic approach that has held sway throughout the twentieth century developed in part in response to prescriptivism, and attempts to approach language "neutrally and scientifically" (Straaijer 2016: 235). Unfortunately, the way in which descriptivism has set itself up in contrast with prescriptivism has been coupled with a general lack of interest in prescriptivism as a topic of (descriptive) linguistic research (Kroch & Small 1978: 45), although there have been exceptions to this.

Prescriptivism is sometimes invoked as a factor in actual language change, but this claim remains speculative. In her book on the influence of nineteenth-century grammar writing on the English verb system, Anderwald (2016) concludes that "both the prescriptive nature and the prescriptive impact of nineteenth-century grammar writing on actual language change have been greatly overestimated by previous scholars" (Anderwald 2016: 245). In their research on future temporal reference in French, Poplack and Dion (2009) discover "a great and growing disconnect between the variable rules governing speech and the normative dictates that underlie the notion of the standard" (Poplack & Dion 2009: 557). In this respect, Woolard and Schieffelin (1994: 69) explain that prescriptive guidelines have little to only pernicious effects on speech, while the same norms might have a less

negligible impact on written registers. Similarly, Curzan (2014) acknowledges that at present there is no conclusive evidence for any significant influence of prescriptivism on speech, but she also warns against an easy dismissal of the phenomenon, underscoring that some evidence exists for its influence on written texts. In their study on variation in relative markers in the Brown quartet of corpora, Hinrichs, Szremcsanyi and Bohmann (2015) show that prescriptivism might be one of the factors involved in conditioning the shift from *which* to *that*. Nevertheless, they conclude that prescriptivism alone cannot be seen as having caused the change, since other factors such as colloquialisation are also involved.[2]

3. Methods and data

3.1 The Hansard: A hybrid spoken-written register

As mentioned, this article draws on a unique corpus consisting of British and Australian Hansard records. Traditionally, 'Hansard' was the name given to transcripts of British parliamentary debates, but through the years, other Commonwealth countries (e.g. Australia, Canada, Fiji, New Zealand) began

[2] Hinrichs, Szmrecsanyi andd Bohmann (2015: 830) also note that while grammar checkers these days do help enforce the *that*-rule, they cannot have set the drift towards *that* in motion, as word processors such as Microsoft Word started implementing grammar checkers only in the 1990s.

to use the same name for similar reports of their own parliamentary debates. The uniqueness of the Hansard as linguistic data resides in the fact that it is an edited written reconstruction of parliamentary debates and speeches. Consequently, one aspect that has occupied linguists is the extent to which this hybrid spoken-written register is an accurate representation of the original recordings (see e.g. Slembrouck 1992; Mollin 2007). Moreover, how did editorial practices for the Hansard (in different countries) change over time? (See also this volume, Chapter 2.)

While the Australian Hansard adopted a 'verbatim' policy from the beginning of the federal parliament (in 1901), the British Select Committee on Parliamentary Debates only adopted such a policy for the British Hansard in 1907. Before this time, the British Hansard was a mixture of reported and direct speech, with substantial paraphrase (see also this volume, Chapter 1). The Select Committee defined the notion of the full, verbatim report as one which, "though not strictly verbatim, is substantially the verbatim report, with repetitions and redundancies omitted and with obvious mistakes corrected, but which on the other hand leaves out nothing that adds to the meaning of the speech or illustrates the argument" (Hansard 1907, quoted in House of Commons Information Office 2010: 3). This is the concept of 'verbatim' that continues to underlie the production of the Hansard to this day.

It is therefore, evidently, not possible to interpret the Hansard as a true verbatim record of spoken production. In a study on British Hansard reports from March 1987 and their corresponding recordings, Slembrouck (1992)

notices “considerable discrepancies between the spoken proceedings and what is reported in direct form” (Slembrouck 1992: 117). Instead of interpreting the Hansard as a verbatim record, we regard parliamentary records as reconstructions of speech, and specifically reconstructions in which various agencies have participated. Kruger, Van Rooy and Smith (2019: 187) explain that three groups of language users participate in this “fused production”:

> (a) the parliamentarians who produce speeches and participate in debates; (b) Hansard reporters, who transform the speeches and debates into written material; and (c) Hansard (sub-)editors, who polish the written record and apply the house style of the publication (Kruger, Van Rooy & Smith 2019: 187).

These groups are, in turn, influenced by three different sets of external mechanisms. Firstly, there is the broader social context in which the parliament itself is also embedded as an institution. The broader social context influences the three groups of language users when, for example, the society becomes less hierarchical and more democratised. Secondly, there are the factors “related to the immediate production context of the Hansard: editorial policy and practice, and production mechanisms” (Kruger, Van Rooy & Smith 2019: 188). This means that changes in the Hansard style manuals or innovations in the existing reporting techniques (e.g. digital recordings) also

have had an impact on the Hansard. The remaining factors deal with "the relationship between the parliamentary debate/Hansard and the context. These are the aim or function of the register and the intended audience" (Kruger, Van Rooy & Smith 2019: 188). The radio and television broadcasts of parliamentary proceedings caused a change in the nature of the debates, as speeches became increasingly prepared and parliamentarians started realising that they were 'performing' in front of a larger audience. It is precisely these interacting factors which make the Hansard a suitable corpus for a study on the influence of prescriptivism or even colloquialisation, but the same interacting factors will also yield some interesting insights into the overlaid processing mechanisms that lie at the basis of this corpus.

The Hansard texts included in the corpus for this study were initially collected from the publicly accessible digital archives of the Australian House of Representatives and the British House of Commons Hansard, for five sample years, 1901, 1935, 1965, 1995 and 2015 (see Kruger & Smith 2018; Kruger, Van Rooy & Smith 2019 for a more detailed description of the two corpora). Our corpus consists of a sample of 50 text files, each containing a single full day's sitting, from these larger corpora. The files were selected in a random fashion for each respective year and dialect, although we ensured that the sittings were distributed over the year, and controlled for the number of tokens to make sure that the ten variational settings contain an approximately similar number of words. In Table 9.1, we give an overview of the distribution of these text files across British and Australian Hansard

materials and across the different sampling years. At the end of each row, the total number of texts for that period is indicated between parentheses. Note that the files are not equally distributed across different years. The reason for this is that certain text files contain fewer words than others. The Australian Hansard material of our corpus consists of 1,362,134 words, while its British counterpart contains 1,555,875 words. The corpus shown in Table 1 was part-of-speech (PoS) tagged using the Stanford Tagger.

Table 9.1. Corpus composition

Variety	Year	Token count	Dates of sittings
AusE	1901	203,713	13 December, 27 September, 28 June, 29 November, 30 August, 31 July, 31 May (7)
	1935	271,572	5 December, 10 April, 26 September, 28 November, 29 March, 31 October (6)
	1965	275,682	10 December, 26 May, 29 April, 29 October, 30 March, 30 November (6)
	1995	300,770	26 October, 28 February, 30 August (3)
	2015	314,081	17 September, 22 October, 26 February (3)
BrE	1901	250,015	1 March, 8 August, 18 February, 25 June, 26 April (5)
	1935	350,634	10 December, 24 October, 28 March, 29 April, 31 July (5)
	1965	285,918	22 December, 22 January, 26 February, 29 October, 30 April, 30 July (6)
	1995	324,096	28 April, 28 February, 30 June, 30 November, 31 October (5)
	2015	347,111	26 March, 28 May, 30 November, 30 October (4)

3.2 The variable context and corpus retrieval

Our focus in this study is on restrictive subject relative clauses with inanimate antecedents which can be introduced by either *that* or *which*. In a first step, candidate relativisers were automatically selected by means of a Perl-script, based on the following two conditions. Firstly, the relativiser had to be preceded by a noun. The noun could be either a single or mass noun, a plural noun, a singular proper noun or a plural proper noun. This implies that the relative markers *that* and *which* cannot be preceded by a comma, and thus non-restrictive relative clauses were excluded by this criterion. The Australian *Hansard Style Manual* (Hansard 2017) adopts the following policy for punctuation with restrictive and non-restrictive relative clauses:

> The insertion or omission of a comma before a relative (…) clause is necessary to inform the reader whether that clause is defining or non-defining. If the relative clause defines – that is, contains information which is an essential part of the meaning of the sentence – it should not be marked off with commas (and should be introduced with "that" or "who"). (Hansard 2017: 82)[3]

[3] The style manual from which this fragment was taken is available on the Internet. It is, however, not the only circulating manual. Another style manual of the Australian Hansard that was created in 2008 (see Hansard 2008) gives more or less the same advice, but it differs on one striking point: the last part "(and should be introduced with 'that' or 'who')" is left out. In other words, editorial guidelines changed sometime between 2008 and 2017, with the '*that*-rule' codified in the later style manual. Such an adaptation in editorial policies may

The second selection criterion involves the word that occurs after either *that* or *which*. A first constraint is that the word which follows the relativiser cannot be a personal pronoun. If this were the case, we would be dealing with a restrictive relative clause on object position instead of one on subject position. A similar constraint was placed on other word classes, such as possessive pronouns, adjectives, nouns and determiners.

The restrictive relative clauses extracted in this way were manually inspected by the first author. If the clauses were indeed restrictive relative clauses on subject position with inanimate antecedent, they were further manually annotated. If not, they were removed from our data set. This process yielded a data frame containing 8283 observations.

3.3 Predicting relativiser choice

We distinguish between three different types of predictors in this study: (a) language-external, (b) prescriptivism-related, and (c) processing-related predictors.

3.3.1 Language-external predictors

have influenced the relative marker alternation, an important point raised in the discussion of the findings (see Section 4.1).

We include two language-external predictors, namely **year (of composition)** and **variety of English**. For the latter variable, we compare Australian Hansard material to British Hansard material. As for **year of composition**, we randomly sampled Hansard materials from five distinct points in time, namely 1901, 1935, 1965, 1995, and 2015 (as described in Section 3.1).

3.3.2 Prescriptivism-related predictors

The analysis in Section 4 includes six prescriptivism-related predictors: ***shall-will* ratio**, **passives**, **preposition stranding**, **split infinitives**, **presence of *you***, and **contractions**. The first four predictors are partially replicated from Hinrichs, Szmrecsanyi and Bohmann (2015), who identify these as four variables about which "prescriptivists were propagating fairly consistent advice throughout most of the twentieth century" (Hinrichs, Szmrecsanyi & Bohmann 2015: 807). Some usage grammars advise their readers to use *shall* as first-person future marker (see e.g. Strunk & White 2000), with *will* proscribed in this context. Many prescriptivist sources continue to discourage preposition stranding and the use of split infinitives. Strunk and White (2000) emphasise that "the active voice is usually more direct and vigorous than the passive" (Strunk & White 2000: 28), and the active voice is therefore regarded as the prescribed form in this study.[4]

[4] We hasten to add that usage of the passive may in principle also be considered a predictor related to language processing, as we know that all other things being equal, processing passive sentences is harder than processing active sentences (e.g. Lee & Doherty 2019).

The two latter predictors are prescriptivism-related predictors particular to the Hansard (Kruger & Smith 2018). In the case of the use of *you*, parliamentary convention dictates that third-person forms of address are to be used when members of parliament refer to one another, except where the Speaker is being addressed. Kruger and Smith (2018) and Kruger, Van Rooy and Smith (2019) show that over time, this rule is enforced less strictly in the Australian Hansard, but maintained in the British Hansard. In the case of the use of contracted forms, this particularly spoken-language feature is largely discouraged in the Hansard.

We arbitrarily chose a sufficiently large window of 2000 words to calculate the frequency of prescriptivism-related variables. This window was placed around the relativiser, so that typically 1000 words before and after the relative marker are considered. In cases where the relative marker occurred less than 1000 words from the beginning or end of the file, we simply adapted the location of the window by including all the words from the beginning/end of the file until the relative marker (window size x) and extending the window on the other side of the relativiser (window size $2000 - x$).

Besides the predictor **presence of *you***, all other prescriptive predictors are calculated as the number of proscribed forms divided by the number of prescribed forms. A small quotient indicates that many prescribed forms occur in the window, while a larger quotient means an increase in the number of proscribed forms. ***Shall-will* ratio** was calculated as the number of *will*-

tokens divided by the number of *shall*-tokens. The predictor variable **passives** presents the proportion of passive conjugated verbs vis-à-vis active conjugated verbs. **Preposition stranding** reflects the fraction of stranded prepositions divided by non-stranded prepositions. Values for the variable **split infinitives** equals the number of split infinitives divided by the number of infinitives that are preceded by *to*. The predictor **contractions** indicates the frequency of contractions vis-à-vis the frequency of the corresponding full forms of the words. For the variable **presence of *you***, we consider the number of times the word *you* occurred in the 2000-word window.

3.3.3 Predictors related to language processing

Language processing refers to the way in which language users choose particular words or constructions to express their ideas, and to the way in which these communicated words are processed and understood by other language users. In this respect, writing differs from speech in that it is not influenced by processing pressures related to time constraints. In addition, in contrast to speech, writers also do not necessarily share the same communicative context as their readers, which has consequences for the linguistic structures selected. As discussed in Section 3.2, we view the Hansard as a hybrid spoken-written register produced by an overlay of the language production of two groups: parliamentarians produce spoken debates and speeches, after which reporters and editors reconstruct their spoken language into written texts. Therefore, the hybrid texts from the Hansard may

shed light on the combined linguistic processing of two groups of language users. The seven processing-related variables included in this study are **persistence**, **persistence distance**, **distance antecedent HN start RC**, **distance AP start RC**, **length RC**, **distance AP end RC**, and **mean word length**.

- **Persistence**: It has been found that once language users have chosen for a particular construction, they are more likely to repeat the same construction further on in the text or speech (Szmrecsanyi 2006). Persistence is a Boolean variable that compares the current variant (i.e. *that* or *which*) to the variant chosen the last time there was a choice, namely *that*, *which*, or *non-applicable* (used in case the current variant is the first relative marker in the file). Concretely, persistence is 1 if the current and the previous relativiser are identical, while it receives a 0 value when this is not the case. Instances for which no previous relativiser exists (*non-applicable*) are not considered for this variable.
- **Persistence distance**: This variable indicates the distance (in words) between the previously used relativiser and the current one in the corpus file. If the distance between these two relativisers increases, the odds that language users will re-use the previous relativiser are supposed to decrease.
- **Distance antecedent HN start RC**: This is the number of words that occur between the antecedent head noun and the relativiser introducing the relative clause, including the antecedent head noun. Hinrichs,

Szmrecsanyi and Bohmann (2015: 826) observe that "[a]s antecedents and relative clauses become longer, and processing load becomes larger, writers increasingly favour (…) *which*".

- **Distance AP start RC**: This variable is the distance (in words) from the start of the antecedent phrase to the respective relativiser. This predictor is a measure of complexity.
- **Length RC**: Length (in words) of the restrictive relative clause. This predictor is another measure of complexity.
- **Distance AP end RC**: The sum of **distance AP start RC** and **length RC**.
- **Mean word length**: This last predictor measures the lexical complexity of the text surrounding the relativisers. The hypothesis is that the relativiser *which* is favoured in linguistic environments with long and complex words, thanks to its formal character. The variable **mean word length** is computed in the 2000-word window. In this window, no distinction is made between punctuation marks and words. Since punctuation marks are included in a consistent way, the predictor **mean word length** can be used for the purpose of statistical analysis. Nevertheless, the actual mean length will probably be higher than the values of this predictor.

3.4 Analysis

Our analysis proceeds in two parts. We first consider a number of contingency tables that model the distribution of the restrictive relative markers *that* and *which* across time and varieties of English (see Section 4.1). Subsequently, we assess the relative importance of predictors by carrying out a conditional random forest analysis, as introduced by Breiman (2001) and implemented in the 'party' package in R (Strobl, Malley & Tutz 2009). In this package, random forests are calculated on the basis of numerous conditional inference trees in order to create a ranking of the importance of each individual predictor in the alternation. Szmrecsanyi et al. (2016: 113–114) concisely summarise the principles behind conditional inference trees and conditional random forest analysis as follows:

> [Conditional inference trees] predict outcomes by recursively partitioning the data into smaller and smaller subsets according to those predictors that co-vary most strongly with the outcome. Informally, binary splits in the data are made by trying to maximize the homogeneity or 'purity' of the data partitions with respect to the values of the outcome [...]. The CRF [conditional random forests] approach is based on conditional inference trees; unlike individual inference trees, however, CRF uses ensemble methods in a forest of trees built on randomly sampled data subsets to arrive at an aggregated estimate of a particular outcome's probability. Additionally, each tree is restricted to a random subset of predictors whose significance is assessed through

> random permutation tests. By amalgamating the results over the entire forest of trees, we obtain a model that is both highly accurate and robust to predictor multicollinearity and data overfitting [...].

4. Results and discussion

4.1 Distributional analysis of the relative markers which *and* that

We begin by inspecting the effect of the language-external predictors on the alternation among subject relative markers *that* and *which*. Three contingency tables were created: Table 9.2 reports the distribution of relativisers in our corpus of Hansard material, Table 9.3 shows the diachronic distribution of the relativisers for the Australian Hansard material, while Table 9.4 shows this distribution for the British English Hansard material. The first line in the cells indicates the raw frequencies, while the second line shows the column percentages. On the basis of the column percentages in Tables 9.3 and 9.4, we constructed two column charts to further visualize the diachronic shift (Figure 9.1 and 9.2).

Table 9.2. The distribution of subject relative markers *that* and *which* across the two Hansard corpora

	Australian Hansard	**British Hansard**	**Row total**
that	2213 52.9%	1878 45.8%	4091
which	1967 47.1%	2225 54.2%	4192
Column total	4180 50.5%	4103 49.5%	8283

Pearson's χ2-test: $\chi 2 = 42.60$; $df = 1$; $p < 0.001$

Table 9.2 shows that there are more instances of *that* introducing subject relative clauses with inanimate antecedents in the Australian Hansard (52.9%), while in the British Hansard, *which* seems to be the preferred relativiser (54.2%). The preference for the relativiser *that* in the Australian Hansard material might reflect the more permissive nature of the Australian Hansard towards "spoken-language features in writing" (Kruger, Van Rooy & Smith 2019: 205). The distribution of the relative markers across the two varieties differs significantly.

Table 9.3. The diachronic distribution of subject relative markers *that* and *which* in the Australian Hansard material

	1901	**1935**	**1965**	**1995**	**2015**	**Row total**
that	185 31.4%	228 30.6%	486 52.5%	515 54.0%	799 82.7%	2213
which	404 68.6%	518 69.4%	440 47.5%	438 46.0%	167 17.3%	1967
Column total	589 14.1%	746 17.8%	926 22.2%	953 22.8%	966 23.1%	4180

Pearson's χ^2-test: $\chi^2 = 603.77$; $df = 4$; $p < 0.001$

Table 9.4. The diachronic distribution of subject relative markers *that* and *which* in the British Hansard material

	1901	**1935**	**1965**	**1995**	**2015**	**Row total**
that	88 14.6%	303 26.2%	110 15.2%	664 79.2%	713 91.1%	1878
which	513 85.4%	853 73.8%	615 84.8%	174 20.8%	70 8.9%	2225
Column total	589 14.1%	746 17.8%	926 22.2%	953 22.8%	966 23.1%	4180

Pearson's χ^2-test: $\chi^2 = 1711.42$; $df = 4$; $p < 0.001$

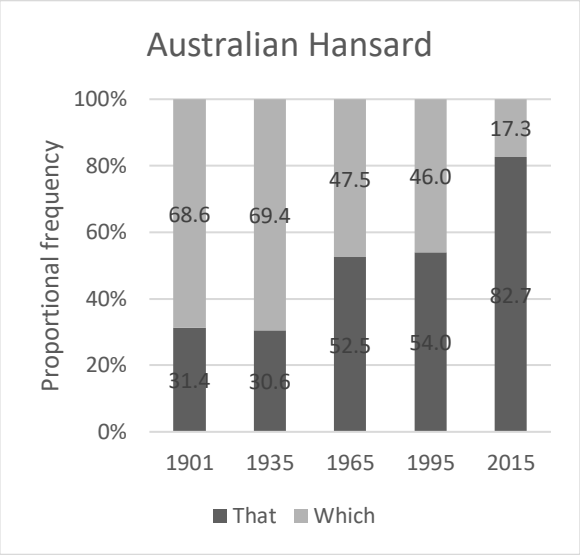

Figure 9.1. The diachronic distribution of subject relative markers *that* and *which* in the Australian Hansard material

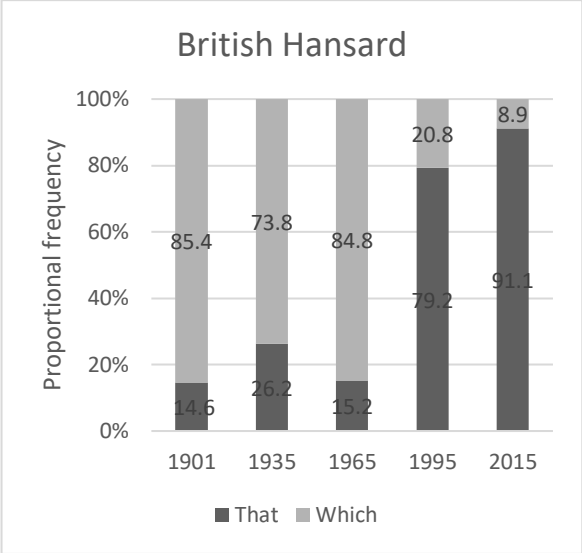

Figure 9.2. The diachronic distribution of subject relative markers *that* and *which* in the British Hansard material

Tables 9.3 and 9.4 and Figures 9.1 and 9.2 further indicate that the relative frequency of *that* has increased over time in both varieties. This shift is moderate in the Australian Hansard: the relativiser *that* increases by 21.9 percentage points and 28.7 percentage points between 1935 and 1965 and 1995 and 2015, respectively. As for the other periods, there are no significant diachronic developments in the alternation. The rise of *that* and the corresponding decline of *which* are, however, particularly pronounced in the British Hansard. In 1965, the relativiser *which* was still the preferred form by far to introduce subject relative clauses with inanimate antecedents. This situation shifted drastically in the next 30 years: by 1995, *that* had increased by 64 percentage points at the cost of *which*. During the next 20 years, the shift continued: in 2015, the relative marker *which* was used in only 8.9% of

choice contexts. In sum, while *that* is more frequent in the British Hansard than in its Australian counterpart in the two last sampling years, *that* occurred twice as often in the Australian Hansard material than in its British counterpart at the beginning of the twentieth century.

Some of these results tie in with the general literature on the variation in relative markers. Firstly, a shift from *which* to *that* does indeed occur in both the Australian and the British Hansard material. Moreover, in the beginning of the twentieth century, the relativiser *that* is more frequently used in the Australian than in the British Hansard material. That being said, Collins (2014) argues that regarding the use of *that*, AusE "maintain[s] a lead over [its] Br[itish] E[nglish] parent" (Collins 2014: 367). This claim does not hold for the Hansard, as *that* becomes more frequent in the British Hansard than in the Australian Hansard. Our results are in line, however, with results from research on the Brown family of corpora (Hinrichs, Szmrecsanyi & Bohmann 2015; Leech & Smith 2006, 2009; Leech et al. 2009), in that a shift from *which* to *that* occurs also in the hybrid written-spoken register. Whereas previous research on written standard English has characterised British English as the conservative variety regarding relativisation, our findings suggest that this trend is not supported for the Hansard materials: in the two last sampling years of our corpus, British English proves to actually be more progressive than Australian English in shifting to *that*.

Several factors that are specific to the Hansard register might have contributed to the shift from *which* to *that*. Firstly, the editorial guidelines of

the Australian Hansard have changed sometime between 2008 and 2017 (see Hansard 2008, 2017), with the '*that*-rule' being codified in the later style manual. Although we do not have access to the precise editorial guidelines that were used in 1995 and 2015, we suspect that the increase of *that* between these years might be the result of such an adaptation of the editorial guidelines underlying the creation of the Australian Hansard. In addition, the production mechanisms of the Hansard drastically changed in this period, due to the gradual shift from shorthand reporting to recordings and voice recognition.

As for the earlier period, the rise of *that* between 1935 and 1965 in the Australian Hansard coincides with the introduction of radio broadcasts of parliamentary proceedings in 1946. The argument is that as broadcasts and recordings started, it became possible to compare the actual spoken records with the written Hansard, which probably led to a more verbatim reporting style. If parliamentarians were using more *that*, as one would expect from speech, it would have been transferred more accurately to the Hansard.

We were unable to find editorial guidelines for the British Hansard. Hence, we can only hypothesise that the strong rise of *that* and corresponding decline of *which* between 1965 and 1995 is the result of some similar changes in editorial style guides. The production mechanisms of the British Hansard only started to change in the early 1980s when, in some contexts, tape recordings were beginning to be used in producing the British House of Commons Hansard (Gravlee 1981: 88). While by 2015, the move to recordings and voice transcriptions would have been made, it remains unclear

whether this was already the case in 1995. In Britain, there initially was also considerable resistance against recording parliamentary proceedings, and it was not until 1978 that regular radio broadcasts began (Hendy 2019).

4.2 Probing the importance of constraints on relativiser choice

Based on conditional random forest modelling, Figure 9.3 plots the variable importance ranking for the Australian Hansard material, while Figure 9.4 plots the ranking for the British material. The vertical dotted line in the plots is the absolute value of the least significant variable in conditioning the variation. By convention, this line represents the threshold between significance and insignificance: all predictors to the right of the line play a significant role in the alternation, while all predictors to the left of the line are insignificant. The indices of concordance indicate that the underlying models have a good fit (*C*-index > 0.85).[5]

The four most important predictors are identical in both varieties: the language-external predictor **year of composition** is the most significant predictor in both British and Australian English, which underlines the massive real time shift towards *that* in both varieties. To get an impression of

[5] The *C*-index (sometimes also referred to as concordance statistic or *C*-statistic) is a measure of goodness of fit. It can be used to determine the robustness or performance of classifiers (including random forests). Values over 0.8 indicate that the model is strong, meaning that it is good at predicting our binary outcome variable (i.e. *that* or *which*) by means of the different predictors.

the role played by the other language-external predictor **variety**, we calculated a supplementary conditional random forest model (*C*-index = 0.92) (not shown here) based on a combined British-Australian dataset. This random forest ranks **variety** as the third highest ranked predictor, after **year** and **persistence**.

Second, in both varieties, predictors that are related to processing constraints and effects seem to play a more significant role in conditioning the choice between *that* and *which* than prescriptivism-related predictors do. One exception to this general trend is the predictor **passives**, which ranks third in both varieties, although we should remark that this predictor is not entirely unrelated to processing constraints (see Footnote 5). The two most important processing-related predictors are the same in both varieties, namely **persistence** and **distance AP end RC**. **Distance antecedent HN start RC**, **mean word length** and **persistence distance** are processing-related predictors that are highly ranked in the random forest for the Australian English parliamentary material, whereas **mean word length**, **length RC** and **distance antecedent HN** are important in the British English material.

While **passives** is the highest-ranked prescriptivism-related predictor in both varieties, other prescriptivism-related predictors also contribute to a certain, somewhat lower degree in conditioning this variation. **Preposition stranding** and ***shall-will* ratio** are respectively the second and third most influential prescriptivism-related predictors in the Australian English data. The same constraints are important in the random forest for the British

English data: ***shall-will* ratio** is ranked as sixth most important predictor, while **preposition stranding** is indicated as the ninth most significant predictor.

In other words, prescriptive recommendations might have contributed to the shift from *which* to *that*, but prescriptivism alone cannot account for the change. Moreover, it is hard to evaluate to what extent general style guides have influenced the editorial guidelines of the Hansards, as we only have two (Australian) editorial guides at our disposal (see Section 4.1). Although the more extensive *Hansard Usage and Editing Guide* (Hansard 2008) gives some suggestions[6] on how to deal with preposition stranding, contractions, and split infinitives, it does not deal with *shall*/*will* or active/passive voice. We suspect, however, that the influence and type of prescriptive advice varies across sampling periods, as the guidelines probably underwent considerable changes during the twentieth century. Nevertheless, we would need more data to make reliable estimations for the different sampling years.

In addition, other processes such as colloquialisation are probably also involved in the shift, with parliamentarians more likely to use *that* in their spoken discourse. Accompanying this, there is a tendency towards a more verbatim rendering in the Hansards, prompted by widespread broadcasting and increasingly synchronised audio/video and written records, as well as

[6] These suggestions first refer to some long-standing 'rules' in the prescriptive tradition, but then underscore the importance of "giv[ing] speakers what they say" (Hansard 2008: 16); in other words remaining close to the original speech.

change towards production mechanisms that encourage verbatim transcription rather than edited reconstruction. It is likely that these factors act in consort in accounting for the observed changes.

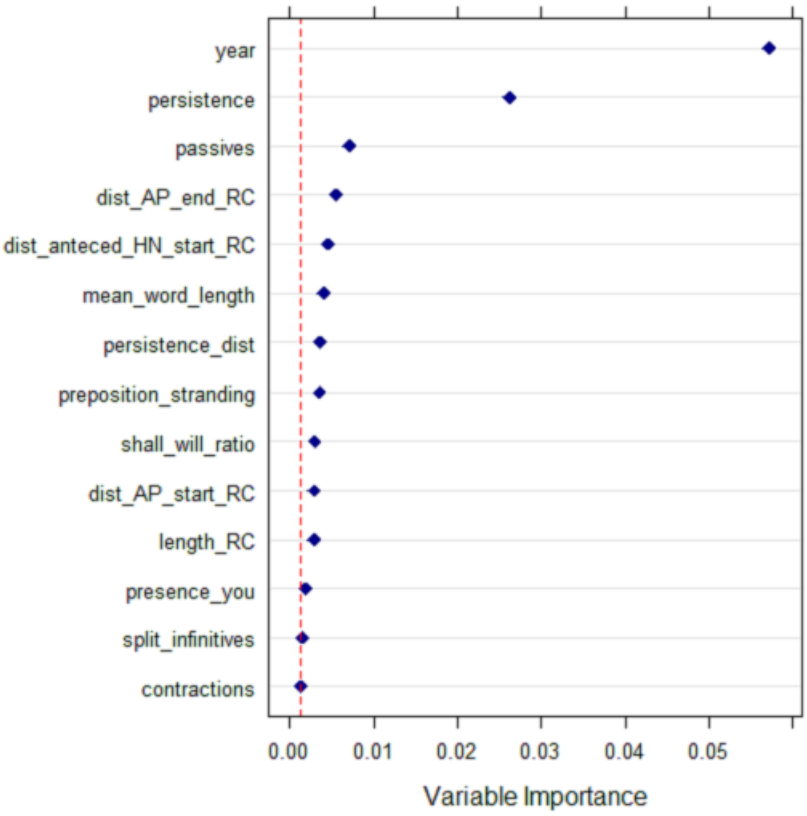

Figure 9.3. Random forest of the Australian Hansard data (*C*-index = 0.90)

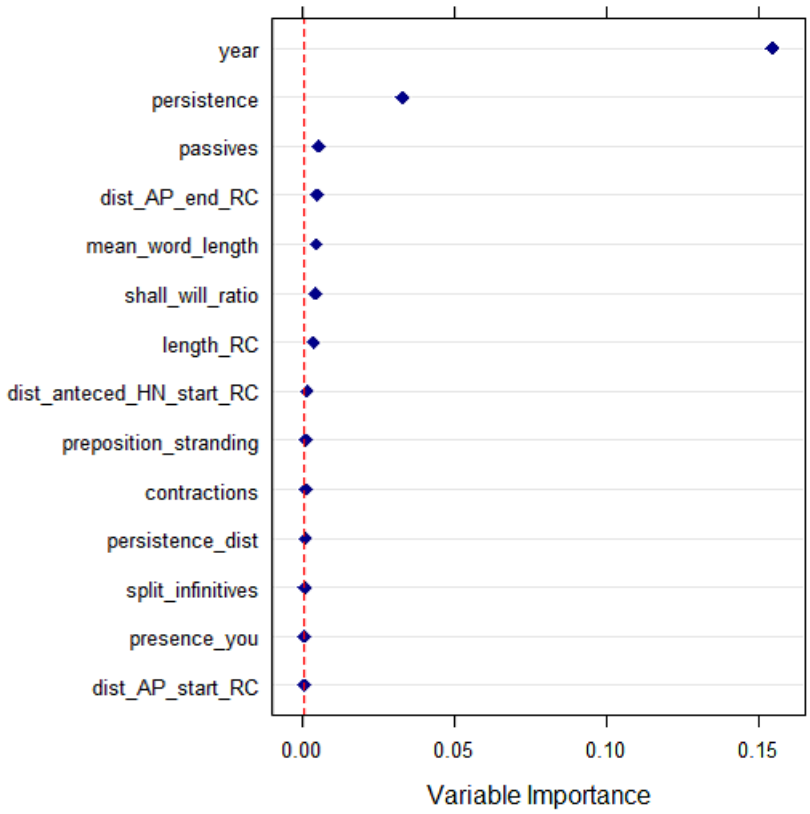

Figure 9.4. Random forest of the British Hansard data (*C*-index = 0.95)

5. Conclusion

In this paper, we investigated variation between *that* and *which* in restrictive subject relative clauses with inanimate antecedents in British and Australian English Hansard material from the period between 1901 and 2015. Taking the approach of comparative variationist analysis, we investigated relative frequencies and explored the conditioning of variation, distinguishing between three types of predictors: (a) language-external predictors, (b) processing constraints and effects, and (c) prescriptivism-related constraints.

Our point of departure was that previous research has shown that the relative marker *that* has been on the rise throughout the twentieth century, at the expense of the relativiser *which* (Collins 2014; Hinrichs, Szmrecsanyi & Bohmann 2015; Leech & Smith 2006, 2009; Leech et al. 2009). We find evidence that a similar drift is underway in the Hansard material under study here. Whereas the shift is relatively moderate in the Australian material, the rise of *that* at the cost of *which* is particularly pronounced in the British material. Although at the beginning of the twentieth century, *that* occurred twice as often in the Australian Hansard (31.4%) as in the British Hansard (14.6%), the situation has shifted drastically by 2015, as the relative marker *which* occurs only in 8.9% of the British and in 17.3% of the Australian restrictive subject relative clauses.

By means of conditional random forest analysis, we set out to determine which predictors are most influential in conditioning the choice between either *that* or *which*. The analysis shows that the language-external predictor of timeframe is the most important predictor in the relative marker alternation in both varieties – and variety itself is an important predictor in an analysis of the full dataset, suggesting real-time differentiated change in the two varieties in question. Language-internal predictors related to language processing play an important role in conditioning the choice, while some individual prescriptivism-related predictors, such as passives, also correlate significantly with the outcome variable. Nevertheless, prescriptivist recommendations alone cannot account for the shift from *which* to *that*.

The findings of the study support the much-highlighted colloquialism of Australian English, in that the more informal *that*-variant is proportionally more frequent in the Australian Hansard from the outset. This informality in all likelihood partially emanates from the speech input of parliamentarians; however, the Australian Hansard editorial guidelines for representing speech are also permissive of reflecting such speech accurately, as evident in the oft-repeated dictum of "give speakers what they say" (Hansard 2008). While starting out from a higher proportional frequency of relativiser *that*, the further increase seems to occur from 1965 onwards, already. While there is some disagreement about the periodisation of endonormative stabilisation of Australian English, Moore (2008) argues that the period from the 1960s to the 1980s was crucial in the "discarding of the external English standard of language" (Moore 2008: xii), and Yao and Collins (2019) find that indeed Australian English experienced substantial (and rapid) grammatical change from 1961 to 1991. Notable is also the publication of the first *Australian Government Style Manual* in 1966 (now in its seventh edition), which marked a clear encoding of Australian grammatical and stylistic standards.

It should be noted, however, that even though the Australian Hansard is more 'progressive' in adopting the more frequent use of the relativiser *that*, by 2015, the British Hansard has overtaken the Australian Hansard in the proportional preference for *that.* The change sets in later (it is only evident from the 1995 data onwards), and it is dramatic – rather than the more gradual change evident in the Australian Hansard. This dramatic change is suggestive

of a change in editorial policies (as also argued by Kruger & Smith 2018); in the absence of access to style manuals for the British Hansard, it is not possible to confirm this hypothesis. More focused investigations of, particularly, the influence of the *that*-rule (see Section 1) in Australian and British writing practice is needed; it may also be that this rule, particularly influential in American English, had more general uptake in Australian English than in British English.

When considered against the general assumption that usage tends to win out over prescriptivism (see Section 3.2), the findings of this study prompt us to more careful reflection: In registers that enforce careful editorial control, like the Hansard (but, in fact, all published genres), prescriptive advice no doubt plays a role. However, it evidently acts in consort with other factors, like colloquialisation, and it seems clear that the rise of *that* needs to be understood as a case where usage and prescription acted in mutual reinforcement.

Future research may expand the study of *that* or *which* by including data from other new Englishes (e.g. New Zealand English), performing regression analyses on the data in order to determine whether a predictor prefers either *that* or *which*, and including restrictive relative clauses in object position. More insight in the relationship between prescriptivism and other factors (like colloquialisation) will be gained by also analysing variables where prescriptive advice and usage pull in opposite directions.

Acknowledgements

We would like to thank the anonymous reviewers and the editors of this volume for their constructive feedback and their valuable suggestions for improvements. This research received funding from the Flemish Government under the AI Research Program – 174L00121.

References

Allan, Keith. 1984. The component functions of the high rise terminal contour in Australian declarative sentences. *Australian Journal of Linguistics* 4(1): 19–32.

Anderwald, Lieselotte. 2016. *Language Between Description and Prescription: Verbs and Verb Categories in Nineteenth-Century Grammars of English*. Oxford: Oxford University Press.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, Finegan, Edward & Quirk, Randolph. 2007. *Longman Grammar of Spoken and Written English*. 7th ed. London: Longman.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45: 5–32.

Bryant, Pauline. 1997. A dialect survey of the lexicon of Australian English. *English World-Wide* 18(2): 211–241.

Collins, Peter. 2014. Relative clauses in Australian English: A cross-varietal diachronic study. In *Selected Papers from the 44th Conference of the Australian Linguistic Society*, Lauren Gawne & Jill Vaughan (eds), 355–371. Melbourne: University of Melbourne.

Collins, Peter & Blair, David. 1989. *Australian English: The Language of a New Society*. Brisbane: University of Queensland Press.

Collins, Peter & Yao, Xinyue. 2018. Colloquialisation and the evolution of Australian English: A cross-varietal and cross-generic study of Australian, British, and American English from 1931 to 2006. *English World-Wide* 39(3): 253–277.

Curzan, Anne. 2014. *Fixing English: Prescriptivism and Language History*. Cambridge: Cambridge University Press.

D'Arcy, Alexandra & Tagliamonte, Sali. 2010. Prestige, accommodation, and the legacy of relative *who*. *Language in Society* 39(3): 383–410.

Gravlee, G. Jack. 1981. Reporting proceedings and debates in the British Commons. *Central States Speech Journal* 32(2): 85–99.

Hansard. 2008. *Hansard Usage and Editing Guide*. Canberra: Commonwealth of Australia.

Hansard. 2017. *Hansard Style Manual*. Canberra: Office of the Legislative Assembly for the Australian Capital Territory.

https://www.parliament.act.gov.au/__data/assets/pdf_file/0008/1507958/Style-Manual-2017-Accessible-version.pdf (20 February 2021).

Harder, Peter. 2012. Variation, structure and norms. *Review of Cognitive Linguistics* 10(2): 294–314.

Hendy, David. 2017. History of the BBC: Broadcasting Parliament. https://www.bbc.com/historyofthebbc/100-voices/elections/broadcasting-parliament (20 February 2021).

Hinrichs, Lars, Szmrecsanyi, Benedikt & Bohmann, Axel. 2015. *Which*-hunting and the Standard English relative clause. *Language* 91(4): 806–836.

House of Commons Information Office. 2010. *Factsheet G17 General Series*: *The Official Report.* https://www.parliament.uk/globalassets/documents/commons-information-office/g17.pdf (27 July 2021).

Kroch, Anthony & Small, Cathy. 1978. Grammatical ideology and its effects on speech. In *Linguistic Variation: Models and Methods*, David Sankoff (ed), 45–55. New York: Academic Press.

Kruger, Haidee & Smith, Adam. 2018. Colloquialization versus densification in Australian English: A multidimensional analysis of the Australian Diachronic Hansard Corpus (ADHC*). Australian Journal of Linguistics* 38(3): 293–328.

Kruger, Haidee, Van Rooy, Bertus & Smith, Adam. 2019. Register change in the British and Australian Hansard (1901–2015). *Journal of English Linguistics* 47(3): 183–220.

Labat, Sofie. 2018. *Relative markers in parliamentary speech: A real-time study of Australian and British English*. KU Leuven, Master's Thesis.

Lee, James F. & Doherty, Stephen. 2019. Native and nonnative processing of active and passive sentences: The effects of processing instruction on the allocation of visual attention. *Studies in Second Language Acquisition* 41(4): 853–879.

Leech, Geoffrey & Smith, Nicholas. 2006. Recent grammatical change in written English 1961–1992: Some preliminary findings of a comparison of American with British English. In *The Changing Face of Corpus Linguistics*, Antoinette Renouf & Andrew Kehoe (eds), 185–204. Amsterdam: Rodopi.

Leech, Geoffrey & Smith, Nicholas. 2009. Change and constancy in linguistic change: How grammatical usage in written English evolved in the period 1931–1991. In *Corpus Linguistics: Refinements and Reassessments*, Antoinette Renouf & Andrew Kehoe (eds), 173–200. Amsterdam: Rodopi.

Leech, Geoffrey, Hundt, Marianne, Mair, Christian & Smith, Nicholas. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.

Leitner, Gerhard. 2004. *Australia's Many Voices: Australian English – The National Language.* Berlin: Mouton de Gruyter.

Mitchell, Alexander George & Delbridge, Arthur. 1965. *The Speech of Australian Adolescents*. Sydney: Angus and Robertson.

Mollin, Sandra. 2007. The Hansard hazard: Gauging the accuracy of British parliamentary transcripts. *Corpora* 2(2): 187–210.

Moore, Bruce. 2008. *Speaking Our Language: The Story of Australian English*. Sydney: Oxford University Press.

Mustanoja, Tauno Frans. 1960. *A Middle English syntax*. Helsinki: Société Néophilologique.

Peters, Pam. 2014. Usage guides and usage trends in Australian and British English. *Australian Journal of Linguistics* 34(4): 581–598.

Peters, Pam & Young, Wendy. 1997. English grammar and the lexicography of usage. *Journal of English Linguistics* 25(4): 315–331.

Poplack, Shana & Dion, Nathalie. 2009. Prescription vs. praxis: The evolution of future temporal reference in French. *Language* 85(3): 557–587.

Ramson, William Stanley. 1966. *Australian English: An Historical Study of the Vocabulary, 1788–1898*. Canberra: Australian National University Press.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.

Sigley, Robert. 1997. The influence of formality and channel on relative pronoun choice in New Zealand English. *English Language and Linguistics* 1(2): 207–232.

Slembrouck, Stef. 1992. The parliamentary Hansard 'verbatim' report: The written construction of spoken discourse. *Language and Literature* 1(2): 101–119.

Straaijer, Robin. 2016. Attitudes to prescriptivism: An introduction. *Journal of Multilingual and Multicultural Development* 37(3): 233–242.

Strobl, Carolin, Malley, James & Tutz, Gerhard. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4): 323–348.

Strunk, William & White, Elwyn Brooks. 2000. *The Elements of Style*. 4th ed. London: Longman.

Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English: A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin: Mouton de Gruyter.

Szmrecsanyi, Benedikt. 2019. Register in variationist linguistics. *Register Studies* 1(1): 76–99.

Szmrecsanyi, Benedikt, Grafmiller, Jason, Heller, Benedikt & Röthlisberger, Melanie. 2016. Around the world in three alternations: Modeling

syntactic variation in varieties of English. *English World-Wide* 37(2): 109–137.

Tagliamonte, Sali, Smith, Jennifer & Lawrence, Helen. 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17(1): 75–112.

Woolard, Kathryn A. & Schieffelin, Bambi B. 1994. Language ideology. *Annual Review of Anthropology* 23(1): 55–82.

Yao, Xinyue & Collins, Peter. 2019. Developments in Australian, British, and American English grammar from 1931 to 2006: An aggregate, comparative approach to dialectal variation and change. *Journal of English Linguistics* 47(2): 120–149.