Article

Corpus-Based Investigation of S-V Concord Patterns of Nouns with Latin Plural Endings

Audrey B. Morallo
College of Education
University of the Philippines
Diliman, Quezon City, Republic of the Philippines

ABSTRACT

English subject-verb concord can pose problems to learners such as omission of third person singular inflection and overextension due to linguistic typology, phonological and perceptual factors, syntax, and the number of nouns acting as the subject of the sentence, among others. A group of nouns that invites confusion is what Huddleston and Pullum (2002) call nouns with Latin plural endings whose singular forms are uncommon compared to their plural counterparts. This investigation aims to provide teachers with guidance on the general and genre-based agreement patterns of these nouns using the balanced Corpus of Contemporary American English (COCA). This study found that based on their agreement patterns these nouns can be classified into three groups: 1) nouns whose tendencies are line with prescriptive rules; 2) noun whose tendencies are ambivalent; and 3) nouns which might have already been "reclassified" by speakers. By genres, there was a strong preference to use "plural" verbs with these nouns in the written and the academic components of the corpus while "singular" verbs are preferred in the spoken components. The implications of the study for teaching and research were also provided.

KEYWORDS: Subject-Verb Agreement, Corpus Linguistics, Agreement Patterns, Genre-Based Agreement Patterns

INTRODUCTION

A child acquiring or learning English will immediately be exposed to the mechanics of subject-verb agreement, also known as concord, in his/her earliest language classes or in any environment where the target language is spoken because this agreement helps in constructing the structure of the sentence (Keeney & Wolfe, 1972). The forms that enable one to encode this concord are supposed to be simple to describe and understand, yet they pose substantial problems at all instructional levels (Celce-Murcia & Larsen-Freeman, 2008) not only to the native speaker but also to the second language learner, suggesting that, far from being straightforward, agreement is a complex and messy area of grammar (Hudson, 1999). This paper attempts to look at one problem area- the equivocal agreement with a group of nouns with Latin plural endings (Huddleston & Pullum, 2002) and to look for guidance to the possibilities offered by corpus linguistics.

Specifically, this paper will attempt to describe the agreement patterns of these nouns with the three primary verbs in English, *be*, *have*, and *do*, (Quirk et al., 1985) in different genres using a balanced corpus.

The general rule for subject-verb concord, which is an obligatory rule in all English third person, present tense sentences (Keeney & Wolfe, 1972), in most cases is: the subject and the verb number choice will agree (Celce-Murcia & Larsen-Freeman, 2008; Hudson, 1999; Keeney & Wolfe, 1972). To be more precise, when the subject is singular, the noun is uninflected and the verb uses the third person singular inflection, *-s* (or the *be* verb *is*); when the noun is plural, it is inflected with *–s* while the verb is uninflected (Celce-Murcia & Larsen-Freeman, 2008; Keeney & Wolfe, 1972). The concept of grammatical person is also considered in agreement, but this applies only to the form of *be* when its subject is the singular form of the first person personal pronoun, *I*. Due to this, this is criticized as problematic and having no function in grammar (Hudson, 1999). In cases where the predicate is a verb phrase (VP) with an auxiliary element, it is the tense-bearing auxiliary element that takes the agreement and not the main verb of the VP. Because of the complexity of the choice facing English-as-second-language (ESL) or English-as-foreign-language (EFL) learners, subject-verb concord poses problems or issues to learners that can result in errors committed especially those at the beginning stages of their language learning/acquisition (Celce-Murcia & Larsen-Freeman, 2008).

These problems or issues can be traced to a number of reasons. First, ESL/EFL learners have been observed to simplify and omit the third person singular inflection, thus producing sentences such as this:

>   *President Duterte prescribe the use of gasoline in disinfecting face masks.

In some cases, learners commit overextension and inflect *–s* on modal auxiliaries or main verbs following modal auxiliaries. Another source of error is language typology as topic-prominent languages such as Korean and Japanese do not mark subject-verb concord, with a possible exception in Filipino which still marks the verb if the subject is plural (C.J. Aguila, personal communication, July 31, 2020). Thus, learners from these linguistic backgrounds might have a hard time mastering this aspect of English grammar. Phonology and perceptual factors might also have an impact on agreement errors of learners. For example, some learners might not produce the *–s* inflection in speech because their linguistic background does not permit its production such as in the case of French speakers. This phenomenon might also be because of its lack of perceptual saliency, its infrequent appearance in native speaker speech (Larsen-Freeman as cited by Celce-Murcia & Larsen-Freeman, 2008), or its low communicative value since person/number is clear from the subject of the sentence anyway. Errors or confusion might also be because of syntactic reasons. For example, Hudson (1999) believes that word classes sensitive to the noun's number, i.e., the determiner and the verb, might lead to conflicting classifications such as in the case of this sentence where the determiner *this* signals that *family* is taken as singular, yet two forms of *be* are possible, *is* and *are*:

>   This family is/are going up to Baguio City at the weekend.

Another source of error or confusion in subject-verb agreement is whether the verb agrees with the grammatical number or the notional number of the subject (Humphreys & Bock, 2005). In the case of the subject *family* above, grammatically speaking, it looks singular as it does not have a

morphological marker for plurality, the *–s* inflection. However, it is not uncommon for one to encounter sentences where *family* as a subject takes a plural verb, especially in British English (Celce-Murcia & Larsen-Freeman, 2008; Huddleston & Pullum, 2002). This notional agreement might be due to the numerosity of the subject's referent in the mind of the speaker, thus triggering what Huddleston and Pullum (2002) call plural override.

In addition to what has been discussed above, nouns that have equivocal number, i.e., those that can be treated either as singular or plural or where usage is divided over whether to treat them as one or more than one (Huddleston & Pullum, 2002; Quirk et al., 1985), might also be a source of learners' confusion over usage. According to Huddleston and Pullum, these nouns have Latin plural endings and have historically unequivocal singular forms:

Table 1. **Nouns with Latin endings and their historical singular forms**

Nouns with Equivocal Number	Historical Singular Forms
1. algae	alga
2. media	medium
3. bacteria	bacterium
4. phenomena	phenomenon
5. criteria	criterion
6. data	datum
7. insignia	insigne

Although these singular forms are undoubtedly still being used, these are uncommon compared to their plural counterparts (Huddleston & Pullum, 2002). They add that this has led speakers to the construal that the plural forms are singular and unrelated to their historical singular forms, especially *alga*, *bacterium*, and *insigne* which are rarely utilized outside their highly specialized fields. This reanalysis of the number of these words is what has taken place in the case of *agenda* and *candelabra*, which were historically plural nouns but are now treated as singulars (Huddleston & Pullum, 2002). The complexity of choices in subject-verb concord facing the language learner and the fact that several nouns have equivocal agreement pose serious pedagogical questions to the language teacher. Some teachers might find themselves teaching "rules" on these agreement areas and then be immediately confronted by examples from students contradicting such prescriptions. On the other hand, some teachers might avoid giving straightforward answers and run the risk of students' not having any firm guidance on these matters. This study aims to bridge this gap and provide teachers with corpus-based descriptions of the agreement patterns of these nouns with equivocal agreement using a balanced corpus that teachers can use in navigating such a treacherous area of grammar.

Corpus linguistics is a branch of linguistics that uses computers to analyze and study large collections of languages. Crucial in this field is the use of a corpus. A corpus is a collection of linguistic data, which can be spoken and/or written, compiled for the primary purpose of research (Liu, 2011) and stored on a computer for analysis (Hunston, 1995; Kayaoğlu, 2013). The collection of linguistic data for a corpus is not arbitrary, however. They should have been selected and ordered according to an explicit set of criteria (Huo, 2014; Kayaoğlu, 2013). Some types of sample discourse which can be collected for a corpus are spoken, written, computer-mediated,

spontaneous, or scripted and may represent a variety of genres such as conversations, lectures, seminars, and mass communication broadcasts (Huang, 2011).

One of the best known and largest corpora in the world is the Corpus of Contemporary American English (COCA) developed by Brigham Young University's Mark Davies. This corpus is considered balanced since it draws linguistic data from several genres: spoken, fiction, popular magazines, newspapers, and academic texts (Huo, 2014). Using COCA for linguistic investigations offers several advantages (Yusu, 2014). First, COCA is large enough to show the lexical and grammatical patterns of a language. COCA can also provide a more realistic picture of how words and structures are used by language speakers since it is a corpus that draws from a variety of registers. COCA can likewise provide insights on how words have been used over time. It is also tagged for word classes and can be used in studying collocates, synonyms, frequency, among others, facilitating the linguistic analysis of the data. In addition, COCA can also simultaneously show frequency and sample sentences which can prove useful in linguistic analysis. Finally, COCA is easy to use and will not require any special linguistic training for its use.

Although corpora were initially developed for linguistic research, they have been playing a significant role in language learning and teaching for a few years now (Biber & Conrad, 2001). One can trace the contribution of corpus linguistics to language learning and teaching along three lines: Direct use of corpus, indirect use of corpus (McEnery et al., 1997; Römer, 2011), and the way teaching of subjects have been approached (McEnery et al., 1997). Direct application involves direct access to corpora and concordances by teachers and learners, thus affecting how a lesson is taught and learned (Römer, 2011). Other direct applications include awareness-raising exercises and in the use of corpus to enhance the teaching of language and linguistics (McEnery et al., 1997), integration of grammar teaching in the teaching of vocabulary (Conrad, 2000), and efforts to include corpus linguistics in initial language teacher preparation to enhance the research skills and language awareness of teachers (O'Keeffe & Farr, 2003). Corpora can also provide authentic language samples to students, helping them raise their awareness on language variety, variations in meaning between language forms, and patterns of use (Kayaoğlu, 2013; Römer, 2011). The second area of contribution of corpus to language education is in the indirect use of corpus data as has been done in their utilization in classroom materials and books such as dictionaries and grammars (McEnery et al., 1997; Römer, 2011) where better and more accurate descriptions of language can be expected (Conrad, 2000). Indirect use is also shown in using corpus insights to inform decisions on what linguistic items to teach, when, and in what order, and on syllabus design (McEnery et al., 1997; Römer, 2011). The third area where corpus has an impact on language learning is in how certain subjects are approached today (McEnery et al., 1997). For example, some courses have become corpus dominated since the late 1960s (McEnery et al., 1997), a trend that continues until this century (e.g., O'Keeffe & Farr, 2003). Corpus data can also provide authentic samples of synonyms, collocates, and customized word lists, which can all be limited by period and genre such as spoken, fiction, popular magazines, newspapers, and academic journals.

A few studies have used corpus data to improve and better inform decisions in language teaching and its various areas. In literature teaching, a study on the structures of similes (Hussein & Sawalha, 2016) suggested a more liberal attitude toward the use of similes by non-native speakers after it found that multiple nouns can follow similes, contrary to prescriptions that these figures of speech are formulaic and only specific nouns can appear with these forms. Another research (Grant, 2007) meanwhile discovered that some idioms were used more in academic spoken

American English compared to general British English, which could mean that these idioms were more American or were part of the idiolect of speakers.

Another area where corpus data have found robust use is vocabulary instruction because its authentic data, context, and frequency can provide learners with critical information simultaneously (Yusu, 2014). This utility has been demonstrated by studies that applied corpora in improving different aspects of vocabulary instruction. For example, an investigation (Kayaoğlu, 2013) found a corpus effective in helping students differentiate between close synonyms, words that have similar meanings but cannot be substituted for each other, even if the students found the experience demanding. Corpora have also been used to address concerns that vocabulary presented in textbooks is simply based on intuition and not on actual use. For example, a study (Liu, 2003) found that the selection of idioms in published materials was inconsistent, with frequent items missing and less frequent ones include in these learning resources. Variations of idioms were also missing in these materials, leading the study to recommend that selection of idioms in these resources be based on authentic language rather than intuition. In another study (Liu, 2012), COCA and the British National Corpus were used to identify the most frequently-used multiword constructions (MWCs) of various types of academic writing and found that there was a mismatch between lists of MWCs in published materials and what the corpus data showed was frequent. Another investigation probed the relationship between native speaker intuition and corpus data, specifically the relationship of corpus-frequency to native speaker self-reported frequency of words (Okamoto, 2015). This paper concluded that native speakers played a role in the selection of words to be included in ESL/EFL textbooks up to the 7000-word and 6000-word levels, which can be described as a realistic goal in teaching. Corpus can also be used to identify words that can be prioritized in instruction such as what another study did (Hsu, 2018). The investigation developed a vocabulary word list for Traditional Chinese Medicine (TCM) by compiling texts about this discipline and measuring their vocabulary levels using the frequency scale of COCA and BNC. A corpus can also be used to probe the pragmatic functions of phrases in discourse. A study on the functions of the phrase "just kidding" in American English (Skalicky, Berger, & Bell, 2015) found that there were four uses for the phrase in discourse: inoculation of the speaker from failure of prior utterance, repair of failed humor, return to a serious frame, and set-up of a new joke.

The teaching of academic English and academic writing has been influenced, too, by corpus insights. Arguing against suggestions that the academic component of COCA was insufficient for some types of searches, Davies (2013) explained that the corpus was functional enough for searches for word, phrase, or construction that might be of interest in academic English teaching. It said that COCA could compare words, phrases, constructions in academic and non-academic English and in domains and discipline in the academic genre; show how these are used, the patterns where they occurred, and the other words they occurred with; generate collocates for comparison within and between texts; and search for semantically-oriented queries and not just strings of words. Another investigation (Charles, 2007) tried to reconcile top-down and bottom-up approaches to graduate writing by using a corpus to teach rhetorical functions. It found that combining discourse and corpus work could offer teachers several options to teach a rhetorical function and provide enriched input to students that would make them see the connection between general rhetorical purposes and specific lexicogrammar choices. In another study (Pham, 2022), a teacher-made corpus of IELTS essays written by IELTS experts was employed to teach grammar in IELTS writing classes. It found that this 1) boosted teachers' confidence in materials preparation as it enabled them to identify structures to prioritize in IELTS writing, 2) promoted student writing

by raising his/her awareness of linguistic features common in IELTS writing, and 3) enhanced students' sense of ownership of learning as it allowed them to investigate other structures using a corpus outside of class. Using a corpus of essays of students from Chinese universities and the Leuven Native Language Composition Corpus (LONCESS) as the corpus of native speaker writings to compare use of hedging, a study (Zhao, 2022) found that Chinese students were assertive in their writing and did not properly employ hedging in their essays. It also reported that Chinese students tended to overuse hedges and employed only a limited number of hedging devices and that proficiency may not be a big factor in these patterns of use.

There are also studies on the use of corpus in teacher preparation courses and teacher training. A case has been made for the inclusion of corpus linguistics in teacher preparation education to enhance teachers' research skills and language awareness (Hunston, 1995; O'Keeffe & Farr, 2003). Teachers should be empowered to use corpora because many teaching materials were already based on them and because these would enable them to evaluate these (O'Keeffe & Farr, 2003). In the case of non-native teachers, corpus training would allow them to use corpora for pedagogic ends and assess ideas presented to them as facts and place them for sociocultural mediation and pedagogic re-contextualization. Corpus training will also allow teachers to carry out research to confirm insights of traditional grammar and understand that grammar learning is about learning to do grammar and not learning about it (Hunston, 1995) or what is called "grammaring" (Celce-Murcia & Larsen-Freeman, 2008). Finally, a grounding in corpus will take out grammar from the head of the grammarian and allow people to have a say on it, thereby making it a less esoteric activity (Hunston, 1995). An example of teacher training for use of corpus in classrooms is provided by a study (Ma et al., 2021) which set out to do this by proposing a two-step training framework for corpus-based teacher training to develop two distinct concepts- corpus literacy (CL) and corpus-based language pedagogy (CBLP)- among TESOL trainee teachers in a vocabulary course. Results showed that the participants rated themselves as having a fairly good level of CL and were aware of the advantages of corpus use in their teaching. They were also able to demonstrate good pedagogical awareness by integrating corpus resources in their practice as evidenced by their lessons.

Corpus linguistics has also influenced grammar teaching. With the opportunities offered by corpus data, teachers can now have a more nuanced and principled approach in contrast to before when they would just dish out grammar rules to students. For example, since corpus data from a variety of genres have shown that grammatical patterns differ across varieties, register-specific descriptions and on patterns of use across registers can start to replace monolithic views on the English language, thereby meeting the needs of students and teachers in specialized contexts (Biber & Conrad, 2001; Conrad, 2000; Conrad, 1999) such as English for Specific Purposes (ESP) courses. Corpus data can also provide more evidence of the interplay between grammar and lexis, which is embodied in the stance on language pedagogy called lexicogrammar (Conrad, 2000; Liu, 2011; Sambursky, 2014). Corpus-informed grammar pedagogy can also push the language teaching field to following case law in grammar where the definition of standard language is based on majority usage (McEnery et al., 1997). When before linguists and ESL practitioners would rely on native-speaker intuition on areas of usage, even if their intuitions usually disagreed, today we can have a better understanding of usage through information on how people use language, enabling teachers and students to appreciate how different linguistic forms are used at different levels of formality and fulfill differing needs in different contexts (Huang, 2011).

Corpus-based grammar studies have proven to be a fertile research ground for the few years. Hunston and Francis (1998) showed how a pedagogic grammar of English that was based on naturally-occurring data, was lexically based, and could easily be lent to text-based and concordance-based consciousness-raising techniques could be made. Probing the effects of integrating corpus and contextualized lexicogrammar in ESL/EFL teaching, Liu and Jiang (2009) uncovered that this approach enhanced language awareness and command of lexicogrammatical patterns; led to appreciation of context in lexicogrammatical choices; increased critical understanding of grammar; and promoted discovery learning and made learning more interesting and effective. Meanwhile, a study (Barbieri, 2005) on the frequency of use and grammatical associate patterns of quotatives such as *be like*, *go*, *be all*, and *say* in four conversation registers found that the type and frequency of quotatives varied across the registers and that direct quotation was less frequently used in academic registers than in casual interactions. A corpus-based approach was also employed in analyzing get-passives in English (Jong-Bok, 2012). This found that get-passives had many peculiar characteristics and different grammatical, semantic, and pragmatic properties compared to be-passives. Corpus data were also used in studying the semantics and grammar of adverbial disjuncts in Philippine English. According to a study (Dita, 2014), disjuncts in Philippine English came in different forms, were found mostly in initial and final positions, and prototypically functioned as a comment on their accompanying clause. It concluded disjuncts were used by Filipinos in ordinary speech, defied the restrictions set by Quirk et al. (1985), and could contribute to cohesion, mitigation, intensification, pragmatics, discourse filling, among others. An experiment on the effects of corpus use in a university grammar course (Estling Vannestal & Lindquist, 2007) concluded that introducing the use of corpus to students would require a great deal of time, support, patience, enthusiasm, and reflection from the teacher. The students would also need clear guidance on drawing conclusions from corpus data, according to the study. Meanwhile, Liu (2011) found that corpus may be helpful in grammar and writing classes; may assist in developing the critical understanding and awareness about lexicogrammatical and language issues; and might enhance the appreciation for context- and register-appropriate grammatical forms. The investigation concluded that corpus may help in making grammar teaching more empowering, contextualized, and progressive.

Despite these aforementioned positive outcomes, there are problems and obstacles that need to be hurdled for corpora to fully realize their potential in the language classroom. First, corpora have been designed with researchers in mind, and this might pose technical difficulties in terms of access to teachers and students (Huang, 2011; McEnery et al., 1997; Römer, 2011). In addition, the use of corpus in language teaching needs to be expanded by introducing it in teacher preparation institutions and by providing more forums for dissemination of insights and knowledge (McEnery et al., 1997). Using a corpus can also overwhelm learners because of the amount and complexity of data it can generate and present to students (Yusu, 2014). This might pose unnecessary cognitive, affective, and psychological burden on students and lead to frustration and disappointment (Samburskiy, 2014; Yusu, 2014). Some have also questioned if language samples taken out of their context, such as those presented in concordance samples, can be considered authentic or natural language (Huang, 2011). Not enough research has also been conducted if beginning students will benefit from corpus-based instruction (Römer, 2011).

Despite these problems and obstacles, corpus linguistics has contributed robustly to language teaching and learning. This study will hopefully make a similar contribution by analyzing subject-verb concord patterns of a selected number of nouns with Latin endings and with equivocal agreement and the primary verbs of English (Quirk et al., 1985) in different genres. This

investigation is important because other corpus studies have shown that grammatical patterns differ across varieties (Conrad, 2000) and language is used in different extents and with different meanings in different varieties and texts (Conrad, 1999). Studies such as the current investigation are needed because they shed light on the value of register variation when analyzing language use and the effect of register on the shape of grammatical features (Conrad, 1999). This paper will also hopefully improve the use of corpus insights in English language teaching and provide more resources to teachers to inform their instruction, a problem discussed by Römer (2011).

METHODOLOGY

This study utilized the Corpus of Contemporary American English (COCA). The list of nouns with equivocal agreement with Latin endings came from Huddleston and Pullum (2002) while the primary verbs in English were based on the classification of verbs by Quirk et al. (1985). According to Huddleston and Pullum (2002), the nouns with equivocal agreement in English are: 1) algae, 2) media, 3) bacteria, 4) phenomena, 5) criteria, 6) data, and 7) insignia. Meanwhile, the primary verbs in English according to Quirk et al. (1985) are: *be*, *do*, and *have.* Each noun in combination with each of the three primary verbs, *be*, *have*, and *do*, was individually searched on the corpus. To do this, the string *noun verb* was searched. For example, to find the frequency of *algae* combined with the present form of the plural *be* verb "are," the string *algae are* was searched. For the data for each string to generate information on its raw frequency, words per million, and its frequency across the text types in the corpus, the chart display of COCA was employed. For example, for the search string *algae are*, the following data were presented by the corpus:

Table 2. **Sample data generated from COCA for the string *data are***

Section	All	Blog	Web	TV/M	SPOK	FIC	MAG	News	Acad
Frequency	46	8	11	1	7	0	7	6	6
Words (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
Per Mil	0.05	0.06	0.09	0.01	0.06	0.00	0.06	0.05	0.05

To ensure that the searches yielded the results relevant to the strings, the keywords in context (KWIC) were also checked. After the checking of KWIC, the words per mil (per mil) data were gathered. Words per million are the normalized frequencies that allow comparisons between parts of the corpus despite having different number of words. The words per mil data for the singular and plural forms of the verbs (for both the present and the past tense in the case of *be*) were then compared to see the patterns of agreement in the corpus. General tendencies in subject-verb agreement concord as well as tendencies per genre were noted and analyzed.

FINDINGS AND DISCUSSION

For this section, the results for each group of nouns will be presented. A discussion of these results will then follow.

General Agreement Patterns

Table 3. **Words per million data on noun agreement with *be* verbs**

Nouns with Equivocal Agreement	Be Verbs (Present)		Be Verbs (Past)		Do Verbs		Have Verbs	
	Is	*Are*	*Was*	*Were*	*Does*	*Do*	*Has*	*Have*

1. Algae	46	81	9	16	5	5	17	23
2. Media	4509	1799	897	581	462	244	2173	1302
3. Bacteria	204	515	61	220	9	62	46	165
4. Phenomena	125	311	15	53	4	14	28	71
5. Criteria	229	658	113	577	8	45	31	104
6. Data	5457	4749	1846	5864	307	429	703	602
7. Insignia	18	7	6	4	No data		2	No data

Prescriptive accounts of these nouns will tell learners that these are plural and therefore should take the "plural" forms of the verbs *be* in both the preset and the past (i.e., *are* and *were*); *do*; and *have*. However, as shown in Table 3, this prescription is not always followed. Generally, we can see three agreement patterns in the data: 1) nouns whose tendencies are line with prescriptive rules; 2) noun whose tendencies are ambivalent; and 3) nouns which might have already been "reclassified" by speakers.

Based on the data from COCA, English speakers still have the tendency to follow this prescriptive rule when using *bacteria*, *phenomena*, and *criteria* with primary verbs. These nouns are followed by *algae* in terms of strength of tendency in following prescription. *Bacteria's* agreement tendencies with primary verbs hue closely with the prescriptivist rule. *Bacteria* has the following wpm data: *be present* (515 vs. 204), *be past* (220 vs. 61), *do* (62 vs. 9), and *have* (165 vs. 46). The same strong tendency to take "plural" verbs in agreement can be observed in the case of *phenomena*. In the case of *be* verbs, *phenomena* has been observed to take the "plural" *be* verbs: (311 vs. 125) for the present tense and (53 vs. 15) for the past tense. For *do*, the wpm data are 14 for "plural" and 4 for "singular." Meanwhile, 71 wpm were found for the string *phenomena have* while there were 28 for *phenomena has*. The final noun with equivocal agreement whose use by English speakers follows prescriptive rules is *criteria*. In agreement with present tense *be,* the wpm is 658 for "plural" and 229 for "singular". For the past tense, it is 577 for "plural" and 113 for "singular." In the case *do*, the wpm are 45 for "plural" and 8 for "singular." Finally, for *have*, the wpm are 104 for "plural" and 31 for "singular." There are 81 wpm of the string *algae are* compared to just 46 for *algae is*. The same can be observed when using the past forms of the *be verb* (16 for plural vs. 9 for singular). Usage is evenly divided when it comes to *algae* being used with do verbs, with *does* and *do* each getting 5 wpm. Finally, there are more words per million for the string *algae have* compared to *algae has*.

Of this group of nouns, *data* demonstrates ambivalence in terms of whether English speakers treat it as a singular or plural noun. In agreement with *be* verbs, tendencies differ depending on the tense of the verb. *Data* takes *is* more than *are* (5457 wpm for *is* and 4749 for *are*) while users seem to prefer *were* to *was* (5864 vs. 1846).

The final nouns in the group, *insignia* and *media*, seem to have already been reclassified by native speakers as singular, similar to the reclassification that Huddleston and Pullum (2002) said happened to *agenda* and *candelabra.* This partly confirms the observation that nouns with plural endings are reanalyzed as plurals by English speakers (Gardelle, 2016). Of the two, it seems that the tendency is the strongest in the use of *media* across the three primary verbs. In agreement with *be*, users prefer the "singular" forms in both the present (4509 wpm vs. 1799) and past tense (897

wpm vs. 581). The same tendency for the "singular verb" can also be seen in *do*. Based on the corpus data, there were 462 wpm for *media does* compared to just 244 for *media do*. Finally, there were more wpm for *media has* than for *media have* (2173 vs. 1302). Users also seem to prefer "singular" verbs in agreement with *insignia* although the wpm gathered from the corpus were limited: *be present* (18 vs. 7), *be past* (6 vs. 4), *do* (no data), and *have* (2 vs. no data).

Agreement Patterns across Genres
- ### *Strongest Tendency to Follow Prescription*

At this point, it will be wise to look more closely at the gathered wpm data to identify any patterns in the subject-verb concord of the concerned nouns with primary verbs across the genres represented in COCA. For this part of the paper, the tendency of each verb across genres will be discussed.

Table 4. **Agreement patterns of *bacteria* across genres**

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Bacteria (are)	515	37	89	9	21	10	171	44	134
Bacteria (is)	204	33	40	8	29	5	42	15	32
Bacteria (were)	220	13	30	5	6	10	50	8	98
Bacteria (was)	61	9	11	0	9	2	14	2	14
Bacteria (do)	62	9	11	2	6	1	21	8	4
Bacteria (does)	9	1	2	0	3	0	3	0	0
Bacteria (have)	165	5	24	4	17	2	62	5	46
Bacteria (has)	46	5	5	2	2	1	17	5	9

Table 5. **Agreement patterns of *phenomena* across genres**

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Phenomena (are)	311	47	66	5	2	6	36	6	143
Phenomena (is)	125	26	24	5	6	0	7	4	53
Phenomena (were)	53	2	9	0	0	3	9	1	29

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Phenomena (was)	15	2	3	3	2	1	0	0	4
Phenomena (do)	14	2	4	1	0	1	1	0	5
Phenomena (does)	4	0	2	1	1	0	0	0	0
Phenomena (have)	71	9	14	1	0	3	13	2	29
Phenomena (has)	28	4	3	2	1	1	3	0	14

Table 6. **Agreement patterns of *criteria* across genres**

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Criteria (are)	658	63	112	5	19	6	42	34	377
Criteria (is)	229	48	54	2	19	0	11	18	77
Criteria (were)	577	12	59	0	4	4	13	13	472
Criteria (was)	113	8	14	2	12	0	0	5	72
Criteria (do)	45	7	5	0	3	2	5	2	21
Criteria (does)	8	1	3	0	1	0	0	0	3
Criteria (have)	104	6	9	1	5	1	15	6	61
Criteria (has)	31	5	3	0	8	0	0	1	14

Tables 4 to 6 show the agreement patterns of *bacteria, phenomena, and criteria* across genres. Of the eight genres, six (blog, web, fiction, magazine, news, and academic texts) are written while the remainder (TV/movie and spoken) involve oral discourse. Based on the data, speakers generally prefer using the "plural" verbs with these three nouns with some exceptions in the TV/movie and spoken components of the corpus. The preference to treat these three nouns as plural is strong enough that, when Huddleston and Pullum (2002) noted the singular use of these nouns with Latin plural endings had become well established that it had been regarded as standard, they cited the cases of *criteria* and *phenomena* as exceptions to this general acceptance. As these three nouns demonstrate the strongest general tendencies to be used with "plural" verbs, it is unsurprising that the same pattern of "plural" verb agreement can be seen across the written genres. Of these genres, most of the tokens of these nouns were found in the academic texts, which is expected since these three nouns are regular features of academic discourse. Since academic discourse is formal and

needs time and reflection, English users can be expected to be stricter in their observance of subject-verb concord which still prescribes the use of "plural" verbs with these nouns which are notionally plural. This opportunity for reflection and editing can also be the reason for this strong observance of subject-verb agreement with these nouns in the other written genres. Since writing affords English users time to edit and think about their sentences (Chafe & Danielwicz, 1987), the expectation that writing will treat these as plural nouns is understandable.

Of the eight genres in COCA, it can be seen that the tendency to use "singular" verbs such as *is*, *was*, *does*, and *has* was the strongest in the spoken and to a lesser extent TV/movie components of the data. For example, *phenomena* is paired more by users with *is*, *was, does*, and *has* in the spoken and TV/movie components as indicated in its wpm measures in Table 5. The same can generally be observed in the case of *criteria*. In the spoken data, it is preferred by users to be used with *was* and *has* while speakers are divided between *are* and *is*. Finally, speakers prefer the plural *do* to *does* in speech. English users meanwhile are divided in spoken discourse in their subject-verb concord with *bacteria*. "Singular" verbs *is* and *was* and "plural" verbs *do* and *have* are preferred by speakers in speech as indicated in the wpm data in Table 4. These agreement tendencies can be explained by the nature of oral discourse. One characteristic of oral discourse is its immediacy that leaves users with little time to craft their linguistic production and to carefully consider word choice and subject-verb agreement (Chafe & Danielwicz, 1987). Thus, when one uses any of these nouns, he/she will be left with little to no time to think about the "right" subject-verb agreement for it. In addition, the plurality of these nouns is notional and not morphological, and this notional plurality is a source of difficulties, if not mistakes, on the part of learners (Humphreys & Bock, 2005). Since these nouns cannot be split into *stem + plural inflection,* there is a tendency for speakers to treat them as non-count singular nouns (Gardelle, 2016) because the -s (plural) inflection that can help the online processing of the plurality meaning of these words is not present unlike in the case of regular nouns such as *boys*. Thus, the absence of this inflection leads speakers to treat these nouns as singular.

Table 7. **Agreement patterns of *algae* across genres**

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Algae (are)	81	7	27	0	3	1	15	1	27
Algae (is)	46	8	11	1	7	0	7	6	6
Algae (were)	16	0	2	0	1	1	4	0	8
Algae (was)	9	0	2	0	0	2	2	0	3
Algae (do)	5	0	1	0	0	1	0	0	3
Algae (does)	5	2	1	1	0	0	1	0	0
Algae (have)	23	0	7	0	0	3	3	0	10

| Algae (has) | 17 | 0 | 2 | 0 | 3 | 1 | 5 | 3 | 3 |

The wmp data from COCA show that English speakers generally prefer to follow prescriptive rules in using *algae*. However, the strength of prescription across genres is not as strong as that which is observed in *bacteria*, *phenomena*, and *criteria*. As shown in Table 7, *are* is preferred with *algae* in written genres such as academic texts, magazine, fiction, and web. *Is* is preferred in blogs, but the wpm data are very close that we can say that usage is divided in this genre. Preference for *is* is also demonstrated in the TV/movie and spoken genres. For the past form of *be* the "plural" form is preferred in academic texts, magazine, and spoken genre though in the latter the wpm data are very small. A small tendency for the "singular" form meanwhile can be seen in fiction. The rest of the genres have either the same number of wpm or none at all. For *do* verbs, the "plural" is preferred in academic texts (3 wpm vs. 0) while a small preference is shown in fiction (1 wpm vs. 0). The "singular" form meanwhile is preferred in the following genres although the wpm are low: blog (2 wpm vs. 0), TV/movie (1 wpm vs. 0), and magazine (1 wpm vs. 0). Finally, users tend to use *algae* with *have* in the following genres: academic texts (10 wpm vs. 3); fiction (3 wpm vs. 1), and web (7 wpm vs. 2). *Has* meanwhile is preferred with *algae* in the following genres: spoken (3 wpm vs. 0), news (3 wpm vs. 0), and magazine (5 wpm vs. 3).

Wpm data for *algae* demonstrate two interesting facts. First, the general tendency is still to follow prescription when using *algae*. Although there are exceptions, this is shown in formal and written genres such as academic texts, fiction, and magazine where writers have time to reflect on and edit what they are writing (Chafe & Danielwicz, 1987). This gives them the opportunity follow the prescriptive rule when using *algae*. The level of formality of these genres might also be a factor as suggested by Huang (2011). Since formal language is expected in these texts, English users might be more conscious of following the prescriptive rule for *algae*. However, a closer observation of the data will also reveal that this is not true for all written genres such as blogs, TV/movie, spoken, and news media. For example, there is a small preference to use *has* with *algae* in magazine articles. Usage is also divided in the spoken and TV/movie components of the corpus. This usage preference might be traced to the oral and more informal nature of these texts. Although more conclusive data are needed, it is possible that *algae's* reclassification by English speakers in their speech and writing has begun, as noted by Huddleston and Pullum (2002) when they said that the singular use of *algae* was already well established enough for it to be treated as standard.

- ***Ambivalent Agreement***

Table 8. **Agreement patterns of *data* across genres**

	All	Blog	Web	TV/ Movie	Spoken	Fiction	Magazine	News	Academic Texts
Data (are)	4749	422	787	11	44	24	402	177	2882
Data (is)	5457	1462	1403	115	242	77	623	339	1196
Data (were)	5864	135	471	4	11	27	124	80	5012

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Data (was)	1846	242	274	40	55	62	167	93	913
Data (do)	429	58	71	0	12	2	51	22	213
Data (does)	307	98	70	10	14	4	41	21	49
Data (have)	602	65	103	4	10	6	75	41	298
Data (has)	703	166	151	30	36	9	90	51	170

Of the nouns being analyzed, *data* seems to show the highest level of ambivalence in subject-verb agreement. Compared to the nouns in the first and the latter groups, *data* exhibits more equivocal agreement with primary verbs overall. It is preferred to be partnered with "singular" verbs *is* and *has* and "plural" verbs *were* and *do*. With present *be*, the preference of English users is to use *data* with *is* generally and across genres with a notable exception in academic texts which are source of more than half of the wpm data for *are*. In the past tense, the tendency of use is to employ *were* generally. By genre, *was* is preferred except in the web and academic text components of COCA. Although *were* was preferred in only two components, this was enough to overcome the preference for *was* in the other components as more than 85% of the wpm came from academic texts. *Do* is also generally preferred by English speakers to use with *data*. *Do* is preferred in web texts, magazines, news, and academic texts while *does* is used more in blogs, TV/movies, spoken texts, and fiction. Finally, English speakers tend to use *has* more than *have* with *data*. *Has* is preferred across all genres except academic texts.

The case of *data* is interesting as it is the noun which can be considered as the prototypical noun with Latin plural ending whose subject-verb agreement is ambivalent. General patterns indicate that use is divided between "plural" and "singular" primary verbs. Upon close inspection of the wpm data, it can be noticed that "singular" nouns are preferred in a wider range of genres compared to their "plural" counterparts. However, this is not enough to decisively skew the wpm data in favor of "singular" nouns because a substantial portion of tokens is from academic texts where users have consistently followed the prescriptive rule. Like the reasons enumerated above, the formality of academic texts might be the reason for the preference to use "plural" verbs with *data* (Huang, 2011). Overall, the wpm data for *data* show that is preferred to be treated as a singular noun in other genres, but it is still treated as plural in academic contexts.

- ***Reclassified Nouns***

Table 9. **Agreement patterns of *media* across genres**

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Media (are)	1801	346	299	35	463	26	163	211	258
Media (is)	4512	1487	1084	82	1026	21	281	262	269

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Media (were)	581	76	73	13	102	24	83	80	130
Media (was)	897	193	152	18	195	19	104	95	121
Media (do)	244	45	47	6	75	2	21	32	16
Media (does)	462	133	107	9	146	2	27	27	11
Media (have)	1302	203	184	11	257	10	183	269	185
Media (has)	2173	623	439	40	542	15	178	184	152

Table 10. **Agreement patterns of *insignia* across genres**

	All	Blog	Web	TV/Movie	Spoken	Fiction	Magazine	News	Academic Texts
Insignia (are)	7	0	2	0	0	0	1	0	4
Insignia (is)	18	4	5	2	0	2	2	2	1
Insignia (were)	4	0	2	0	0	1	0	0	1
Insignia (was)	6	0	3	0	0	2	0	1	0
Insignia (do)	0	0	0	0	0	0	0	0	0
Insignia (does)	0	0	0	0	0	0	0	0	0
Insignia (have)	0	0	0	0	0	0	0	0	0
Insignia (has)	2	0	1	1	0	0	0	0	0

The final group of nouns that demonstrate a unique pattern of subject-verb concord is composed of *media* and *insignia*. It can be seen in the wpm data that users generally prefer to treat these nouns as singular despite being notionally plural. In the case of *media*, it has the general tendency to take the "singular" *be* (both in present and past), *do*, and *have*. By genre, English speakers use *media* with "singular" *be* (present and past), *do*, and *have* in the samples from blogs, web, TV/movie, and spoken texts. There is also a trend towards "singular" verbs in magazines and news with some exceptions (e.g., *have* for magazines, and *do* and *have* for news). For fiction and academic texts, however, we can see that there are equivocal agreements with *media*. In fiction, users prefer to use "plural" *be* and *has* with media. In academic texts, we can see that *media* is not

consistently treated as plural, with users preferring to use *is* with it together with the other "plural" verbs. The case of *insignia* is less straightforward as the wpm data from the corpus are limited. Despite this, one could still see some patterns although the strength of these observations is limited. First, in genres where data for *insignia* are available, one could see that the consistent preference is to use the "singular" forms of verbs. The notable exception to this pattern is in the case of academic texts, where the preference is to use "plural" *be* verbs. There are no data for *do* and *have*.

Based on the wpm data presented in Tables 9 and 10, one might be tempted to say that *media* and *insignia* have already been classified as singular nouns. At the very least, the process for their reclassification has already been commenced. If one arrays these seven nouns along a continuum, these two nouns are the closest to the end where nouns are already treated as singular. This conclusion, although still tentative, is supported by the fact that the overall data (aggregate of the data from all genres) demonstrate a strong preference for the use of "singular" verbs for *media* and *insignia*, whenever data are available for the latter. Across blogs, the internet, TV/movies, and spoken texts, the preference is to treat *media* and, where data are available, *insignia* as singular. There is a little ambivalence, with notable preferential advantage to "singular" verbs, when it comes to usage in the spoken, magazine, news, and even academic text components of the corpus. The latter is especially interesting as one can expect writers to adhere more to traditional subject-concord rules in formal contexts such as academic texts (Chafe & Danielwicz, 1987). Like the explanations proffered for the first group of nouns, the singular treatment of *media* and *insignia* might have been due to the reclassification of these nouns in the minds of English speakers. This reclassification can be because the plurality of these nouns is not encoded morphologically unlike other English nouns such as *boy-boys* as explained by Gardelle (2016). Because notional and morphological plurality can be a source of confusion (Humphreys & Bock, 2005), it is not unsurprising that English users reclassified these nouns as singular despite their original plural meaning. Since these nouns are generally abstract and their numerosity is not registered in the minds of speakers, one might be observing "singular override," the reverse of what Huddleston and Pullum (2002) call plural override. The strength of this reclassification is demonstrated by the fact that these nouns are treated as singular generally across genres (written and spoken) and formality levels. This "singular override," which Quirk et al. described as recently as 1985 as a widely condemned usage, seems to have become the fact among English speakers of today.

- ***Implications for language teaching***

The collected data for this study show that the subject-verb agreement of these nouns with Latin plural endings is not as straightforward as language manuals and stylebooks would like us to believe. This has implications for how these nouns should be discussed or introduced in the language classroom and in language teaching materials. Language teachers ought to avoid sweeping "rules" on how these nouns behave in subject-verb concord as their behaviors vary depending on the genre, context, and level of formality as this investigation demonstrates. Instead of airtight "rules," teachers can provide students with explanations on how these nouns behave in different genres and contexts, in line with the general prescription for grammar teaching given by Celce-Murcia and Larsen-Freeman (2008). These explanations might be based on the level of formality or text type. For example, the data suggest that the more formal the context is (e.g., academic texts) the more conservative English speakers can be in following subject-verb agreement. In the classroom, ESP teachers can highlight these tendencies to their students writing their research papers or technical reports where certain words are regularly employed. The same tendencies can also be explained to students in journalism classes where certain articles might call for the use of these nouns. Doing this ensures that students are taught how to use these nouns with

Latin plural endings in the same way they are used by individuals in their target discourse communities, thus helping them ease into these communities as new members. In addition, language classrooms should continuously strive to employ authentic, real-world examples of language use to show how speakers of English actually utilize the language. This can then be used by teachers to demonstrate the dynamic relationship between language and social contexts and how they impact each other. For example, teachers can show students a blog post and a magazine article with enhanced nouns with Latin plural endings. Teachers can make students analyze the purpose, audience, topic, and context of each text and how these might have contributed to the SV concord patterns they employ. Teachers can further this aim by allowing students to analyze texts from various fields and English-speaking communities and see the differences in how these domains employ language and structure their texts. This can enhance student language awareness, command of lexicogrammatical patterns, appreciation of the role of context in these patterns, and critical understanding of grammar, language variety, variations in meanings, and patterns of use (Kayaoğlu, 2013; Liu & Jiang, 2009; Römer, 2011). By combining discourse and corpus work, teachers can reconcile top-down and bottom-up approaches to teaching writing, similar to what Charles (2007) did. In materials development, textbook authors can use more corpus-drawn insights in writing their materials to ensure that these reflect real-world use of the language and to address the need to revise pedagogical language descriptions which have been found to be mismatched with naturally-occurring English (Römer, 2010). By ensuring that classroom lessons and materials reflect authentic use of the language, teachers hopefully will be able to arouse the interest and motivation of students to learn the target language by demonstrating clear classroom-real world connection.

CONCLUSION

The corpus data gathered for the seven nouns with Latin plural endings show that their agreement behaviors can allow their classification into three groups: 1) nouns whose tendencies are line with prescriptive rules; 2) noun whose tendencies are ambivalent; and 3) nouns which might have already been "reclassified" by speakers. *Bacteria*, *phenomena*, and *criteria*, and to some extent *algae*, belong to the first group. Wpm data on these nouns demonstrate that English speakers still prefer to treat them as plural nouns. By genre, "plural" verbs are still preferred to be used with these nouns with some exceptions in the TV/movie and spoken components of COCA. The preference for "plural" verbs is more strongly manifested in the written components of the corpus especially academic texts while a notable preference for "singular" verbs has been observed among the spoken components of the corpus. The preference to follow the prescriptive rule is also seen in *algae* although it must be noted that this strength by genre is not as strong as that observed in the first three nouns. Prescriptive rule is followed, in general, when using *algae* in some formal and written genres such as academic texts, fiction, and magazine. The noun that demonstrates the highest equivocation in terms of subject-verb agreement is *data* as shown by its overall wpm data. Preference across genres is to use "singular" verbs, but this is negated by the number of tokens from academic texts where preference is still to use "plural" verbs. These data suggest that *data* is already slowly being classified as a singular noun by general English users and is being treated plural only in more formal, academic situations. Finally, two nouns have shown evidence suggesting that they might have already been reclassified by English speakers: *media* and *insignia*. Another possible reason for this singular treatment is singular override and the fact that these cannot be split into *stem + plural inflection* (Gardelle, 2016). Even in academic texts, which are traditionally more conservative in their subject-verb concord, one can observe that users equivocate in using *media*. More data, however, are needed for *insignia* although the little data available on this noun show singular treatment.

The equivocal treatment of some of these nouns can be traced to structural and contextual reasons. The plurality of these nouns with Latin plural endings is encoded semantically. This means that their being plural is notional and not morphological, a source of error or confusion among speakers (Humphreys & Bock, 2005). Hence, upon initial glance, one would not see any plural marking an English speaker can expect in regular nouns such as *animal*. As Gardelle (2016) explained, speakers cannot split these nouns with Latin plural endings into *stem + plural inflection*. This lack of morphological plural marking removes one clue that can help English speakers to process their meaning and treat them as plural especially in an online processing of the language where speakers do not have the time to reflect on the sentences they will utter just like in oral discourse (Chafe & Danielwicz, 1987). This might lead English speakers to reclassify some nouns and to be affected by singular override. Another possible factor for their equivocal agreement might be in the nature of the components of the corpus. The components can be generally divided into written and spoken. Although with exceptions, the data show that the written, formal genres exhibit more conservative treatment of these nouns as it is expected that writers will have time to reflect on and follow prescriptive rules in what they are writing (Chafe & Danielwicz, 1987). For example, a strong tendency to use "plural" verbs is observed in using these nouns in academic texts. The less formal and spoken genres meanwhile demonstrate equivocal agreement to preference for "singular" verbs in their use of these nouns.

The findings of these study have several implications for language teaching and research. Primarily, the results of this investigation show that language teaching and materials development should ensure that corpus-based insights are used to inform their preparation and development. A more nuanced explanation of subject-verb agreement tendencies should be provided by teachers, taking into consideration various factors such as context, register, and genre. Students can also be made to explore readily available data to heighten their linguistic sensitivity and awareness of the dynamic nature of what is traditionally presented as grammar "rules." On the research front, this paper suggests that other problematic subject-verb agreement structures be investigated using corpora to ensure that insights on these are based on how speakers actually use the language. Genre-based differences in the linguistic behavior of these structures may also be investigated. These structures may also be analyzed from the perspective of local English varieties as these may point to possible unique features of these dialects.

REFERENCES

Barbieri, F. (2005). Quotative use in American English: A corpus-based, cross-register comparison. *Journal of English Linguistics*, *33*(3), 222-256.

Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly, 35*(2), 331-336.

Celce-Murcia, M., & Larsen-Freeman, D. (2008). *The grammar book*. Thomson Learning Ltd.

Chafe, W., & Danielwicz, J. (1987). Properties of Spoken and Written Language. Technical Report No. 5.Center for the Study of Writing, California University, Berkeley.

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for academic purposes*, *6*(4), 289-302.

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century?. *Tesol Quarterly*, *34*(3), 548-560.

Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System*, *27*(1), 1-18.

Davies, M. (2013). Google Scholar and COCA-Academic: Two very different approaches to examining academic English. *Journal of English for Academic Purposes*, *12*(3), 155-165.

Dita, S. (2014). The semantics and grammar of adverbial disjuncts in Philippine English. In M.L.S. Bautista (Ed.), *Studies in Philippine English: Exploring the ICE-Philippines* (pp. 33-50). Vibal Publishing.

Estling Vannestal, M., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, *19*(3), 329-350.

Gardelle, L. (2016). Lexical plurals for aggregates of discrete entities in English: why plural, yet non-count, nouns?. *Lingvisticae Investigationes*, *39*(2), 355-372.

Grant, L. E. (2007). In a manner of speaking: Assessing frequent spoken figurative idioms to assist ESL/EFL teachers. *System*, *35*(2), 169-181.

Huang, L. S. (2011). Corpus-aided language learning. *ELT journal*, *65*(4), 481-484.

Hsu, W. (2018). The most frequent BNC/COCA mid-and low-frequency word families in English-medium traditional Chinese medicine (TCM) textbooks. *English for Specific Purposes*, *51*, 98-110.

Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the English language*. Cambridge University Press

Hudson, R. (1999). Subject-verb agreement in English. *English Language and Linguistics*, *3*(2), 173-207.

Humphreys, K., & Bock, K. (2005). Notional number agreement in English. *PsychonomicBulletin & Review, 12*(4), 689-695.

Hunston, S. (1995). Grammar in teacher education: The role of corpus. *Language Awareness, 4*(1), 15-31.

Hunston, S., & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics, 19*,(1), 45-72.

Hussein, R.F., & Sawalha, M. (2016). A corpus-based study of similes in British and American English. *Arab World English Journal, 7*(2), 49-60.

Huo, Y. (2014). Computer aided design of Chinese college English teaching materials based on COCA corpus. *Applied Mechanics and Materials, 590*, 916-919.

Jong-Bok, K. (2012). English *get*-passive constructions: A corpus-based approach. *Studies in Generative Grammar, 22*(2), 437-457.

Kayaoğlu, M.N. (2013). The use of corpus for close synonyms. *The Journal of Language and Linguistic Studies, 9*(1), 128-144.

Keeney, T., & Wolfe, J. (1972). The acquisition of agreement in English. *Journal of Verbal Learning and Verbal Behavior*, *11*, 698-705.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31*, 25-35.

Liu, D. (2011). Making grammar instruction more empowering: An exploratory case study ofcorpus use in the learning and teaching of grammar. *Research in the Teaching of English, 45*(4), 353-377.

Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly, 37*(4), 671-700.

Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal, 93*, 61-78.

McEnery, T., Wilson, A., & Barker, P. (1997). Teaching grammar again after twenty years: Corpus-based help for teaching. *ReCALL 9*(2), 8-16.

Okamoto, M. (2015). Is corpus word frequency a good yardstick for selecting words to teaching? Threshold levels for vocabulary selection. *System, 51*, 1-10.

O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly, 37*(3), 389-418.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *The comprehensive grammar of the English language*. Longman.

Pham, Q. H. P. (2022). A Corpus-Based Approach to Grammar Instruction in IELTS Writing Classes. *RELC Journal, 53*(3), 723–730.

Ma, Q., Tang, J., & Lin, S. (2022) The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration, *Computer Assisted Language Learning, 35*(9), 2731-2760.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205-225.

Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present, and future. In M.C. Campoy-Cubillo, B. Belles-Fortuno, & M.L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18-35). Continuum.

Samburskiy, D. (2014). Corpus-informed pedagogical grammar of English: Pros and cons. *Procedia – Social and Behavioral Sciences*, 263-267.

Skalicky, S., Berger, C.M., & Bell, N.D. (2015). The functions of "just kidding" in American English. *Journal of Pragmatics, 851,* 18-31.

Yusu, X. (2014). On the application of contemporary American English in vocabulary instruction. *International Education Studies, 7*(8), 68-73.

Zhao, Q. (2022). A Study on the Use of Hedges in English Writing of College Students at Different Levels. *Theory and Practice in Language Studies, 12*(10), 2211.