Relative Clauses in a Modern Diachronic Corpus of Singapore English

Kit Mun Lee
(National University of Singapore)

Lee, K. M. (2020). Relative clauses in a modern diachronic corpus of Singapore English. *Asia Pacific Journal of Corpus Research, 1*(1), 31-60.

This paper investigates changes in relativization in Singapore English broadsheet newspapers from 1993 to 2016. One of the first diachronic studies in Singapore English (SgE), it also explores corresponding data from the diachronic Siena-Bologna (SiBol) news corpus. As SgE is in the endonormative stabilization phase in Schneider's (2007) Dynamic Model of postcolonial Englishes, divergence from British English (BrE) is to be expected. In this study, the dataset is a new Singapore English Newspaper (SEN) corpus compiled from local news articles in 1993, 2005 and 2016, and the corpus tool employed is Sketch Engine. The results reveal changes in relativization practices in SEN over the given period, many of which occur in a similar pattern as those identified in SiBol, albeit at varying rates of change. Most significant of these include a sharp decline in the *which* relativizer in restrictive relative clauses with non-animate antecedents, complemented by a rise in *that*. The change has been so rapid that although *which* relative clauses were more common than *that* clauses in 1993, *that* has subsequently overtaken *which* for both the corpora. One shift in SEN that is different from SiBol is the increase in frequency of non-restrictive relative clauses in SgE. The likely motivators for the changes in the two varieties are identified as colloquialization, densification and prescriptivism. The effect each of these factors could have had on the varieties are discussed, as well as the implications that the findings have on our understanding of the evolutionary status of SgE as a postcolonial variety.

Keywords: Diachronic Corpus, Relative Clauses, News Writing, Postcolonial Variety, Singapore English

1. Introduction

According to Schneider's widely accepted Dynamic Model of the evolution of postcolonial English (Schneider, 2007), Singapore English (SgE) is among the more advanced varieties in Phase 4 of the model, which is endonormative stabilization. This phase is characterized by the emergence of a stable, local norm that "is beginning to be codified and to be accepted in society" (Schneider, 2010, p. 381). Much previous research has been done on SgE, and other postcolonial varieties to test or demonstrate this model. Most of such studies on SgE are synchronic and do not chart the development of SgE in real time. Recently, scholars such as Hoffmann (2018), Mair (2015), Mukherjee and Schilk (2012) and Gries et al. (2018) have opined that diachronic data are necessary for a more informed understanding of the processes at work in structural nativization.

Syntactic change is notably slow and can take long periods to reach completion, but grammatical preferences can change significantly within a period of just 30 years (Hundt & Szmrecsanyi, 2012), as short-term diachronic research by Leech et al. (2009), *inter alia*, on the Brown family of corpora, and by Partington (2010) on the SiBol (Siena-Bologna) Modern Diachronic Corpus has shown. However, few studies have been conducted to trace such changes over time in the development of new varieties of English, including SgE. The current paper reports results from an ongoing study that aims to address this lack by exploring some of these changes in SgE. It is reasonable to surmise that the 20th century linguistic developments in the colonial source variety may be mirrored to some extent in a postcolonial variety such as SgE (Mair, 2015). At the same time, an advanced Phase 4 variety as SgE may also be expected to exhibit more divergence from the source variety (Mukherjee & Gries, 2009). Since relativization is an area that has demonstrated interesting diachronic trends in 20th century English, particularly relevant to news writing, the present study focuses on relative clauses in postmodification of noun phrases.

Relative clauses are finite clausal postmodifiers comprising three major components: the head noun phrase or antecedent, the relativizer and the gap (Biber et al., 1999). The relativizer can be a

relative pronoun or adverb while the gap is the missing grammatical constituent in the relative clause and corresponds in meaning to the antecedent. Thus, in Example (1) below,

(1) the first question **which** companies always ask [ST 2016-11-19[1]]

the antededent is *the first question*, the relativizer is *which*, and the gap is in the direct object position of the relative clause, which corresponds in meaning to the antecedent *the first question*.

Previous research on relative clauses in BrE and SgE has mostly been based on data from the Brown family of corpora and the International Corpus of English (ICE). In this study, the dataset is a new Singapore English Newspaper (SEN) corpus compiled from local news articles at three temporal points, 1993, 2005 and 2016. Comparisons will be made with data from the SiBol corpus of English broadsheets of 1993, 2005 and 2013.

This paper will first provide the major features of relativizer choices, then briefly outline previous work that sets the background for the current research. Next, the methodology employed in this study will be described, followed by an analysis of the findings, ending with a discussion of the results and implications.

2. Background

2.1. Choice of Relativizer in Relative Clauses

In Example (1) above, the relative pronoun *which* alternates with *that* in standard varieties of English, or can be omitted. However, the variation does not occur freely, but only in restrictive relative clauses (RRC) as in (1). For non-restrictive relative clauses (NRRC), the relativizer *which* is used, and it is not interchangeable with *that*.

---

[1]   This example is a token extracted an article published in *The Straits Times* (ST), the national broadsheet newspaper in Singapore, on November 19, 2016.

In most cases, such NRRC are signalled by a comma preceding the relative pronoun, as in (2), or sometimes by a pair of parenthetical commas, as in (3). There are certain exceptions, as in (4), where the commas are absent but the context makes it clear that the relative clause is non-restrictive.

(2) Then, there is the bond market, **which** several bankers also recommended. [ST 1993-01-10]

(3) The National Productivity Board (NPB), **which** plans to take the lead in introducing this management tool here, said it will develop a database listing of companies which have gone into "benchmarking". [ST 1993-03-20]

(4) … in a museum complex called Bede's World **which** opens tomorrow [SiBol-G 1993-05-19]

Most relativizer variation studies (e.g. Grafmiller et al., 2018; Hinrichs et al., 2015) tend to limit their investigation to RRC because only RRC permit the variation between *which*, *that* and *zero* relativizers.

Another constraint affecting the choice of the relativizer is that of animacy of the antecedent, or more specifically, whether or not the antecedent is human or personal. Relativizers such as *who*, *whom* or *whose* are reserved for personal antecedents, *which* tends to be used predominantly for inanimate antecedents (Gut & Coronel, 2012; Ball, 1996), while the use of *that* is available for both animal and inanimate antecedents (Grafmiller et al., 2018). The animacy constraint is apparently one that emerged recently and operates differently in spoken and written English (Grafmiller et al., 2018).

2.2. Previous Work on Relative Clauses

The most common relativizers in English overall are *which* and *that*, as (Biber et al., 1999) have found in their data from the Longman Corpus of Written and Spoken English (LWSE), with *which* being more common than *that* in written genres, especially news,

and *who* being by far more common than either *which* or *that*. One of the most remarkable recent developments in English relative clauses is the increasing popularity of *that*-relatives, accompanied by the decline in *wh*-relative clauses (Leech et al., 2009). This development is remarkable probably because it demonstrates a reversal of the trend that Ball (1996) identifies in her diachronic corpora of 16th to 19th century English, where *which* was seen to be preferred over *that*. The modern trend is corroborated by previous work on various varieties of English and it is even predicted that the use of *wh*-relativizers[2] may be reduced significantly enough to "disappear in the near future" (Suarez-Gomez, 2014, p. 4) .

One of the main factors said to be responsible for the complementary trends in the use of *which* is colloquialization (Leech et al, 2009; Collins, Yao & Borlongan, 2014, p. 135). This is supported by the use of *wh-* relative pronouns in more formal contexts, being "more literate and appropriate to careful language" while   *that* and the *zero* relativizer, having "a more colloquial flavor", are used in more informal contexts and preferred in conversation (Biber et al., 1999, p. 612; Sigley, 1997). Previous studies (e.g. Suarez-Gomez, 2014; Gut & Coronel, 2012) have shown the preference for *that* in spoken English and for *which* in written English in most varieties examined, including the 'native speaker' and the postcolonial varieties.

Besides colloquialization, another possible contributing factor to this trend, which applies only to RRC, has been said by several authors (e.g. Leech et al., 2009; Hinrichs et al., 2015) to be the influence of prescriptivism. There is a rule in prescriptive literature dating back to the early 90s, particularly in the United States, stating that only *that,* and not *which,* can be used to introduce an RRC with a non-human antecedent (Leech et al., 2009). The rationale, it seems, has to do with the maintenance of grammatical symmetry and convenient mapping of form to function: since *that* can only be used in RRC, then only *which* should be used for NRRC (Collins, Yao & Borlongan, 2014; Hinrichs et al., 2015; Leech et al., 2009).

---

[2]   The relativisers *which*, *who*, *whose* and *whom* are also often referred to as the *wh*-relativisers

This tradition is still pervasive today, as can be seen from the following extract from a blog on Grammarly, a prominent digital writing assistant developed in 2009 by an American technology company:

- In a defining clause, use *that*.
- In non-defining clauses, use *which*.
  Remember, *which* is as disposable as a sandwich bag. If you can remove the clause without destroying the meaning of the sentence, the clause is nonessential and you can use *which*.
  (www. Grammarly.com/blog/which-vs-that/)

Similar advice is offered on a Dictionary.com site, but with the proviso that "the above distinction [between *that* and *which*] is a rule of formal American English, and is not as strictly observed in BrE or in informal English of any type".

Leech et al. (2009) argue that the influence of prescriptivism was a strong contributing factor for the increase of *that* at the expense of *which* because of the sharp changes observed in the American Brown and Frown corpora, as compared to the more gradual fluctuations in the parallel British LOB and FLOB corpora. Although the rule is more strongly prescribed in usage and style guides in the US than in the UK, the influence has spread to other countries throughout the world, possibly through "its incorporation in internationally marketed word processors and grammar checkers" (Leech et al., 2009, p. 230).

The use of relative clauses in SgE has been the subject of a few studies, for instance, Alsagoff and Ho (1998), which describe relativization in Colloquial Singapore English or Singlish, Newbrook (2003), which gives a qualitative description of relative clauses in SgE, and more recent studies (Gut & Coronel, 2012; Suarez-Gomez, 2014) which compare relativization strategies in SgE with other postcolonial varieties of English. These studies, based on data from ICE, report generally similar findings, some of which are highlighted below.

Suarez-Gomez (2014) report that the relativizer *that* is the most frequent in SgE, followed closely behind by *who* (also reported in

Newbrook, 2003), and further behind by *zero*, with *which* in fourth place, and *whom* and *whose* trailing behind at extremely low frequencies – a situation which is similar to British and American English. The distribution varies according to register, so it is not surprising that Gut and Coronel (2012) observe that there are more instances of *which* than *that* in written SgE, whereas the reverse is true of spoken SgE, a pattern that is similar to the 'standard' English varieties. The relativizer with the lowest frequency both in spoken and written SgE, is *whose* (Gut & Coronel, 2012).

As for the animacy of the precedent, all the three studies on SgE report a preference for *that, which* and *zero* with non-human antecedents and a preference for *who*, *whom* or *whose* with human antecedents. More specifically, NRRC (which do not use *that* as relativizer) with a non-human antecedent almost exclusively have *which* as a relative marker, while those with a human antecedent have *who/m/se* (Gut & Coronel, 2012). In addition, Deterding (2007) in his description of SgE, makes the observation that "there is a widespread belief among speakers of English in Singapore English that, for a relative pronoun with a human antecedent, use of *that* is ungrammatical and only *who* or *whom* can be used (Newbrook, 2003), even though that is in fact very common in this context in other varieties of English" (p. 54). It is indeed the case that in spoken British and American English, *that* is used for both animate and inanimate antecedents, though it is mostly used for inanimate antecedents in written English (Biber et al., 1999).

Some of the few diachronic studies on postcolonial varieties are on Philippines English (PhilE) based on a corpus that parallels the Brown family of corpora in design and sampling dates, i.e. early 1960's and 1990's. Collins, Borlongan and Yao (2014) and Collins, Yao and Borlongan (2014) explore aspects of diachronic change in PhilE compared to BrE and AmE, with the former study focusing on modality and the latter on relative clauses. In both the studies, it is reported that PhilE appears to be patterning closely with its colonial source, AmE, or "striving to 'catch up' with AmE" (Collins, Borlongan & Yao, p. 68). As a result of this close alignment, Collins et al. conclude that PhilE may not be ready to enter the endonormative

stabilisation phase.

The current study sets out to investigate three research questions:

(a) How has the overall distribution of relative clauses in SEN changed over the past 20-25 years, compared to the trends in SiBol-Guardian?

(b) Has the replacement of *wh*-relativizers by *that* in RRC continued to increase during the sampled period in the two corpora changed over the sampled period?

(c) What effect is there, if any, of animacy on the choice of relativizer (which or that) on RRC in the two corpora?

Based on diachronic research on relative clauses so far on British, American and Philippine English, we expect to find a continuing decrease in the use of *which* and *whom* over the years, happening concurrently with an increase in the use of *that*, even in news writing. Possible reasons driving the changes will be discussed, as well as the implications they have on the evolutionary status of SgE as a postcolonial variety.

3. Methodology

3.1. Compilation of the Singapore English Newspaper (SEN) Corpus

The SEN corpus comprises texts from selected issues of the Singapore national English newspaper *The Straits Times* (ST for short) published in 1993, 2005 and 2016. The texts are all in electronic form and retrieved from the ST database available on the LexisNexis Academic platform. Only articles that have been written locally have been included. *The Straits Times* was selected for this study because it is the main national broadsheet newspaper and therefore more authoritative and influential than the other local English newspapers.

Ideally, a diachronic study of language changes in SgE should begin at least from the time when Singapore became independent

from her colonial rulers, which is the mid-1960s. Practically, however, obtaining access to early texts is currently too difficult a task to undertake. Some of the challenges are detailed in Hoffmann (2018) on the practical and methodological issues involved in compiling the corpus of historical SgE.

LexisNexis Academic provides free access to fully digital text articles from mid-1992 onwards. So this study may be said to be a "short-term diachronic comparable" corpus-based one, to borrow the terminology from Leech et al. (2009, p. 44), covering a short period of 23 years. The first sampling point for the present study is 1993 because that is the earliest year when daily issues for the entire year can be obtained online. The most recent sampling point available for the purpose of this study is 2016. Right in the middle between the range of 23 years between these two points are 2004 and 2005. The latter was chosen to facilitate comparison between SEN and the SiBol corpus[3], which also uses 1993 and 2005 as its two main sampling points, before the more recently added 2013 data.

Although the number of issues in each sampled year is kept at 52 (one issue for each week of the year), the amount of text expands from year to year, so the number of words increases as well. The size of the SEN corpus is as follows:

Table 1. Number of tokens in the SEN corpus

SEN Subcorpus	1993	2005	2015
Number of tokens	1,041,910	1,444,149	1,683,901

Only normalised frequencies (per million words) rather than raw frequencies are used in the analysis of the results. Statistical significance is calculated in terms of Log Likelihood using the Excel

---

3   The Siena-Bologna Modern Diachronic Corpus (the SiBol Corpus) when introduced by Partington in 2010 (Partington, 2010) comprised only SiBol 93 and SiBol 05, containing all the articles published by the three main UK broadsheet newspapers in 1993 and 2005. It was subsequently expanded to include newspapers in 2010 and 2013 (Department of Modern Languages, Literature and Culture, 2016).

spreadsheet designed by Rayson[4].

3.2. The Siena-Bologna English Broadsheet Newspaper Corpus

In order to compare recent developments in SgE with developments in the 'source' or 'parent' variety, BrE, over the same period, data from the SiBol Corpus was also retrieved and analyzed. Only data from the Guardian was used, forming SiBol-Guardian subcorpora corresponding to the years sampled in SEN, except for 2016. The Guardian was selected for its similarity to *The Straits Times* in terms of it being a quality national broadsheet in Britain. The SiBol corpus is extensive, consisting of the entire output of the newspaper in each year, so the size of each subcorpus is many times larger than the corresponding subcorpus in SEN (see Table 2). The SiBol corpus was edited and made available through the Sketch Engine interface, so the corpus tool used in this study is Sketch Engine (Kilgarriff et al., 2014).

Table 2. Number of tokens in the SiBol-Guardian corpus

SiBol-G Subcorpus	1993	2005	2013
Number of tokens	29,674,745	45,088,742	34,364,936

3.3. Relativization in SEN

The dataset for this section was obtained firstly by identifying all instances of the main relativizers in standard English – *that, which, who, whom* and *whose*. In Sketch Engine, this involved entering the tag [WDT (*wh*-determiner)] to retrieve the occurrences of *which* and *that* as relative pronouns, [WP (*wh*-pronoun)] for *who* and *whom*, and [WPZ (*wh*- possessive pronoun)] for *whose*. Other relativizers, such as relative adverbs *why, where* or *when*, were excluded because of their low frequencies. Unfortunately, zero relativizers had to be excluded for practical reasons – it was difficult to locate, let alone extract tokens of their occurrence. RRC and NRRC were

---

[4]   http://ucrel.lancs.ac.uk/people/paul/SigEff.xlsx (latest version, 4 July 2016)

distinguished based on the presence or absence of the comma, dash or parenthesis preceding the relative pronoun, as shown in (2) and (3). This method of "punctuation-separation" (p. 602) was also used by Biber et al. (1999) to distinguish RRC from NRRC.

Another type of relative clause that was downloaded as a separate set is that containing a pied-piping construction, where the relativizer is preceded immediately by a preposition. These are excluded from the dataset examined in this paper.

A random sample of the tokens of relative clauses (20% from SEN and 2% from SiBol-Guardian) was then manually edited to exclude those that are irrelevant or incorrectly tagged, for instance interrogative *wh*-pronouns, *wh-* determiners, *that* as subordinating conjunction or demonstrative pronoun. As with related previous studies, sentential relative clauses such as (5) and (6), where the antecedent is not a noun phrase but a sentence, are also excluded from the analysis.

(5) But the underlying trend for the biomedical industry appears strong, ***which*** could provide a counter-balance to slowing growth in the electronics industry, said [...] [ST 2005-01-27]

(6) One rabbit had bite wounds and blood oozing out of its mouth, ***the cause of which*** is unclear. [ST 2016-04-26]

Frequencies are normalized to occurrences per million words (pmw) to facilitate comparison across the subcorpora of different sizes. However, where the focus is on the variants *which* and *that* in RRC, another approach is used which considers frequency changes in each variant as a percentage of the total number of RRC. According to Aarts et al. (2013), this measure is more accurate when studying changes in variation where there is a choice between the variants as it reduces the probability that the changes are caused by confounding factors. For this reason, 'knock-out' contexts such as relative clauses with pied-piping were excluded from the frequency count of RRC since they only occur with *which* and not with *that* as a relativizer.

It could be argued that since the choice of relativizer in RRC is not binary but includes the zero-relative, as illustrated in (3), this investigation on the variation will have limited value unless the zero-relative is included. However, Leech and Smith (2006) have shown on a limited set of data from the Brown quartet of corpora that the zero relativizer has appeared to remain stable diachronically, and therefore has little effect on changes in the *which-that* alternation. Furthermore, a decade later, Hinrichs et al. (2015) confirm this finding that the zero relativizer is largely unaffected by the changes in the *which-that* variation. Based on the assumption that this neutral role of the zero relativizer remains unchanged for the following two decades, the results of the current study where only *that* and *which* RRC are examined can be taken to be relevant.

The tokens were individually inspected for animacy of the antecedent, to determine if animacy could be a contributing factor to the choice in the *which* vs *that* variation. Unlike most, if not all, of the previous studies on relativizers, the current study uses the coding scheme devised by Zaenen et al. (2004, as cited in Rosenberg, 2014) for animacy.

4. Research Findings

4.1. Overall Distribution of Relative Clauses

The overall distribution of relative clauses introduced by the different relativizers *which*, *who*, *whose*, *whom* and *that* in the SEN corpus is shown in Figure 1.
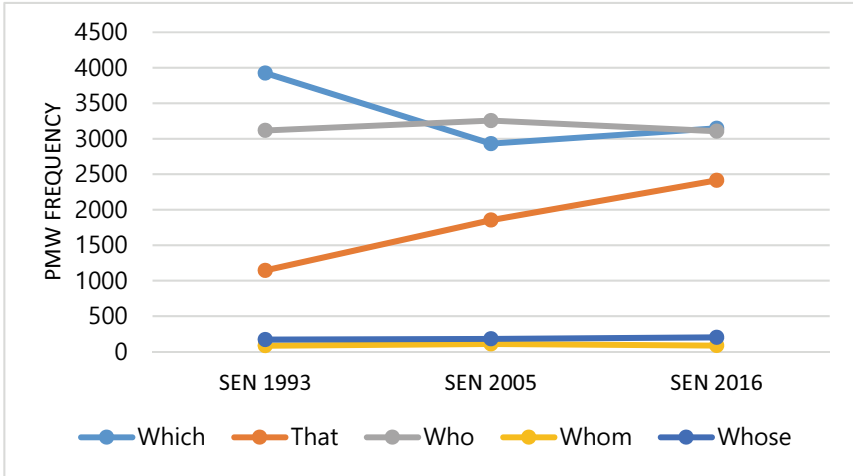
Figure 1. Distribution of Relative Pronouns by Relativizer in SEN

From Figure 1, we note that the *which* relativizer in SEN is declining in frequency over the entire sampling period while the *that* relativizer is increasing steadily. There is a small bend at 2005 where the *which* relative clauses stopped decreasing but even increased a little, though not significantly, after 2005. The relativizer *who* was less frequently used than *which* in 1993 but overtook *which* in 2005 and the two came close in 2016. This seems to be consistent with Biber et al.'s (1999) finding that *which* is more common than *that* in news writing, and *who* is even more common than *which* or *that*, which is true of SEN in 2005. As Biber et al. (1999) explain, this is not unexpected since news is generally focused on human activity. However, the result is different from Gut and Coronel's (2014) finding that *whose* is the least frequently occurring relative marker in spoken and written SgE. As can be seen in Figure 1, *whose* is more than twice as common as *whom* in SEN, although the magnitude is small to begin with.

A similar plot in Figure 2 shows the normalised pmw frequencies of the five relativizers in the SiBol-Guardian corpus. The axes for Figure 1 and Figure 2 use the same scale to facilitate easier comparison between the corpora.
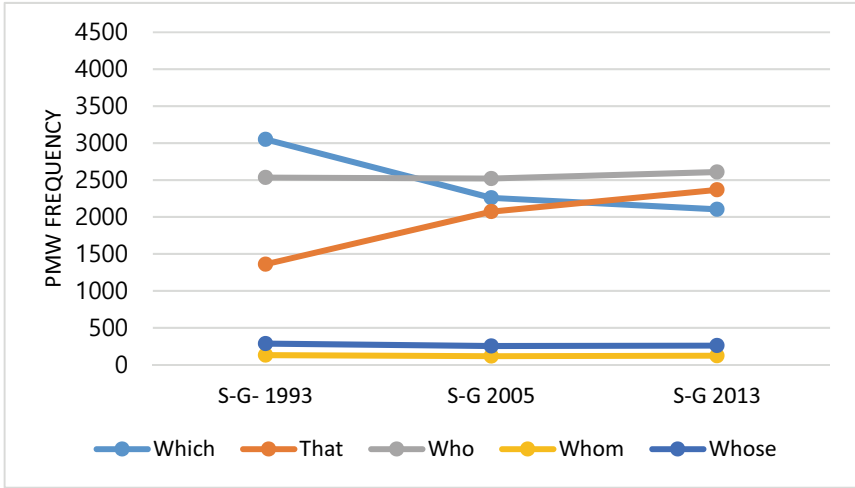
Figure 2. Distribution of Relative Pronouns by Relativizer in SiBol-
        Guardian

Comparing Figure 1 with Figure 2, we notice striking similarities and differences. Firstly, we see the frequency of *which* and *that* relative clauses changing in opposite directions, with *which* declining in contrast to *that* increasing. Both changes seem to be occurring at a similar pace. This shift is most interesting because it signals a continuation of the dramatic shifts occurring in the use of these two relative markers from 1961 to 1991 observed by Leech et al (2009), among other scholars. Secondly, we note the frequency of *that* has surpassed that of *which* in the British corpus, a phenomenon that has been observed in American English and Australian English in 1999 but has yet to occur in SEN. Though the pattern of change for each relative marker is the same across the two corpora, there is a higher frequency of *which* and *who* relative clauses in SEN than SiBol-Guardian. Thirdly, as seen in SEN, the relativizer *whose* occurs more often than *whom* in SiBol-Guardian, even though there is little change in the use of these two relative markers in both corpora across the sampling periods.

4.2. Restrictive and Non-restrictive Relative Clauses

The normalised frequencies of RRC and NNRC in the two corpora

are compared in Figure 3. All relative clauses introduced by *that* are restrictive. Those introduced by the *wh*-relativizers include both RRC and NNRC.
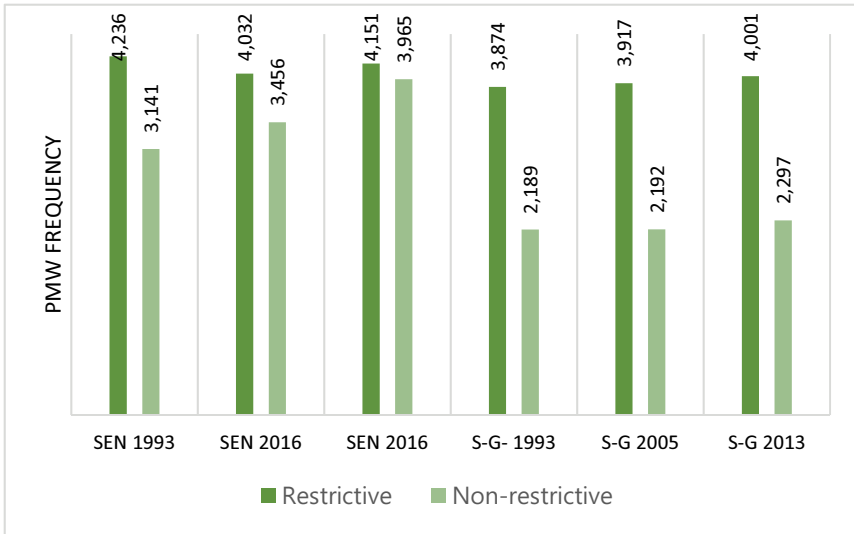

Figure 3. Normalised Frequencies (pmw) of Restrictive and Non-restrictive Clauses

The most striking difference between the two corpora is the higher (and significantly increasing) frequency of NRCC in SEN as compared to SiBol-Guardian corpus. Interestingly, despite some of the extreme changes seen in the choice of relativizers, the relative proportion of RRC to NRRC has hardly changed from 'about 30%' of the total, as reported in Leech et al. (2009) and Biber et al. (1999), whose data were collected in the 1990s. The ratio in SEN has changed slightly, but the more striking observation is that the relative proportion of NRRC is a significant 10% above that in the SiBoL-Guardian corpus.

It is only to be expected to find that NRCC are common in news writing. According to Biber et al. (1999), they occur more often in news than in other genres, and are used to add information that may be newsworthy, though incidental and not necessarily directly related to the news story. Sometimes, the additional information

may serve as background that helps the reader in the interpretation of the main story. Examples (7) and (8) illustrate the two main uses of NRCC:

(7) The additional flight, **which** will leave Singapore at 5.35pm and return from the Thai capital at 8.05pm, will add 2,000 seats a week to the route. [ST 2016-01-10]
The relative clause adds information that is not essential for the purpose of informing the reader of the additional seats, but may be of interest to some readers.

(8) Her neighbour in Isleworth was the scientist and explorer Sir Joseph Banks, **who** had sailed to Australia with Captain James Cook, and **who** knew the Raper family. [SiBol-G 2005-011-17]
The relative clause is not necessary in identifying the referent, whose full name and vocation have already been provided, but adds interesting background information about him.

However, Biber et al. (2009) have found that postmodifiers, including relative clauses, are generally scarce in conversation. Therefore, it is somewhat puzzling to find that such NRCC are becoming more common in news writing, indicating a shift opposite to colloquialization, away from similarity to speech. Besides the significant jumps in SEN, the increase in NRCC in SiBol-G from 2005 to 2013 is statistically significant too (Log Likelihood value of 97 at p<0.001). Perhaps this is the case of densification being more dominant than colloquialization as a motivating factor for change. Biber (2003) argues that "the 'informational explosion has resulted in pressure to communicate information as efficiently and economically as possible, resulting in compressed styles that depend heavily on tightly integrated noun-phrase constructions" (p. 170), among which are NRCC.

A clearer picture of the distribution of RCC versus NRCC emerges as we examine it according to the individual relativizers *which* and *who*, as shown in Figure 4 and Figure 5. Only these two relativizers are discussed here because they are the two most frequently

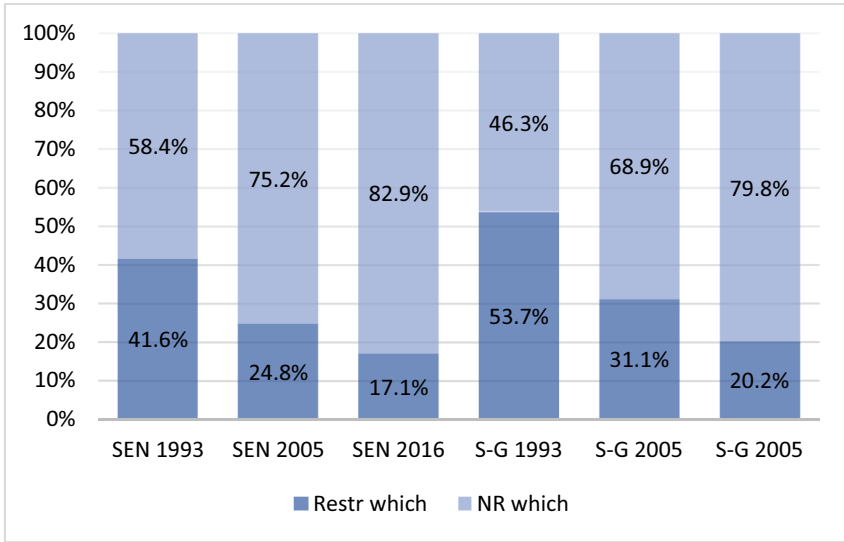occurring, and *that* is only used for NRCC in this dataset.

Figure 4. Frequency of Restrictive vs. Non-restrictive *which* Relative
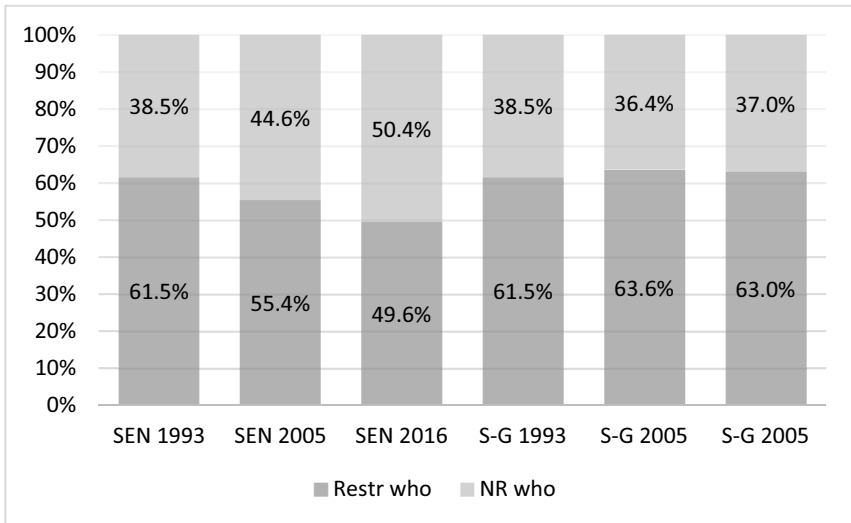Clauses by Subcorpus

Figure 5. Frequency of Restrictive vs. Non-restrictive *who* Relative
Clauses by Subcorpus

In Figure 4, the declining ratio of restrictive to non-restrictive *which* relative clauses is starkly evident in both corpora. In fact, in the SiBol-Guardian corpus, we note that the relative proportion has changed from a majority of over 53% restrictive in 1995 to about 31% in 2005, and just 20% in 2013. Yet the ratio in Figure 3 comparing the overall frequency of all five types of relative clauses remains almost unchanged, with a variance smaller than 5% in the span of 18 years. This suggests that *which* as a relative marker is being replaced by one or more alternatives, the most likely candidate being *that*. Figure 6 below demonstrates how the frequency of *that* relative clauses has risen significantly in both corpora, possibly making up for the decline in *which* relative clauses, since *who* clauses do not seem to have changed much.
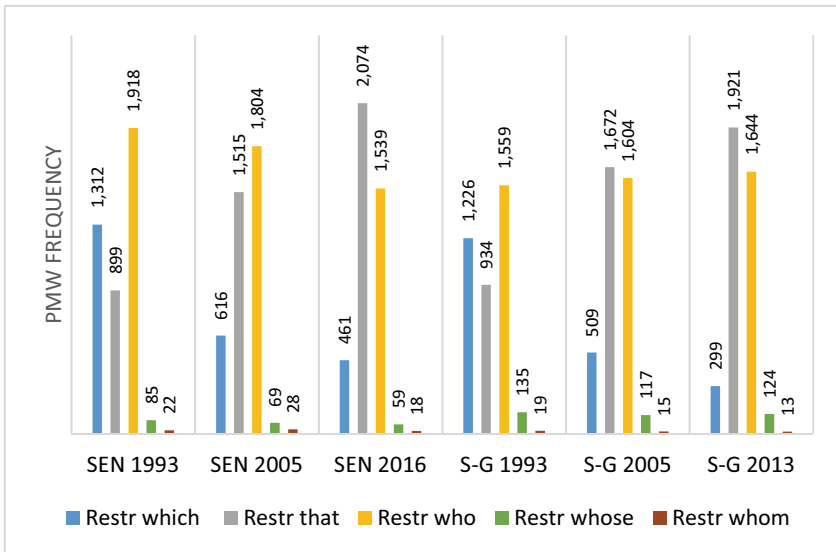

Figure 6. Comparing Frequency of Restrictive Relative Clauses by Relativizer

Turning to the alternation between *which* and *that* in RCC, which has been the focus of many previous studies on relative clauses, we see the proportionate variation in Figure 7 below.
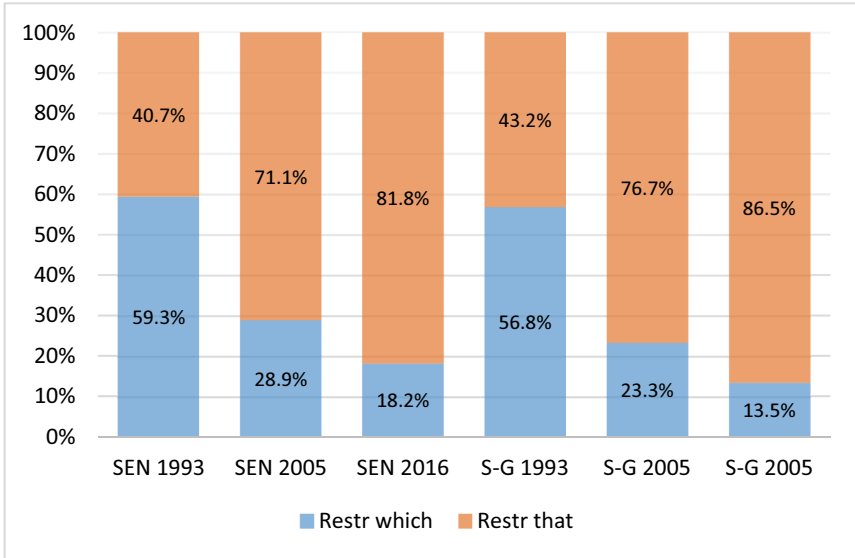
Figure 7. Variation of *which* and *that* in Restrictive Relative Clauses

The patterns in the two corpora are remarkably similar, demonstrating the clearly complementary trajectories of the *which* and *that* relativizers, continuing the diachronic trend that was evident in the Brown family of corpora. The results obtained in the current study partially mirror the findings of Gut and Coronel (2012) from ICE (SIN) that show how *which* is more frequent than *that* in RRC in the ICE (SIN) data of the early 1990s. The drastic change in the usage of relative clauses in the short span of just over 10 years is quite remarkable.

The results from the two corpora suggest that the American prescriptive tradition of the *that*-rule proscribing the use of *which* for RRC has indeed spread to other varieties over the years, perhaps catching on because of the 'convenience' of the distinction in the use of *that* and *which* when writers are faced with a decision to choose one. Hinrichs et al. (2015) conclude that the *that*-rule was firmly planted within British and American English editorial practice by the early 1990s. Since then, this tradition could have spread to other varieties, including the postcolonial Englishes, helped by the pervasiveness of online grammar checkers. It is noted from Figure

6 that there is a higher frequency of *which* relatives in SEN as compared to SiBol-Guardian in all three sampling periods, which could point to SgE trailing behind BrE, which in turn lags behind American English, which is leading this change.

However, this American-led '*which*-hunt' is not the only factor that favours the use of *that* in RRC. We have seen from other evidence that the formality of the context also has a bearing on the variation – *that* is used in more informal contexts and in speech, while *which* is preferred in writing, especially the more formal genres. So this is a case where there is a "rare show of unity" by prescriptivism and colloquialization, which both motivate the change in the same direction (Hinrichs et al., 2015, p. 831) towards a more informal form.

4.3. Animacy of the Antecedent

Tables 3 and 4 show the distribution of RRC by animacy of antecedents for *which* and *that* relativizers in the two corpora.

Table 3. Animacy of Antecedents by Percentage of Total Frequency of Restrictive *which* Relative Clauses

Restrictive *which*	SEN 1993	SEN 2005	SEN 2016	SiBol-G 1993	SiBol-G 2005	SiBol-G 2013
Human	0%	1%	0%	0%	0%	0%
Animal	0%	1%	1%	1%	1%	1%
Collective	13%	12%	17%	9%	9%	15%
Inanimate	87%	87%	81%	90%	90%	84%

Table 4. Animacy of Antecedents by Percentage of Total Frequency of Restrictive *that* Relative Clauses

Restrictive *that*	SEN 1993	SEN 2005	SEN 2016	SiBol-G 1993	SiBol-G 2005	SiBol-G 2013
Human	1%	1%	1%	1%	1%	3%
Animal	2%	1%	2%	2%	1%	1%
Collective	11%	14%	14%	11%	10%	12%
Inanimate	87%	85%	83%	85%	88%	84%

The results shown in Tables 3 and 4 appear to be in agreement with previous studies on the use of *which* and *that* being almost exclusively used with non-human antecedents in both varieties of English. The similarity between the two varieties in Tables 4 indicates that SgE is not so different from BrE in not using *that* for human antecedents in relative clauses, unlike what previous studies on earlier data from the individual varieties (Biber et al., 1999; Newbrook, 2003; Gut & Coronel, 2012) have shown. A likely reason for *that* being used with human antecedents more commonly in conversation than in written prose could be a matter of formality, and as the data shows, colloquialization has not affected this situation although it is believed to be the driving force for *that* increasingly being preferred over *which* in relative clauses in written English.

Using the one-to-one mapping of form to function, it may be surmised that since the *who* relativizer exclusively takes a human antecedent, there is no motivation for a writer to choose *that* in the same context. As noted earlier from Figures 5 and 6, *who* relative clauses have not changed in frequency over the years within the sampling period, and are unlikely to have been affected by the ascendancy of *that* in relative clauses.

It may be posited that prescriptivism could play a part in *that* being rarely used for animate antecedents in SEN, if the following rule in the *Straits Times* handbook was adhered to and communicated to journalists and sub-editors through the years: "Use *who* and *whom* to refer to people, and to animals with names … . Use *which* and *that* to refer to inanimate objects, and to animals without names." Lee (1997). Back in the 18th century, according to Ball (1996), "to use *that* for subjects, all other things being equal, is to mark the referent as nonhuman; to mark a human (adult) referent as nonhuman is improper (Dryden, 1691), impolite (Fowler, 1927, p. 716), or worse" (p. 250), and it was this social sentiment of that era that sent *who* rising at the expense of *that*. However, a check with several current online grammar guides originating from the US and UK does not reveal any site that prescribes this rule strongly, although there was some mention of how *that* is used more

informally for human antecedents.

There are only a few tokens of *that* being used with a human antecedent in the SEN and SiBol-Guardian corpora, and most are either part of quoted speech, as in (9), or/and certain constructions that conventionally go with *that* rather than *who*, whether the antecedent is human or otherwise (see examples (10) to (12)).

(9) "Google demands employees **that** are 90th percentile material, so what's with the 50th percentile compensation?" he asked … . [SiBol-G 2005-02-10]

(10) "It's about an innocent character who is innocently defeating someone **that** is trying to take advantage of him." [ST 2005-03-14]
   – Pronominal antecedent *someone*

(11) Everybody knows, of course, that Bobby Jones was also *the finest amateur* **that** ever drew breath, and the game's greatest sportsman. [SiBol-G 1993-08-18]
   – Postmodifying a noun phrase that contains a superlative adjective

(12) A *baby* **that** is older than 24 weeks can survive, but *one* **that** is less than 24 weeks cannot survive without medical support. If the baby survives, *it* would be disadvantaged by the short gestation. [ST 2016-09-03]
   – A "baby" is not perceived as a fully formed human, which explains the anaphoric "it" used in the following sentence.

It is worth noting from Tables 3 and 4 that a significant proportion (about 10% or more) of the *which* and *that* relatives are used with collective antecedents, a class of nouns that lie in the ambiguous space between what is perceived as human and non-human. The change in the proportion of *which* and *that* relatives for collective antecedents may not be considerable, but it appears to be creeping into the share of the inanimate antecedents. Examples of

such antecedents in the corpus data are *Islamic groups* and *a Japanese company*. These antecedents are collective groups of humans often attributed human actions, such as "launched" and "owns". Yet they are used with *that* or *which* relativizers instead of with *who*. Perhaps the human vs non-human distinction that is used in previous studies should be refined to include this division of collective nouns so as to better understand the effect of semantic properties of the antecedent on relativizer variation.

Another relativizer that can also used with both human and non-human antecedents is *whose*. Table 5 shows the distribution of antecedents in *whose* relative clauses (both RRC and NRRC) by animacy.

Table 5. Animacy of Antecedents by Percentage of Total Frequency of Restrictive "Whose" Relative Clauses

Animacy	SEN 1993	SEN 2016	SEN 2016	SiBol-G 1993	SiBol-G 2005	SiBol-G 2013
Human	62%	65%	71%	70%	74%	70%
Animal	0%	0%	0%	1%	0%	2%
Collective	31%	23%	23%	24%	16%	20%
inanimate	7%	11%	6%	5%	9%	7%

Unlike the case of *that* and *which*, up to 30% of the *whose* relative clauses follow non-human antecedents, as shown in the following examples:

(13) For American multinational AT&T, **which** enjoys tax incentives under the pioneer status and **whose** exports would be zero-rated under the GST, the concern is not with the corporate tax cut but the GST's inflationary impact on wages. [ST 1993-02-13]

(14) "Ask yourself - what was the last consumer app **whose** popularity depended on being available for Windows?" [SiBol-G 2013-12-19]

The majority of the non-human antecedents fall under the collective category though a healthy proportion is used with inanimate nouns. Example (13) was singled out to show that the antecedent "American multinational AT&T" was treated as non-human by the use of the relativizer *which* in the first relative clause postmodifying it. Apart from a slight increase in the use of *whose* with human antecedents in SEN from 1991 to 2016, there is no other discernible trend across the sampling periods. Before any further conclusions can be made about other changes in the use of *whose* relative clauses, a finer analysis of the data is required, which has to be left for future work.

5. Discussion

5.1 Summary of Changes Identified

This paper aimed to identify and compare the use of relative clauses in SEN and SiBol-Guardian corpora – the distribution of the five most common relativizers *which*, *who*, *whose*, *whom* and *that*, the distribution of RCC vs NRCC, and the effect of animacy of the antecedent on the choice of relativizers. There is a scarcity of previous diachronic research in this aspect of SgE, so the results are highly informative.

Firstly, as predicted by previous studies, the *which* relativizers continue to decline in frequency of use, gradually being replaced by the *that* relativizer. Relative clauses introduced by *who* remains fairly stable. A similar trend is seen in SEN and SiBol-Guardian, although the complementary movement of *which* and *that* seems more advanced in the latter in such a way that in SiBol-G 2013, *that* relative clauses have overtaken the *which* relative clauses. Further analysis shows that in RRC, the complementary movement has been so rapid that although *which* relative clauses were more common than *that* clauses in 1993, *that* overtook *which* in 2016 (2013 for SiBol-G) for both the corpora.

The relativizers *which* and *that* are almost exclusively used for non-human nouns in both corpora throughout the sampling periods.

However, more antecedents in SEN are collective nouns, and there is a minor increase in this category of antecedents in *which* relative clauses when comparing 2016 to 1993, in addition to the majority of inanimate antecedents. While *who* and *whom* predominantly follow only human antecedents, about 30-35% of *whose* relative clauses are used with non-human antecedents, including collective and inanimate ones.

5.2 Possible Factors Driving the Changes

Besides reporting on the recent changes in Singapore news writing with regard to relative clauses, this paper has also considered some of the main factors that may have been implicated in these developments. Firstly, changes such as the rapid rise of *that* replacing *which* in RRC point to colloquialization. These shifts are movements towards features more commonly found in speech, and mirror developments in verbs and verb groups in other diachronic studies.

The second factor likely to be involved is densification or economisation, as discussed by many scholars, most notably Biber (2003) but also Leech et al. (2009) and Hinrichs and Szmrecsanyi (2007). This factor might explain the increase in NRRC in SEN, and to a lesser extent, SiBol-Guardian.

A third factor that has been discussed is prescriptivism, which has been cited by other researchers as being responsible for the accelerated replacement of *which* by *that* in RCC. Also known as the *that*-rule or *which*-hunt, the development is reported as "a case of AmE-led colloquialization with institutional backing" or "colloquialization from above" (Hinrichs et al., 2015, p. 831).

With these driving forces acting on news writing, it is no wonder that it is often considered the most "agile" genre of all (Hundt & Mair, 1999). However, it is unlikely for these trends to continue unchecked in broadsheet news writing because such news sources, unlike the tabloids, are expected to maintain a level of formality and prestige above that of conversation. As noted in the *Straits Times* Handbook, "we want to be economical with words but not offend grammarians" (Lee, 1997, p. 159).

6. Conclusion

Having answered the research questions and discussed the possible motivations for the changes identified in the two corpora, it is fitting that the implications for endonormative stabilization of SgE be considered at this point.

An important observation that is made from the findings is that regarding the use of relative clauses, there appears to be little divergence in SgE from BrE. As BrE shows dramatic changes in areas such as the rapid replacement of relativizer *which* with *that* over the last 25 years, SgE appears to be lagging just a step behind, "striving to catch up" with the changes in BrE, to use a phrase by Collins, Yao and Borlongan et al. (2014) in describing modality in Philippine English vis-à-vis American English (p. 68). The apparent alignment of SgE with the source variety could suggest that the linguistic orientation of SgE is still largely exonormative. However, it has to be noted that both varieties are subject to the driving global forces of change described in 5.2 above, resulting in a similar trajectory of changes, like the "diachronic drift pulling along BrE and AmE in broadly the same direction" mentioned by Mair (2015, p. 7). Any conservatism or lag observed in SgE may not be due to vestigial allegiance to BrE but a result of the prevailing status of SgE in broadsheet news writing as a model of 'standard' written English where the effect of colloquialization is limited. For instance, the SEN data shows a divergence from the SiBol-G data in the higher and increasing frequency of NRRC (Figure 3) compared to RRC over the sampled period, indicating that densification has a stronger impact on SgE than the opposing pull of colloquialization for news writing. In other areas, for instance, modality and multi-word verbs, SEN is found to display characteristics that are distinctly different from SiBol-G. The point to be made here is that conclusions on the evolutionary status of SgE, or any postcolonial variety, cannot be drawn simply from the similarity or differences in patterns of change of a specific grammar feature displayed between the source and postcolonial variety. Other factors to be considered include the genre and the factors driving the changes causing global diachronic drifts.

References

Aarts, B., Close, J., & Wallis, S. (2013). Choices over time: Methodological issues in investigating current change. In Aarts, B., Close, J., & Wallis, S. (Eds.), *The Verb Phrase in English: Investigating Recent Language Change with Corpora* (pp. 14-45). Cambridge: Cambridge University Press.

Alsagoff, L., & Ho, C. L. (1998). The relative clause in colloquial Singapore English. *World Englishes, 17*(2), 127-138.

Ball, C. N. (1996). A diachronic study of relative markers in spoken and written English. *Language Variation and Change, 8*(2), 227-258.

Biber, D. (2003). Compressed noun-phrase structures in newspaper discourse. In Aitchison, J., & Lewis, D. M. (Eds.), *New Media Language* (pp. 169-181). London/New York: Routledge.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Collins, P., Borlongan, A. M., & Yao, X. (2014). Modality in Philippine English: A diachronic study. *Journal of English Linguistics, 42*(1), 68-88.

Collins, P., Yao, X., & Borlongan, A. (2014). Relative clauses in Philippine English: A diachronic perspective. In Vandelanotte, L., Davidse, K., Gentens. C., & Kimps, D. (Eds.), *Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora* (pp. 123-146). Amsterdam: Brill-Rodopi.

Deterding, D. (2007). *Singapore English.* Edinburgh: Edinburgh University Press.

Grafmiller, J., Szmrecsanyi, B., & Hinrichs, L. (2018). Restricting the restrictive relativizer. *Corpus Linguistics and Linguistic Theory, 14*(2), 309-355.

Gries, S. T., Bernaisch, T., & Heller, B. (2018). A corpus-linguistic account of the history of the genitive alternation in Singapore English. In Deshors, S. C. (Ed.), *Modeling World Englishes: Assessing the Interplay of Emancipation and Globalization of ESL Varieties* (pp. 245-279). Amsterdam: John Benjamins.

Gut, U., & Coronel, L. (2012). Relatives worldwide. In Hundt, M., &

Gut, U. (Eds.), *Mapping Unity and Diversity World-wide: Corpus-based Studies of New Englishes* (pp. 215-241). Amsterdam: John Benjamins.

Hinrichs, L., & Szmrecsanyi, B. (2007). Recent changes in the function and frequency of standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language & Linguistics, 11*(3), 437-474.

Hinrichs, L., Szmrecsanyi, B., & Bohmann, A. (2015). Which-hunting and the Standard English relative clause. *Language, 91*(4), 806-836.

Hoffmann, S. (2018). I would like to request for your attention. In Kaunisto, M., Höglund, M., & Rickman, P. (Eds.), *Changing Structures: Studies in Constructions and Complementation* (pp. 171-196). Amsterdam: John Benjamins.

Hundt, M., & Mair, C. (1999). "Agile" and "uptight" genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics, 4*(2), 221-242.

Hundt, M., & Szmrecsanyi, B. (2012). Animacy in early New Zealand English. *English World-Wide, 33*(3), 241-263.

Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfeit, J., Rychly, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*, 7-36.

Lee, S. L. (1997). *Handbook on the Use of English*. Singapore: Singapore Press Holdings.

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.

Leech, G., & Smith, N. (2006). Recent grammatical change in written English 1961-1992: Some preliminary findings of a comparison of American with British English. In Kehoe, A., & Renouf, A. (Eds.), *The Changing Face of Corpus Linguistics* (pp. 185-204). Amsterdam: Brill-Rodopi.

Mair, C. (2015). Cross-variety diachronic drifts and ephemeral regional contrasts: An analysis of modality in the extended Brown family of corpora and what it can tell us about the New Englishes. In Collins, P. (Ed.), *Grammatical Change in English*

World-wide (pp. 119-146). Amsterdam: John Benjamins.

Mukherjee, J., & Gries, S. T. (2009). Collostructional nativisation in new Englishes: Verb-construction associations in the *International Corpus of English. English World-Wide, 30*(1), 27-51.

Mukherjee, J., & Schilk, M. (2012). Exploring variation and change in new Englishes: Looking into the International Corpus of English (ICE) and beyond. In Nevalainen, T., & Traugott, E. C. (Eds.), *The Oxford Handbook of the History of English* (pp. 189-199). Oxford: Oxford University Press.

Newbrook, M. (2003). Features of the relative clause in Singapore English. In Deterding, D., Low E. L., & Brown, A. (Eds.), *English in Singapore: Research on Grammar* (pp. 67-76). Singapore: McGraw Hill.

Partington, A. (2010). Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: An overview of the project. *Corpora, 5*(2), 83-108.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics, 13*(4), 519-549.

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

Schneider, E. W. (2010). Developmental patterns of English. In Kirkpatrick, A. (Ed.)*, The Routledge Handbook of World Englishes* (pp. 372-384). London: Routledge.

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

Suárez-Gómez, C. (2014). Relative clauses in Southeast Asian Englishes. *Journal of English Linguistics, 42*(3), 245-268.

THE AUTHOR

K. M. Lee is a Lecturer at the Centre for English Language Communication, National University of Singapore. She currently teaches critical thinking and communication to students from the School of Computing. Her research interests include corpus linguistics, academic writing, and scholarship in teaching and learning for Higher Education. She is currently

writing a PhD thesis on a diachronic corpus-based study on English news writing in Singapore.

THE AUTHOR'S ADDRESS

First and Corresponding Author
Kit Mun Lee
Lecturer
Centre for English Language Communication
National University of Singapore
10 Architecture Drive, #03-13, 117511, SINGAPORE
E-mail: elclkm@nus.edu.sg