A Corpus Analysis of the *Were*-Subjunctive in English:
A Comparative Study of Four Discourse Genres

Min Young Song
(University of California, Los Angeles)


This corpus-based study investigates the usage patterns of the subjunctive *were* and its alternative *was* in four discourse genres (an academic writing, an informal personal essay, an academic lecture, and an informal conversation) and with four subject types (1st person singular, 3rd person singular pronouns, 3rd person singular proper/common nouns, and existential *there* followed by a singular noun). The findings show the overall usage patterns of *were* and *was* and identify several genre-specific structures and lexical bundles. This study demonstrates that despite the uncertain status of the subjunctive in modern English, the *were*-subjunctive is frequently used in some discourse genres and lexical bundles, and therefore it is necessary to provide more detailed descriptions of its usage patterns in the literature. Moreover, it is suggested that more attention should be paid to local patterns of the English subjunctive occurring in specific linguistic and extralinguistic contexts as well as its general patterns. Discussed are some pedagogical implications of this study.

I. INTRODUCTION

For most of the twentieth century, many linguists claimed that the subjunctive had died out in English (Harsh, 1968; Kruisinga & Erades, 1941; Palmer, 1974) or that its use was in decline (Jespersen, 1924; Leech, 1971) and its few remaining forms were being replaced by other constructions (Fowler, 1965; Quirk, Greenbaum, Leech, & Svartvik, 1972, 1985). Quirk, et al. (1985, p. 155) suggest that the subjunctive “is generally an optional and stylistically marked variant of other constructions”. Furthermore, most recent grammatical research has tended to ignore the subjunctive or give only a brief comment on it. For example, Biber, Johansson, Leech, Conrad, and Finegan (1999), a seminal corpus-based grammar of English, does not provide a substantial description of the subjunctive.
Some researchers, however, note the importance of the subjunctive in English. Turner (1980, p. 271) claims that “it is a mistaken exaggeration to conclude that the subjunctive is on the verge of dying out completely”, and Quirk et al. (1985, p. 155) state that “it is not so unimportant as is sometimes suggested”. As a matter of fact, work on the subjunctive has grown over the past twenty years, especially with regard to the mandative subjunctive\(^1\) (e.g., Fillbrandt, 2006; Givón, 1993, 1995; Turner, 1980), which may be motivated by the supposed higher frequency of the mandative subjunctive in American as opposed to British English (Johansson & Norheim, 1988; Quirk et al., 1985). Quite recently, Frazier (2003) demonstrates that would-clauses in counterfactual/hypothetical environments occur more often separated from or entirely without corresponding if-clauses.

The uncertain status of the subjunctive in modern English is also true of the were-subjunctive, the focus of the current study. Quirk et al. (1985, pp. 1012-1014) describe the were-subjunctive as hypothetical or unreal in meaning, being used in adverbial clauses introduced by such conjunctions as if, as if, as though, and in nominal clauses after verbs like wish and suppose. Whereas the indicative form shows a contrast between the two past tense forms of BE, i.e., was (1\(^{st}\) and 3\(^{rd}\) person singular) and were (plural and 2\(^{nd}\) person singular), the subjunctive is invariably were, and thus breaks the agreement rule of the indicative verb BE in the 1\(^{st}\) and 3\(^{rd}\) person singular of the past tense. They also comment that the subjunctive were is replaced by the hypothetical indicative form was in less formal styles and the former is nowadays a less usual alternative to the latter (ibid., p. 1013).

Such brief descriptions of the were-subjunctive, however, are too simplistic, especially considering that several linguistic and extralinguistic or situational criteria may be involved in selecting either the subjunctive were or its alternative was in real discourse. Quirk et al. (1985) briefly mention some of these linguistic criteria. In some fixed phrases, such as if I were you or as it were, it is customary to use were instead of was. In addition, the present indicative is a possible alternative after as if and as though or after imperatives suppose and imagine. With regard to extralinguistic criteria, however, Quirk et al. merely mention that were is replaced by was ‘in less formal style’. They do not provide any detailed descriptions of the ‘less formal style’, nor do they mention any other variables, such as discourse genres/registers or sociolinguistic contexts. Such descriptions of the were-subjunctive have rarely been accompanied by authentic data supporting them empirically. Even Biber et al. (1999) does not provide any empirical data showing the actual usage patterns of the were-subjunctive.

Likewise, many ESL grammar books (e.g., Celce-Murcia & Hilles, 1988; Celce-Murcia & Larsen-Freeman, 1999; Riggenbach & Samuda, 1997) simply explain the traditional prescriptive rules for the were-subjunctive, rarely commenting on the possibility that the

\(^1\) Quirk et al. (1985) categorize the mandative subjunctive as a form of the present subjunctive used in that-clauses introduced by expressions of demand, recommendation, proposal, resolution, intention, etc.
subjunctive *were* may be replaced by *was* in real discourse. However, considering that English learners may encounter both *were* and *was* in real discourse and will need to select one form in a wide range of spoken and written discourses, it would be helpful to provide them with more detailed, practical information about the actual usage of the *were*-subjunctive – the general patterns of its usage, or some specific linguistic and extralinguistic contexts in which either *were* or *was* is preferred.

Despite the apparent usefulness of such information, only a few empirical studies have investigated the actual usage patterns of the *were*-subjunctive. For instance, Deutschmann (1998), in a corpus-based study using the British National Corpus (BNC), demonstrated that not even standard English uses the *were*-subjunctive invariably in hypothetical contexts. Hardie and McEnery (2003), combining corpus and questionnaire data, examined the distribution of the *were*-subjunctive in British rural dialects, and concluded that the question of the *were*-subjunctive’s existence in English rural dialects remains open. Ihalainen (1988) also discussed the use of *were* versus *was* in the speech of dialect speakers in the South of England, even though he did not consider the *were*-subjunctive as a possible motivation for the two forms.

In sum, as seen in most grammar books, brief descriptions of the *were*-subjunctive rarely provide any practical information with regard to its usage patterns in real discourse, and few empirical studies addressed this issue in standard English context. Considering that the *were*-subjunctive is still alive and English learners encounter its variants in real discourse, more detailed descriptions of its usage patterns could provide valuable information in terms of both theoretical and pedagogical perspectives, and only corpus-based approaches can address this issue appropriately.

II. THE STUDY

1. The Purpose of the Study

Following the supposition that the usage of the *were*-subjunctive is closely related to discourse genre/register variation, the current corpus-based study hopes 1) to find out the overall usage patterns of the *were*-subjunctive, by comparing the relative frequencies of the subjunctive *were* and its alternative form *was* as they occur with different types of subjects in four discourse genres, i.e., academic essay, informal personal essay, academic spoken, and informal conversation, and 2) to find out genre-specific linguistic features including lexical bundles regarding the usage of the *were*-subjunctive. It is hoped that the findings of this

---

2 Biber found that lexical bundles are a reliable indicator of genre/register variation (e.g. Biber, 2003; Biber & Conrad, 1999).
corpus-based study will contribute to a better understanding of the current status of the 
were-subjunctive in modern English as well as provide pedagogically valuable information.

2. The Structures Investigated

To find out the occurrences of the subjunctive were and its alternative was, several 
hypothetical structures were initially searched in the corpora listed in TABLE 1. However, 
structures [4] and [5] occurred too infrequently to be included in the study. Therefore, the 
current study focuses on structures [1], [2], and [3], i.e., adverbial clauses beginning with if ~, 
as if, and as though ~, and nominal clauses after the verb wish, including adverbial 
clauses with no overt if and subject-verb inversion (e.g. Were it a sure cure for obesity, ~).

| Table 1 |
| Main Hypothetical Structures in English |

[1] If I were/was rich, I would buy you anything you wanted.
[2] The stuffed dog barks as if it were/was/is a real one.
[3] I wish the journey were/was over.
[4] I’d rather I were/was in bed.
[5] Just suppose everyone were/was/is to give up smoking and drinking.

Extracted from Quirk et al. (1985)

It is generally accepted that “the person and number of the subject are factors that are 
known to have an effect on the form a verb takes in English” (Hardie & McEnery, 2003, p. 
216). According to Hardie and McEnery (ibid.), the corpus used in their study showed that 
while <I were> is more common than <I was>, the <she was>/<she were> distribution is 
much more even. Also, Quirk et al. (1985) state that it is more common to use were in the 
fixed phrase, <If I were you>. Thus, assuming that subject type may have an effect on the 
selection between were and was in hypothetical structures, the current study investigates 
the aforementioned three hypothetical structures in the following four contexts: first, when 
the subject of the clause which contains the subjunctive were or was is I; second, when the 
subject is a third person singular pronoun including he/she/it/this/that/one; third, when the 
subject is a third person singular proper or common noun, such as Tom or dog, etc.; and 
finally when existential-there is followed by a singular noun. Third person plural 
pro/proper/common nouns were not considered in this study, since were is not a marked 
selection for these plural subject forms. Summarized in TABLE 2 are the hypothetical 
structures and the subject types investigated in this study.
A Corpus Analysis of the *Were*-Subjunctive in English: A Comparative Study of Four Discourse Genres

### TABLE 2

<table>
<thead>
<tr>
<th>Structures</th>
<th>Types of subjects</th>
<th>Verb forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>If</td>
<td><em>I</em> (1st singular)</td>
<td></td>
</tr>
<tr>
<td>As if</td>
<td><em>HE</em> (3rd singular pronoun)</td>
<td><em>were/was</em></td>
</tr>
<tr>
<td>As though</td>
<td><em>TOM/DOG</em> (3rd singular proper/common)</td>
<td></td>
</tr>
<tr>
<td>Wish (that)</td>
<td><em>THERE</em> (~ + singular common)</td>
<td></td>
</tr>
</tbody>
</table>

### 3. The Corpora

For the present study, four different discourse genres were selected: for the written registers, academic written discourse and non-academic/informal personal essays were examined, and for the spoken registers, academic spoken discourse and non-academic/informal conversation were explored. More specifically, the academic written part of the British National Corpus (BNC) was used to represent the academic written discourse genre, while narrative personal essays written by English native speakers ranging from late teens to sixties were collected from websites, providing the informal, non-academic written discourse corpus. For this study, narrative personal essays are defined as short - generally less than 1500 words - pieces of nonacademic, miscellaneous, informal writing on a variety of personal subjects, such as special moments in life, travel experiences, relationships with others, life goals and dreams, lessons learned from experiences, and so on. Generally, no topic or subject is off-limits for this genre of writing, and writers usually wish to share something of their own personality or life with others by expressing their own experiences, ideas, and feelings. Therefore, narrative personal essays are primarily speculative and emotional rather than informative and argumentative. To investigate academic spoken discourse, the Michigan Corpus of Academic Spoken English (MICASE) is combined with the academic spoken lecture portion of the BNC, and for the informal conversation genre, the conversation part of the BNC is combined with Santa Barbara Corpus of Spoken American English (SBC), Parts I to IV. The corpora investigated for each discourse genre are summarized in TABLE 3.

### TABLE 3

<table>
<thead>
<tr>
<th>The Corpora Investigated in the Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written</strong></td>
</tr>
<tr>
<td>Formal</td>
</tr>
<tr>
<td>BNC academic writing</td>
</tr>
<tr>
<td>(Appr. 15 million words)</td>
</tr>
<tr>
<td>Informal</td>
</tr>
<tr>
<td>Narrative personal essays*</td>
</tr>
<tr>
<td>(Appr. 410,000 words)</td>
</tr>
<tr>
<td><strong>Spoken</strong></td>
</tr>
<tr>
<td>Formal</td>
</tr>
<tr>
<td>MICASE + BNC academic lectures</td>
</tr>
<tr>
<td>(Appr. 3 million words)</td>
</tr>
<tr>
<td>Informal</td>
</tr>
<tr>
<td>SBC + BNC conversation</td>
</tr>
<tr>
<td>(Appr. 4.7 million words)</td>
</tr>
</tbody>
</table>

* This is a mini corpus built for this study.
Originally, only the MICASE and the SBC were going to be used for the academic spoken genre and the informal spoken genre respectively. From the preliminary analyses of the two corpora, however, it was found that the subjunctive *were* and its alternative *was* do not occur frequently enough for this study in these two corpora. Therefore, it was decided that each corpus should be combined with its counterpart of the BNC, i.e., the academic lecture portion and the conversation portion. To check if the two combined corpora are comparable with regard to the hypothetical structures concerned, the distributions of the four hypothetical structures were investigated for each corpus as seen in TABLE 4 and TABLE 5.

**TABLE 4**

**Distributions of the Structures in MICASE & BNC Academic Lecture**

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Number of words</th>
<th>Number of tokens /10,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>MICASE</td>
<td>1,538,801</td>
<td><em>if</em> 62 <em>as if</em> .6 <em>as though</em> .2 <em>wish</em> .4</td>
</tr>
<tr>
<td>BNC Lecture</td>
<td>1,544,884</td>
<td><em>if</em> 65 <em>as if</em> .8 <em>as though</em> .2 <em>wish</em> .5</td>
</tr>
</tbody>
</table>

**TABLE 5**

**Distributions of the Structures in SBC & BNC Conversation**

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Number of words</th>
<th>Number of tokens /10,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBC</td>
<td>532,222</td>
<td><em>if</em> 19 <em>as if</em> .08 <em>as though</em> 0 <em>wish</em> .3</td>
</tr>
<tr>
<td>BNC Conversation</td>
<td>4,206,058</td>
<td><em>if</em> 41 <em>as if</em> .7 <em>as though</em> .4 <em>wish</em> .9</td>
</tr>
</tbody>
</table>

As shown in TABLE 4, it seems that the MICASE and the BNC academic lecture part can be regarded as comparable in terms of the distributions of the four hypothetical structures as well as the corpus size. TABLE 5, however, shows that the SBC is not comparable to the BNC conversation part in terms of the distributions of the four hypothetical structures. Moreover, as suggested by previous research, American English and British English might be incomparable in terms of the frequency of certain subjunctive structures (e.g. Johansson & Norheim, 1988; Quirk et al. 1985). Nevertheless, it was decided to use both corpora for the following reasons: first, both the SBC and the BNC conversation part are generally accepted as balanced corpora of daily conversation, and there is no special reason to doubt this supposition⁵; second, based on the results of the preliminary analyses, it seems that the overall patterns of the occurrences of *were* and *was*,

---

⁵ Although the majority of the SBC transcripts are based on daily conversations, there are also some transcripts based on university lectures, religious sermons, and business meetings. The current study investigated only the conversation transcripts in the SBC.
which will be discussed in the following section, are consistent in the SBC and the BNC conversation subsections; finally, for the current study, the full contexts from which the examples of were and was were extracted were available for the SBC data, but not for the BNC data. Given that some occurrences of were and was can be clearly judged as a subjunctive form or its alternative by considering the full contexts, including the SBC in this study might improve the accuracy of the analyses.

4. Analysis

The BNC data obtained for academic writing and conversation genres, and the narrative personal essays collected from websites were all saved in text files and analyzed for the structures concerned using a concordance program, Concordance 3.2. Likewise, the SBC data, which are available in text files, were also concordanced using the same software. Finally, the MICASE was searched for the structures concerned using its own concordance tool available in the website.

III. RESULTS AND DISCUSSION

1. Overall Patterns

Given that the main purpose of this study is to find out the overall patterns of the were-subjunctive, all tokens of as if, as though, and wish in each corpus were investigated to identify the occurrences of the subjunctive were and its alternative form, was. Because of the high frequency of if, however, only the first 10,000 tokens were selected and examined from each part of the BNC (i.e., the academic writing, the academic lecture, and the conversation sections). As for the other corpora, all tokens of if were investigated. In this way, the raw and relative frequencies of were and was occurring with each type of subject were calculated for each genre, as seen in TABLE 6. However, the relative frequencies were not calculated when the raw frequencies of were and was are extremely low; the relative frequencies for such cases could be misleading and meaningless.

---

4 This study used the BNC data available from the website http://corpus.byu.edu. This website provides only the concordance lines for any searched word or phrase, but does not provide the full contexts for it.
TABLE 6

Relative Frequencies of were & was Occurring with Each Subject Type in the Four Genres (%)

<table>
<thead>
<tr>
<th>Subject Type</th>
<th>W_ACA are / was</th>
<th>W_ESS were / was</th>
<th>S_ACA were / was</th>
<th>S_CON were / was</th>
<th>Mean are / was</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>66.7 / 33.3</td>
<td>54.1 / 45.9</td>
<td>68.6 / 31.4</td>
<td>21.1 / 78.9</td>
<td>52.6 / 47.4</td>
</tr>
<tr>
<td></td>
<td>(12 / 6)</td>
<td>(20 / 17)</td>
<td>(24 / 11)</td>
<td>(16 / 60)</td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>79.5 / 20.5</td>
<td>52 / 48</td>
<td>44.8 / 55.2</td>
<td>14 / 86</td>
<td>47.6 / 52.4</td>
</tr>
<tr>
<td></td>
<td>(186 / 48)</td>
<td>(26 / 24)</td>
<td>(60 / 74)</td>
<td>(18 / 111)</td>
<td></td>
</tr>
<tr>
<td>TOM/DOG</td>
<td>58.6 / 41.4</td>
<td>56 / 44</td>
<td>39.5 / 60.5</td>
<td>13.3 / 86.7</td>
<td>41.9 / 58.1</td>
</tr>
<tr>
<td></td>
<td>(170 / 120)</td>
<td>(14 / 11)</td>
<td>(32 / 49)</td>
<td>(4 / 26)</td>
<td></td>
</tr>
<tr>
<td>THERE</td>
<td>54.3 / 45.7</td>
<td>NA</td>
<td>21.6 / 78.4</td>
<td>0 / 100</td>
<td>25.3 / 74.7</td>
</tr>
<tr>
<td></td>
<td>(19 / 16)</td>
<td>(3 / 1)</td>
<td>(8 / 29)</td>
<td>(0 / 19)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>67.4 / 32.6</td>
<td>54.3 / 45.7</td>
<td>43.2 / 56.8</td>
<td>15.0 / 85.0</td>
<td>49.6 / 50.4</td>
</tr>
<tr>
<td></td>
<td>(387 / 190)</td>
<td>(63 / 53)</td>
<td>(124 / 163)</td>
<td>(38 / 216)</td>
<td>(612 / 622)</td>
</tr>
</tbody>
</table>

(W_ACA: academic writing, W_ESS: narrative personal essay, S_ACA: academic spoken, S_CON: conversation)

* Raw frequencies in the parentheses

Moreover, for more meaningful comparison of the occurrences of were and was depending on subject type, the mean relative frequencies, instead of the accumulative frequencies, were calculated for each subject type. This was based on the judgment that the total frequencies of were and was occurring with each subject type across the discourse genres may disguise the real distribution patterns, since the total frequencies of were and was occurring with a specific subject type were dominantly affected by the frequencies of were and was occurring in a specific discourse genre. For example, the total frequency of <I was> (= 94), was dominantly affected by its frequency in conversation (= 60), while such a dominant case was not observed for <I were> in any discourse genre. Accordingly, comparing the total frequencies of <I were> and <I was> across the discourse genres may misrepresent their real distribution pattern for subject type, I. This suggests the necessity of considering discourse genres in investigating the occurrences of English structures in corpus-based studies.

To find out the overall patterns of the occurrences of were in relation to its alternative was, the relative, rather than raw, frequencies of the two forms are firstly compared among the four discourse genres and then among the four subject types, as follows.

1) Comparison among the discourse genres

Overall, the occurrences of were and was were almost identical (49.6% vs. 50.4%). However, meaningful variations are found when their occurrences in different genres are compared. As clearly seen in TABLE 6 and FIGURE 1, the selection between were and was seems to be closely related to the register and the formality of discourse. The relative frequencies of were are gradually decreasing as the register changes from written to spoken and the formality from high to low. The relative frequency of were is highest in academic
writing (about 67%) and decreases to approximately 54% in narrative personal essays, still slightly higher than the relative frequency of was. This pattern is reversed in the two spoken genres. The relative frequency of were is somewhat lower than that of was in academic spoken discourse (43.2% vs. 56.8%), and were is no longer a competing alternative to was in conversation (15% vs. 85%). Based on this observation, it may be argued that were is more common than was in written discourse, especially in academic written discourse, but was is more common in spoken discourse. In particular, were seems to be nearly nonexistent as a means to mark the subjunctive in informal conversation.

These findings are compatible with Trudgill’s (1999) observation that forms tend to spread from non-standard dialects to the standard one. As confirmed in this study, the phenomenon that the subjunctive were is gradually being replaced by was occurs first in spoken discourse, which may allow more non-standard or newly emerging forms than written discourse, and gradually spreads to written discourse that is likely to retain more standard forms than spoken discourse, although this does not make spoken discourse a non-standard dialect.

To see if the distribution of were and was in each genre occurs by chance or not, a Chi-squared test was conducted for the raw frequencies presented in TABLE 6, and the result shows that the distribution is significant (Chi-squared = 198.09, df=3, p < .000). Based on this, therefore, it may be said that the distribution of were and was is strongly associated with discourse genre.

FIGURE 1
Frequencies of were and was in Each Genre (%)
2) Comparison among the subject types

To uncover the patterns of were and was as they related to subject types, the relative frequencies of were and was occurring with each subject type were compared for each genre, and then the mean frequencies were compared for each subject type (see TABLE 6). FIGURE 2 shows the overall relative frequencies of were and was occurring with each subject type.

As seen in TABLE 6 and FIGURE 2, <I were> is more common than <I was> (52.6% and 47.4%, respectively), even though the difference between the two relative frequencies is slight. However, <HE were> is a little ‘less’ common than <HE was> (47.6% and 52.4%, respectively). A similar pattern was found for <TOM/DOG were> and <TOM/DOG was> (41.9% and 58.1%, respectively), although the gap between the two relative frequencies is larger than that between the relative frequencies of <HE were> and <HE was>. That is, was is a little more likely to occur with third person singular proper/common nouns than with third person singular pronouns. Finally, <THERE were> is much less common than <THERE was> (25.3% and 74.7%, respectively). Compared with the other three subject types, existential THERE shows quite a strong preference for was. In particular, no tokens of were were found with existential THERE in conversation, as shown in TABLE 6. A Chi-squared test was not carried out in this case, since as previously discussed, the total raw frequencies of were and was occurring with each subject type were not calculated.

![FIGURE 2](image.png)

So far, the overall patterns of the occurrences of were and was have been discussed with
regard to discourse genre and subject type. To summarize, were is more common than was in the written genres and in formal style. Therefore, were is much preferred to was in academic writing; this preference, however, decreases in narrative personal essays. In contrast, was is preferred to were in spoken discourse, even in more formal genres such as academic lectures and is almost exclusively selected in informal conversation. Given the relative distributions of the two forms in the four discourse genres, at least in conversation, was can be said to have become a major form of the were-subjunctive, while were is still competing with was in the other discourse genres. With regard to subject type, were is slightly more common than was with I. The other three subject types preferred was, especially THERE. Generally speaking, however, except for existential THERE, the other three subject types do not show a noticeably strong preference for one form. These results indicate that, contrary to some linguists’ claim that the subjunctive has died out in modern English, the subjunctive were is still alive and competing with was at least in some specific discourse genres and linguistic contexts. Thus, it is necessary to focus on local patterns of the were-subjunctive as it occurs in various linguistic and extralinguistic contexts such as discourse genre and subject type rather than briefly describe its general patterns, as done in most ESL grammar books.

2. Genre-specific Features and Lexical Bundles or Fixed Phrases

Based on the current analyses, several genre-specific features of the were-subjunctive were identified that are not necessarily compatible with its overall usage patterns discussed previously. In addition, several lexical bundles containing were or was were identified; some may be considered genre-specific features while others are found across genres.

1) It were in academic written genre

As discussed above, in academic writing, the relative frequency of were is highest when the subject type is HE, (i.e., third person singular pronouns) (79.5%), and second highest when the subject is I (66.7%). This result is not consistent with the overall patterns of the were-subjunctive discussed above, which show that the relative frequency of <I were> is higher than that of <HE were>. Moreover, the relative frequency of <HE were> in this genre is remarkably high, compared with its overall relative frequency, 47.6%, with its relative frequencies in the other genres, and with the relative frequencies of were occurring

---

5 In this study, a lexical bundle refers to certain words that occur together as a set and that set is open to addition of other words to complete the meaning and the grammar, while a fixed phrase is a set of words that complete the meaning and the grammar in itself. By this definition, <as if to say> or <I wish I were> is referred to as a lexical bundle, while <If I were you> as a fixed phrase.
Song, Min Young

with the other subject types in this genre (refer to FIGURE 1 and FIGURE 2 for details). A possible reason for this may be that among the various types of 3rd person singular pronouns represented by HE, it is much more commonly followed by were than was in this genre. The numbers of tokens and relative frequencies of <it were> and <it was> in this genre are 139 (83.7%) and 27 (12.3%), respectively. Such a high relative frequency of <it were> is observed only in the academic written genre, but not the other genres: 17 (60.7%) for personal essays, 27 (36.5%) for academic speech, and 9 (15.5%) for conversation. Thus, it might be argued that it prefers were much more to was or <it were> is used as a lexical bundle in academic written discourse. On the other hand, based on the relative frequencies of <it were> and <it was> in conversation (15.5% vs. 84.5%), the latter appears to be a lexical bundle in conversation. However, considering that as indicated in TABLE 6 and FIGURE 1, the relative frequencies of were and was in this genre (15% vs. 85%) are quite similar to those of <it were> and <it was> in this genre, the high relative frequency of <it was> in conversation should be attributed to the genre-specific overall pattern for the were-subjunctive. Thus, unlike <it were> in academic written genre, <it was> cannot be regarded as a lexical bundle in conversation.

2) If I was you in conversation

As Quirk et al. observe (1985), <if I were you> is generally considered a fixed phrase. This seems to be the case for all the genres explored in this study except for conversation. In this study, nine cases of <if I were you> were found and five of them occurred in conversation. As for <if I was you>, only six cases were found, and interestingly, all six cases occurred in conversation. Although the raw frequencies of the two alternative forms are not high enough to draw a definite conclusion, it seems that <if I were you> is just an alternative to <if I was you>, rather than a fixed phrase in conversation. Thus, <if I was you> might be considered a genre specific feature of conversation. Moreover, considering that new forms tend to spread from non-standard dialects to the standard one (Trudgill, 1999), this observation might be an indication that <if I were you> is losing its status as a fixed phrase in modern English and instead <if I was you> could be regarded as a fixed phrase, one that may eventually become fixed in written discourse as well.

3) I wish (that) ~ were / was in informal discourse

A nominal clause after the verb wish can also be considered a genre-specific structure of informal discourse. As seen in TABLE 7, few examples of this structure were found in the two academic discourse genres. Most examples were found in conversation (46 cases) and a few examples in narrative personal essays (7 cases). As indicated by the numbers of
tokens per 1 million words in TABLE 7, however, the magnitude of the 7 is even greater than that of the 46. Therefore, this structure may be a genre-specific structure for both informal discourse genres. Biber et al. (1999) categorize <wish+ that-complement clause> as an attitudinal stance marker which shows personal feelings or emotions. Therefore, this structure may be common only in personal conversations or literary genres in which speakers or writers express their emotions and feelings, rather than academic discourse genres. This supposition probably also explains the reason why this structure mostly occurs with the subject I.

<table>
<thead>
<tr>
<th></th>
<th>W_ACA</th>
<th>W_ESS</th>
<th>S_ACA</th>
<th>S_CON</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>were</td>
<td>1</td>
<td>3 (7)</td>
<td>0</td>
<td>3 (0.6)</td>
<td>7</td>
</tr>
<tr>
<td>was</td>
<td>1</td>
<td>4 (9.8)</td>
<td>1</td>
<td>43 (9.1)</td>
<td>49</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>7 (16.8)</td>
<td>1</td>
<td>46 (9.7)</td>
<td>56</td>
</tr>
</tbody>
</table>

* # of tokens/1 million words in the parentheses

On the other hand, both were and was occur quite evenly in the <I wish (that) ~> structure in narrative personal essays, even though their raw frequencies are quite small (see TABLE 7). In contrast, was is much more common than were in this structure in conversation. In particular, <I wish I was> is more frequent than <I wish I were> (22 cases and 3 cases, respectively). Therefore, it might be concluded that <I wish I was> has become a lexical bundle in informal conversation.

4) If ~ were to inf.

As shown in FIGURE 3, <if ~ were to inf.> is not only much more common than <if ~ was to inf.>, but the relative frequencies of <if ~ were to inf.> are quite similar to one another across genres (approximately 74%), which is not consistent with the overall patterns of the occurrences of were and was discussed in terms of discourse genre. Moreover, TABLE 8 shows that the relative frequencies of <if ~ were to inf.> are quite similar to one another across subject types. Based on these observations, it might be argued that <if ~ were to inf.> is a lexical bundle found across genres and subject types. That is, regardless of genre and associated subject type, were is not very likely to be replaced by was in this structure.
As seen in TABLE 9, the relative frequencies of <as if it were> are very high in the two written genres (i.e., academic writing and personal essay) and moderately high in academic lectures. These relative frequencies are even higher than the relative frequencies of <it were>, as previously discussed. Therefore, it might be argued that <as if it were> is an example of lexical bundles found in most genres including academic lectures, although <as if it were> is sometimes replaced by <as if it was> in this genre. This trend is reversed in conversation, where <as if it was> is used more frequently than <as if it were>. Considering that <as if it were> is disappearing in conversation, this phrase may eventually disappear from written discourse as well. For now, however, it can be regarded as a lexical bundle in most genres including academic lectures.

**TABLE 9**

Frequencies of <as if it were/was> in Each Genre (# / %)

<table>
<thead>
<tr>
<th></th>
<th>W_ACA</th>
<th>W_ESS</th>
<th>S_ACA</th>
<th>S_CON</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>were</td>
<td>81 / 89%</td>
<td>8 / 72.7%</td>
<td>11 / 61%</td>
<td>1 / 12.5%</td>
<td>101 / 78.9%</td>
</tr>
<tr>
<td>was</td>
<td>10 / 11%</td>
<td>3 / 27.3%</td>
<td>7 / 39%</td>
<td>7 / 87.5%</td>
<td>27 / 21.1%</td>
</tr>
</tbody>
</table>
Based on the results of the current study, it seems that overall, subject *it* is more readily combined with *were* than any other type of subject, even more than *I*. Presumably, this phenomenon might be related to the fact that <*as it were* is frequently used as a fixed phrase in modern English. As described by Quirk et al. (1985, p. 158), *were* cannot be replaced by *was* in this fixed phrase. The existence of this fixed phrase might affect people’s selection between *were* and *was* when the associated subject is *it*. On the other hand, contrary to Quirk et al.’s (1985) description that *were* is generally used in the fixed phrase, *If I were you*, *were* seems to have been almost replaced by *was* in conversation.

To sum up, based on the current analyses, some genre-specific structures and lexical bundles or fixed phrases were identified. They provide authentic examples of the *were*-subjunctive in which one form, i.e., *were* or *was*, is not likely to be replaced by the other in a specific genre, and therefore confirm the necessity for paying attention to local usage patterns of the *were*-subjunctive as it occurs in various linguistic and extralinguistic contexts.

### IV. CONCLUSION AND IMPLICATION

Contrary to the general view that the *were*-subjunctive has almost disappeared in modern English, and consequently, is not worthwhile to examine in detail, this corpus-based study has shown that the *were*-subjunctive still exists and is frequently used in some discourse genres and lexical bundles/fixed phrases. Based on these findings, therefore, it may be argued that despite its uncertain status in modern English, the actual usage of the *were*-subjunctive should be investigated in detail in corpus-based studies. In addition, informed by the findings from such studies, more detailed descriptions of the current status of the *were*-subjunctive in contemporary English grammar should be provided in the literature. In addition, as suggested by the current study, more attention needs to be paid to local patterns of the English subjunctive including the *were*-subjunctive in specific linguistic and extralinguistic contexts rather than only its general usage patterns. Especially, as confirmed by this study, in investigating local patterns of English structures, discourse genre should be considered a crucial context that affects distribution of the structures.

This study provides several implications for English as a second or foreign language education. Probably due to the unavailability of practical information on the *were*-subjunctive, English teachers have provided students with only brief theoretical descriptions of the usage of the *were*-subjunctive as found in most primary ESL/EFL grammar texts, hardly aware of the remarkably frequent use of *was* in some specific discourse genres and lexical bundles or fixed phrases. Therefore, English teachers and learners would benefit from the use of such
practical information on the subjunctive that has emerged from a close analysis of authentic written and spoken language. More specifically, English teachers may need to be well aware of the actual usage patterns of the were-subjunctive and its substitute, was, and provide students with more detailed descriptions of the contexts in which one form is preferred to the other. In particular, given that were has almost been replaced by was in informal conversation, and that secondary English education in Korea has been paying more and more attention to informal spoken discourse, English textbooks might need to present was, instead of were, as the major form in conversation, or at least raise students’ awareness of the actual usage patterns for the were-subjunctive by presenting both forms. Moreover, English teachers should present students with such genre-specific lexical bundles or fixed phrases as identified in this study, in order to help students select one form rather than the other for specific subjunctive expressions.

Finally, considering that compared with other English structures, the subjunctive has not been paid enough attention to by researchers, more corpus-based research should be conducted. Thus, it would be advisable to complement a corpus-based quantitative analysis with a qualitative analysis such as a functional or discourse analysis. The current study may also be followed by an intensive functional or discourse analysis of the contexts affecting the selection of were or was.

REFERENCES

A Corpus Analysis of the Were-Subjunctive in English: A Comparative Study of Four Discourse Genres


Applicable levels: secondary and tertiary education
Key words: were-subjunctive, indicative was, English subjunctive, corpus linguistics, discourse genre

Min Young Song
Applied Linguistics and TESL at University of California, Los Angeles
68-26, Yangji Apt. B-201, Sinsu-Dong, Mapo-Gu, Seoul, Korea
Email: songmy@ucla.edu