A Corpus Analysis of *Girl* and *Boy* in Spoken Academic English and Teaching Activities to Raise Awareness About Gendered Discourse

*Reina Hirose Horton**
ELS Honolulu

Abstract
This corpus analysis of spoken academic English investigates the use of gendered language in higher education, particularly focusing on the terms *girl(s)* and *boy(s)*. I will first review previous corpus research on certain gendered terms in various contexts. Then, I will conduct an analysis of two corpora which specialize in spoken discourse in academic settings, MICASE and BASE, analyzing target word frequency, sentential context, and larger context, such as speaker's gender and the type of course in which utterances took place. My analysis confirms the findings of previous research and extends such research by indicating that the use of gendered terms is also prevalent in spoken academic discourse. Drawing from the insights of the study, I go on to provide a set of ESL/EFL teaching activities designed to raise awareness of gendered language and its effects in the ESL/EFL classroom.

Introduction

The expression of gender in language is a topic that has received much attention over the years, and corpus analysis has made significant contributions to this field. By providing statistical analysis of data, the corpus approach can "serve to corroborate the findings of a more impressionistic approach, to confirm—or disconfirm—hunches, and to suggest new directions for further interrogation of the texts themselves" (Thornbury, 2010, p. 280). In other words, corpus analysis can provide the factual backbones to what researchers have claimed based on their experiences or intuitions and can lead them to new findings that they did not expect. The purpose of the present article is to investigate the use of gendered language in spoken academic discourse by using academic English corpora. I first review the findings of previous corpus-based studies on the issue of gender and language. Then, I describe the research that I conducted and discuss the results. Finally, based on my findings, I introduce some teaching activities that can be applied to the ESL/EFL classroom in order to raise awareness of gendered language in the ESL/EFL classroom.

Corpus Analysis of Gendered Language

Most of the recent research on the issue of gender and language focuses on four categories of gender in language: grammatical gender, lexical gender, referential gender, and social gender (Fuertes-Olivera, 2007). Grammatical gender refers to a grammatical feature of a language that specifies gender, e.g., masculine nouns and feminine nouns in Spanish. Lexical gender refers to the words that carry a gendered semantic property, such as *sister* and *father*. Reference gender refers to linguistic expressions that involve the non-linguistic reality associated with gender, e.g., *prostitute* does not linguistically specify gender but generally refers to a woman.

Horton, R. H. (2018). A corpus analysis of *girl* and *boy* in spoken academic English and teaching activities to raise awareness about gendered discourse. *TESOL Working Paper Series, 16*, 1-18.
Website: Hawaii Pacific University http://www.hpu.edu
* Email: reinahirose@gmail.com. Address: ELS, 1188 Fort Street Mall, Honolulu, HI 96816, USA.

Social gender is related to socially imposed stereotypical gender roles and characters. For example, *nurse* is typically followed by pronoun *she*, whereas *surgeon* is followed by *he*. As this study is concerned with the behavior of *girl* and *boy* in academic settings, much of the following survey of the literature will be confined to behaviors of lexical gender. In this section, I will review the findings from the previous corpus-based studies on three sets of gendered lexical items: 1) *man/woman*, 2) *girl/boy*, and 3) *Mr./Miss/Mrs./Ms.* Then, I will discuss the potential of specialized corpora to provide richer contextual information which enables deeper analysis.

Studies on *woman/man*

One of the most common topics of previous studies in this field is frequencies of the grossest terms used to classify gender, the lemmas *man* and *woman*. In their research on five corpora with a million-word data size from three countries, Sigley and Holmes (2002, as cited in Pearce, 2008) studied the comparative frequency of *man/men* and *woman/women*. They found that the reference to plural *women* in writing doubled between 1961 to 1991. Also, the ratio of references to plural *woman* compared to that of singular *man* went up from 1:5 in 1961 to 1:2 in 1991. These results may suggest that the visibility of women in the society raised in that time period. However, references to women as individuals are still fewer than references to men as individuals. Moreover, while the increase in the frequency of *woman/women* in the corpora appears to be a positive sign for gender equality, singular *woman* is less frequent than plural *women* (Holmes & Sigley, 2001; Tylor, 2013). This reflects the tendency for women to be addressed as collectives, while men receive more attention as individuals. Further, the increasing reference to *woman* is not necessarily made in a positive context. Holmes and Sigley (2001) spotted several derogatory references to female Prime Ministers in the Freiburg–LOB Corpus of British English, describing women Prime Ministers using words such as "troubling" and "absolute tragedy." Thus, simple frequency data do not tell the whole story; one needs to look further into the context to see how a word is being treated.

An analysis of the adjectives and verbs that collocate with the lemmas *man/woman* shows that there is a general tendency for men to be represented as strong (*stocky, climb, dig*) while women are represented as weak (*vulnerable, abuse, oppress*). Men tend to be portrayed as subjects exercising forms of power, such as legal execution, violence, and ownership. Women, on the other hand, tend to be depicted as the objects of power, particularly related to sexual violence, limitation, and categorization (Pearce, 2008). Gesuato (2003) claimed that lemmas *man* and *woman* tend to occur in complementary distribution; women are typically associated with domains of physical attractiveness, civil rights, religion and involuntary actions, whereas men are associated with non-physical attractiveness, violence, the military, and voluntary actions. Also, it has been found that while men are prolifically categorized in terms of status, capacity, and behavior, women's categorizations are commonly based on physical appearance and sexuality (Caldas-Coulthard and Moon, 2010; Pearce, 2008).

The prepositional phrases that are headed by *of*, such as "he is a man of humor" or "she is a woman of intelligence," can be found after both genders. However, only the referent of *woman* is depicted as a man's 'property,' as in "He's the sort that needs a woman of his own." On the other hand, the referent of *man* is largely described as related to his work, as in "the curator of the museum had to be a man of the pen and of the book" (Gesuato, 2003). This contrast clearly exemplifies the underlying social conception of men as active participants in public realms such as business and politics, while women belong to the private

realm, dependent on and taken care of by men.

One unique perspective to illustrate the gender bias in language is 'male firstness' (Freebody and Baker, 1987, as cited in Baker, 2014). When the speaker or the writer tries to be inclusive by mentioning both sexes, connecting two gender forms with *and* or *or*, generally the male forms are given the precedence. For instance, in the Corpus of Historical American English (COHA), *men* is 263 times more likely to precede *women* than the reverse. This is also true for *man* and *woman*, *boy(s)* and *girl(s)*, *male(s)* and *female(s)*, and *he* and *she*. The only exception is *ladies* and *gentlemen*, in which *ladies* are seven times more likely to precede *gentlemen* (Baker, 2014).

Another aspect of male bias in language is the generic *man*, that is, usage of *man* referring not only to males but also to females but to human beings in general. Fuertes-Olivera (2007) discovered that 79 out of 208 instances of *man* in the Corpus of Business English were generic. Looking into COHA, Baker (2014) found that after the peak in the early nineteenth century, there has been an overall decline in the use of generic *man*. In the last decade of the twentieth century, only 6% of the use of *man* were generic. Job and role titles such as *spokesman* and *chairman* contain generic *man*. However, while gender non-specific versions of these, e.g., *spokesperson* and *chairperson*, have been introduced, these are typically used exclusively for women, ironically continuing to mark them with gendered language.

Studies on *girl/boy*

It is natural to assume that *girl* and *boy* refer to male children and female children respectively, but research shows that this is not necessarily true. In spite of this intuition, *girl* is often used to refer to an adult female. Indeed, the use of *girl* to refer to an adult female is so normalized that it is often not perceived as insulting, while use of *boy* to refer to an adult male would constitute an insult (Caldas-Coulthard & Moon, 2010; Tyler, 2013). From their study on a million-word corpora from New Zealand, the United States, and the United Kingdom, Sigley and Holmes (2002) found that *girl* is three times more likely to refer to an adult than *boy*. The adjective *young* commonly precedes both *girl* and *boy*, but it is used more frequently with *girl*. Virtually interchangeable with *young woman*, *young girl* generally refers to a teenager or a woman in her twenties, while *young boy* is certainly not an adult. Girls in their twenties can be 'young' because the word *girl* encompasses much older women. Even though the usages that refer to an adult woman as *girl* may be a way to express affection by some speakers, it should not be ignored that they may have the negative effect of belittling women.

One typical situation in which the reference to a woman with *girl* can be observed is the workplace. In the workplace, where status is important, using *girl* to refer to an adult woman can be taken as denial of her competence and independence. For instance, when *boy* is a word part used to describe an occupation in the workplace, virtually all cases refer to low-status, entry-level occupations which could be carried out by teenagers (*delivery boy*, *paperboy*, etc.). Usage of *boy* is age-specific, and it is unlikely that an adult man would be called *delivery boy*. Similarly, *girl* in the workplace is used to refer to a female worker in a subordinate status; however, in this case, the age does not matter. A woman referred to as an *office girl* can be in her teens or can be eighty years old. Such girls are "not allowed to 'grow up' to take on more responsibilities (office woman? shop woman?), and so these labels provide a linguistic version of the 'glass ceiling'" (Holmes & Sigley, 2002).

Further, the lemma *girl* tends to collocate with words that are related to sex, sexuality, and physical appearance (Caldas-Coulthard & Moon, 2010, Taylor, 2013). While both

gendered terms are associated with some aspects of sexual relationships, this tendency is stronger for *girl* than *boy*. For instance, the collocations of *boy* that refer to sexual relations, *molest*, *rape*, *sexual*, *love*, and *marry*, also collocate with *girl*. However, there are a number of sexual words that collocate with *girl* but not with *boy*, such as *pubescent*, *naked*, *single*, *pregnant*, *prostitute*, and so on. Girls' physical appearances are described with words such as *beautiful*, *good-looking*, *pretty*, *gorgeous*, *sexy, stunning*. From these collocations related to sexuality and appearance, a strong tendency to associate women via the label 'girls' with sexual appeal can be observed. Moreover, Sigley and Holmes (2002) argued that *girl* is often used interchangeably with *girlfriend*, as in the phrase "I always hoped for a girl like you." This use of *girl* reflects the notion that the women are viewed as an object of male desire.

By analyzing the verbs that take *girl* and *boy* as arguments, Baker (2014) discovered that girls are more likely to be represented as expressing certain emotions, feelings or cognitive states (*smile*, *want*, *suffer*, *love*, *decide*), while boys are depicted as being the subjects of physical actions or states (*grow*, *play*, *fall*, *die*). Turing to the verbs for which *girl* and *boy* appear as the object, girls are more often described as victims in a various ways (*rape*, *abduct*, *murder*, *assault*, *seduce*); while boys are described as being killed more frequently, they generally are not strongly associated with verbs that position them as victims.

Studies on *Miss/ Mrs./ Ms./Mr.*

The use of terms of address reveal a gender asymmetry from a different viewpoint. While all men are addressed by *Mr.* regardless of his marital status, for a woman, there are address options that are differentiated by her marital status, namely, *Miss* and *Mrs.* The inbuilt inequality of categorizing women according to their marital status and forcing them to reveal their marital status began to be questioned with the rise of feminist consciousness. The introduction of *Ms.* in the English speaking world in the 1970's "was intended to eliminate linguistic discrimination by providing a term for women which, like *Mr.*, did not signal marital status" (Holmes, 2001a, p. 119, as cited in Fuertes-O, 2007), and it was a flagship of feminist reform. Baker (2010) found that the frequencies of *Mr.*, *Miss*, and *Mrs.* have decreased over time, particularly since 1961. On the other hand, the frequency of *Ms.* has been increasing slowly but steadily. There were only 6 cases of reference to *Ms.* in 1991 in Baker's corpus study, but it's presence increased to 30 cases in 2006. However, it was revealed that decades after *Ms.* was introduced, its use is still far from prevalent. In the British National Corpus, *Ms.* was used in only 5% of the total references of *Miss/Mrs./Ms.* In New Zealand and Austria, young, well-educated women prefer refer to themselves as *Ms.*, but it remains an uncommon form of address chosen by others when referring to a woman (Romaine, 2001). The good news for advocates of gender equality is that the trend seems to be the gradual abandonment of gendered terms of address. If the current tendencies continue for several decades, all gendered titles may become rare in the future (Baker, 2014). In such a situation, the struggle to replace *Miss* and *Mrs.* with *Ms.* will be left to the past.

Studies of Gendered Lexical Items in Specialized Corpora

While the data size of a corpus is an important factor in supporting the reliability of search results, it is also true that in many large-size corpora data is not restricted to one domain, which makes it less ideal for a deeper analysis of phenomena in a specific domain. Hence, using smaller, more localized corpora which are compilations of texts of a specific domain may enable deeper interpretation (Thornbury, 2010).

In his research of lexical gender in the Wolverhampton Corpus of Written Business English (WBE), which is a specialized corpus concerned with Business English collected from 23 websites related to business, Fuertes-Olivera (2007) found results that counter those of Romaine (2001). Fuertes-Olivera found that *Ms.* was used more than nine times more frequently as the other two titles (*Miss* and *Mrs.*) combined in written business discourse, suggesting an increased willingness to use non-sexist language in business contexts. Nonetheless, *Mr.* occurs 14 times more frequently than all of the women titles put together, which is an even more striking gap than the results of general corpora. Address terms are used to show respect. The far higher ratio of *Mr.* to *Mrs., Miss, and Ms.* suggests that such markers of respect are often absent when addressing women, pointing to possible gender inequality in the business world.

Caldas-Coulthard and Moon (2010) conducted a study comparing gender representation in two corpora of different media: A British tabloid newspaper, *The Sun*, and the British newspapers *The Guardian, The Independent, The Times,* and *The Sunday Times*. Both types of media tended to describe women/girls with their appearance and sexuality in some way. However, they found that *girl* occurs three times more frequently in *The Sun* than in *British Broadsheets*. Moreover, while the sexually charged expressions described women in *British Broadsheets* were indirect and distanced, those of *The Sun* were more overt. These results reflect the contrast of two distinct discourse worlds about women in British print media: the "world where sexiness is associated with overt display, and one where it is implicit or allusive" (p. 109). This illustrates the range of ways by which women can be sexualized within similar domains.

Specialized corpora that collect data from a long time period enables us to analyze texts with ample contextual information and to compare results from different time periods. Macalister (2011) conducted diachronic research on how the description of gender roles have changed in writings for children. He used materials published by *The New Zealand School Journal,* which has been provided to schools in New Zealand since 1907, and investigated the usages of lemmas *boy* and *girl* in four indicator years with thirty-year intervals throughout the twentieth century. He found that until the third indicator year (1969-1970), *boy/s* occurs three times more than *girl/s*; however, in the final indicator year (1999-2000), the numbers are roughly equal, with *girl/s* slightly exceeding *boy/s*. This reflects the relative success of feminism in the New Zealand started in the 1970's. However, he also found that even in the final indicator year, girls are described as doing less than boys. In other words, girls are less likely to be portrayed as active agent compared to boys.

The Cringe Text Corpus is a small corpus of teenage written narratives taken from online teenage magazines. By analyzing the discourse characteristics of the texts in this specialized corpus, Thornbury (2010) discovered a distinct discourse pattern in teenage narratives. Most of them followed the same pattern, which involved a girl encountering some unlucky event in front of the boy she likes and being embarrassed. Unlike the typical structure of narratives, the resolution part is completely absent. The reason for this can be explained from the perspective of feminist studies; by sharing stories of their misfortune, women try to elicit feelings of mutual empathy and affirm their joint femininity. Thornbury claims that by spreading such discourse practices which reproduce asymmetrical power relations in society, the teenage girl's magazines are "complicit in a process of discursively positioning their readership as the helpless and disempowered objects of male derision" while "a corpus search provides evidence of the 'objectification' of the protagonist" (p. 281).

Researchers have so far studied the use of gendered terms in corpora of various era, countries, and registers. In particular, the results of Fuertes-Olivera's (2007) study on written business language corpus showed that while the underlying 'Male-As-Norm' principle (evidenced by the use of the generic *man* and the disproportionate use of *Mr.* compared to address terms used for women) was found in the corpus, the use of sexist language is still less prevalent in business discourse than in general language use. One possible explanation for a decrease in sexist language in business may be the advances in this field by women.

This paper builds on these studies by expanding the reach of corpus analysis of gendered terms to academic discourse, particularly spoken academic discourse. Spoken language is an especially fertile domain for studying gendered language and the attitudes supporting it because utterances cannot be revised after they are made, unlike sentences produced in written language. I will limit my focus in the following to the lemmas *girl* and *boy*.

Research Questions

1) How are the lemmata *girl* and *boy* used in spoken academic discourse? Is there any gender bias that is observable?
2) Are the lemmata *girl* and *boy* used differently in academic English when compared to their use in general English?

Methodology

For the present study, I used two corpora: MICASE (Michigan Corpus of Academic Spoken English) and BASE (British Academic Spoken English Corpus). MICASE (https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple) is an online corpus of spoken academic discourse created by the University of Michigan. There are currently 152 transcripts from diverse academic disciplines available, and its size is about 1,800,000 words.

Search results can be specified by speaker attributes (gender, age, academic position/role, native/non-native, and first language), and transcript attributes (speech event type, academic division, academic discipline, participant level, and interactivity rating). While users can search by specific attributes of the transcripts, only a simple word search is available; no lemma search and no collocation search are available. Still, it is a useful corpus for analysis since full context is available.

BASE is the result of a project that took place at the University of Warwick and Reading between 2000 and 2005, and it is also available online (https://warwick.ac.uk/fac/soc/al/research/collections/base/). It consists of transcriptions of 160 lectures and 40 seminars recorded in a variety of departments, and the data size is about 1,600,000 words. BASE can be analyzed using Sketch Engine, a software for corpus query and management, and parts of speech are tagged in the corpus.

Findings

Frequency

Table 1 shows the frequencies of the lemmata *girl* and *boy* within MICASE. A point which deserves attention is the high frequency of *girls*. While the frequency of *girl* and *boy* is relatively low, *girls* is used three times more frequently than *boys*. Although these results appear to correspond with the findings of Holmes and Sigley (2001) and Tylor (2013), who found that women are more likely to be addressed in plural, a closer look reveals a different facet of this

issue. Among the 142 references to girls, 87 of them occurred in a speech event titled 'Women's Studies Guest Lecture,' while 10 of them occurred in 'Women in Science Conference Panel,' both of whose content are specifically about women.

Table 1
Frequencies of lemmata girl and boy in MICASE

word/lemma	matches (transcripts)
girl	89 (32)
girls	142 (24)
girl/girls	231(56)
boy	77 (54)
boys	53 (19)
boy/boys	130 (73)

Table 2 shows the frequencies of the same words in Table 1, but without the two speeches on women. In this case, the frequency of *girl* and *boy*, and that of *girls* and *boys* are surprisingly close. However, the word *boy* used as a part of phrases such as *oh boy* and *yeah boy*, that is, not referentially, accounts for 37 of the total 77 tokens of *boy*. Hence, the total number of referential uses of *boy* is significantly exceeded by that of *girl*. However, *girl* is used in spoken academic discourse to refer to adult females (see more details below). With the difference in frequency between *girl* and *boy* that emerges once non-referential uses of *boy* are excluded in mind, the substantially greater number of tokens of *girl* may indicate that even in spoken academic domains *girl* is often used to refer to adult women. This suggests an inequality in the way that women are positioned and portrayed in and through talk in these contexts.

Table 2
Frequency of lemmata girl and boy excluding two speeches on women in MICASE

Word	Matches (number of transcripts)	Referential usage (number of transcripts)
girl	73(30)	73(30)
girls	45(22)	45(22)
girl/girls	118 (52)	118 (52)
boy	77(54)	37 (28)
boys	49(17)	49(17)
boy/boys	126(71)	76(45)

The second corpus that I explored is BASE. I will start from looking into the simple search on the frequencies of the lemmas *girl* and *boy* (Table 3). The frequency ratio of each lemma is quite similar to those of MICASE (withholding the two speech events in MICASE whose topic was women), except for the fact that *girl/girls* are slightly more frequent than *boy/boys*. On the other hand, while the frequencies of singular *girl* and *boy* were higher compared to plural *girls* and *boys* respectively in MICASE, in BASE, the frequencies of the singular and plural forms of these words are almost the same, with only one non-referential use of *boy*.

Table 3
Frequency of lemma girl and boy in BASE

Word	Matches	Referential usage
girl	49	49
girls	35	35
girl/girls	84	84
boy	35	34
boys	30	30
boy/boys	75	74

Context of Use

Now I will look into the statistics of the attributes of the speakers and the transcripts. The total number of tokens of each word by speaker gender shown in Table 4 also shows interesting contrasts.

Table 4
Frequency of lemma girl and boy by the gender of the speaker in MICASE

Word/lemma	Female	Male
girl	70	19
girls	138	4
girl/girls	208	23
boy	41	36
boys	35	18
boy/boys	76	54

Regarding *girl/girls*, there is large disparity between the number of male and female speakers who token these words: women are highly predisposed to use the terms *girl* and *girls*. The imbalance of speakers' gender according to the use of *girls* is striking, with 138 of the

references made by women and only 4 of them made by men. Even without the two speeches on women's studies and women in science, the use of *girls* by female speakers is 41, which is still ten times more than that of male speakers. The fact that there are only 4 uses of *girls* by male speakers and 19 uses of *girl* by male speakers is striking. The following excerpts are the four instances that *girls* were mentioned by male speakers.

Excerpt 1

> Karen Deolavaris's um, work from David Hereford after he died and, yes the dissertation was on nice *girls* uh do fight, on female female fighting in high schools.

Excerpt 2

> in May so toward the end of April i decided there o- uh i was going to ask one of two *girls* whom i'd admired, probably for four years, without really, interacting with them at all, for a date.

Excerpt 3

> and female pros- professors also, yeah. are always a- nicer to the *girls*, cuz i feel like

Excerpt 4

> they just feel most of the guys really won't care and won't, do the stuff and *girls* are more likely to, complete the work and, i don't know but i've known, i think that male professors are much, nicer to females.

Excerpts (1) and (2) are from utterances made by senior faculties, and (3) and (4) are utterances made by undergraduate students. These excerpts are far from constructive discussions on gender issues; rather, they largely revolve around personal memories or complaints about female students being treated better than male students.

Unfortunately, I cannot investigate the collocations of these lemmas since MICASE does not have an advanced search function to enable collocate search, and a manual search of hundreds of word tokens and the calculations that would be required to determine if a word is a true collocate are outside the scope of this paper. However, looking into the specific contexts in which these lemmas are used as an adjective can yield deeper insights. One illuminating example is found in the same lecture that excerpts (3) and (4) are taken from. In (5), a male student refers to female professors as *girl professors*, while another student uses the term *male professors* rather than *boy professors* immediately afterward.

Excerpt 5

> S9: even the *girl professors* (do that)
> S11: but i maybe it's for like, the men, i don't know maybe that, they just feel most of the *guys* really won't care and won't, do the stuff and *girls* are more likely to, complete the work and, i don't know but i've known, i think that *male professors* are much, nicer to females.

Following Holmes and Sigley's (2002) claims, by using *girl* to refer to a professional woman, this male students belittled the woman.

On the other hand, female speakers often used *girls* to discuss social issues related to

women. Excerpt (6) was taken from "Women's Studies Guest Lecture," while (7) and (8) were taken from "Media Impact in Communication Lecture" and " Brazilian Studies Student Presentations" respectively, which are both subjects which might not prime participants to discuss social issues related to women.

Excerpt 6

> it all sounds a clear alarm, for American *girls* more so than boys adolescence is a time of declining confidence, silence, and danger.(Women's Studies Guest Lecture)

Excerpt 7

> whereas sexist *girls* may have a lotta friends because they're fitting in with this, you know a desired stereotype for them. (Media Impact in Communication Lecture)

Excerpt 8

> because some *girls*, particularly in high school, they feel, a bit, not that they feel harassed but, they, they feel that um, the presence of the guys, will no- not permit them, to participate as much. (Brazilian Studies Student Presentations)

Overall, the majority of references to these two lemmas are made by female speakers. A possible insight that can be drawn from these results is that male speakers in academic settings do not actively discuss gender as much as their female counterparts, at least not while using *girls*.

Girls* and *Guys

In addition to the pair *boys–girls,* the data suggest that another pair, *girls–guys,* deserves analytical attention (see Excerpt 5)). This observation of the data prompted me to compare the usage of *guys* and *girls* in the corpora. In MICASE, *girls* was mentioned 142 times in 24 transcripts while *guys* was mentioned 823 times in 103 transcripts, almost six times as much as *girls.* This is a significantly higher frequency compared to that of *boys*, which was merely 52. However, in 649 of 823 utterances, the word *guys* was part of the phrase *you guys* to refer to the students in the classroom. Assuming that most of the classes consist of both female and male students, this usage of *guys* can be considered as a variation of 'generic *man*' discussed above. The examination of the concordances lines of *guys* showed that when it is not used with the meaning of 'generic *guys*,' *guys* was used as a gendered term, as in "guys type their paper cuz they are sloppier" or "they just feel most of the guys really won't care and won't, do the stuff and girls are more likely to, complete the work."

Through an analysis of the BASE corpus using Sketch Engine's Sketch Difference function, which can contrast the usage of two words, I investigated how the two words *girl* and *boy* are used differently in this corpus. The results show that girls are more likely to be described in the role of something acted upon than boys: there are 22 uses of *girl(s)* in which girls are portrayed as being acted upon compared to 16 for *boy(s)*. This may indicate that girls are more likely to be depicted as a passive targets of actions or emotions than boys. In the concordance lines, girls were "liberated" and "frightened" by someone else. One the other hand, the number of uses of *girl(s)* and *boy(s)* which depict girls or boys as agents were equal. However, *girl* is used to represent girls as agents of actions such as maturing, breastfeeding, and feeding, whereas *boy* is used to represent boys as the agent of actions such as getting,

wanting, and choosing, which seems to indicate that girls are more likely to be portrayed as a provider of care or nurture while boys are portrayed as active individuals who take agentive actions.

Girl(s) are generally more frequently preceded by a modifier than *boy(s)*. Modifiers such as *teenage* and *old* (used as a part of the phrase "X-year-old") occurred before *girl(s)* but not before *boy(s)*, indicating girls are more likely to be specified regarding their age than boys. However, as Pearce (2008) mentioned in his study on the use of Sketch Engine for corpus analysis, focusing only on differences may lead one to overlook important similarities. The results of Sketch Difference show that the modifiers *little*, *young*, and *younger* are used commonly for both *girl(s)* and *boy(s)*, and their frequencies are higher than the other modifiers that occur only before one of the terms. Therefore, both *girl(s)* and *boy(s)* are preceded by modifiers which may denote a state of childhood, but only girls' ages are described more specifically. With Sigley and Holmes' (2002) study finding that *young girl* is often used to refer to adult females in mind, I searched for cases in which *young girl* was used to refer to adult women in BASE and MICASE, finding that such cases were entirely absent in BASE and very few in MICASE. This may suggest that tokening *young girl* to refer to a woman is a practice which is less frequent in academic domains.

In sum, once non-referential uses of *boy* are excluded from the data set, analysis showed that *girls* is three times more frequent than *boys*. An examination of the context in which *girls* occurs (e.g., Excerpts 1-8) along with the high frequency of *girls* suggests that it is widely used to refer to women in spoken academic discourse. Especially since the use of *boys* to refer to adult males did not reach parity (e.g., Excerpt 5), this finding indicates an inequality in the way that women are positioned and represented in and through talk. Comparing the findings to those gleaned from general purpose corpora, it would appear that this subtle form of gender oppression extends to classrooms at institution of higher education. Moreover, the analysis found that women are more than men likely to use *girl(s),* and when men use *girl(s),* it is often to derogate women. An examination of the context of *girl(s)* found that it is often used by women to discuss issues related to women. In a sense, the implications of this finding are reminiscent of those of Thornbury's (2010) study, which found that the elicitation of empathy from other girls through teenage narratives reproduces gender asymmetries of power. In this case, part of the linguistic means (use of *girl(s)*) through which women collaboratively question or confront forms of gender oppression (Excerpts 6-8) themselves may reinforce gender power asymmetries. The ubiquity of generic *guys* in the corpus dovetails with Fuertes-Olivera's (2007) study which showed the prevalence of generic *man*, pointing toward the presence of the hegemonic Male-as-Norm standard.

Analysis of BASE found that girls or women are often portrayed as passive targets of actions or emotions while boys are portrayed as active individuals who take agentive actions. This finding is parallel that of Pearce (2008), who found that while women tend to be depicted as objects of power through collocates of *woman/women*, men are depicted as subjects exercising power through the collocates of *man/men*. Finally, I found that referring to women as *young girls* was not prevalent in spoken academic discourse. Barring this exception, in total, the analysis indicates that the patterns of language use with respect to *girl(s)* and *boy(s)* in corpora of spoken academic English are strikingly similar to the type of gendered language found in general corpora and other specialized corpora in the sense that use of these terms simultaneously allow for a diagnosis of gender inequality and serve to perpetuate it.

Teaching Gendered Language

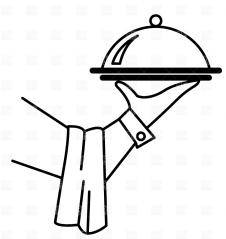
Now that I have looked into the usage of gendered terms *girl(s)* and *boy(s)* in Spoken academic discourse, I will propose some ESL/EFL activities that can be used to raise awareness of the existence and usage of gendered language for the students at the intermediate level. All of the sentences used in the activities (except for the gap filling in the Activity 3) are taken from corpora, which means they are authentic instantiations of target language features.

Activity 1: Stereotypes about Gender and Job Titles (60 minutes)

1. What are the job titles associated from the following pictures? Do those titles specify gender, or are they neutral?
2. In pairs, write down gendered job titles such as, *waitress*, *deliver boy*, etc. in the table below.

A

B



C

D

(images from clipartmart.com, rfclipart.com, 123rf.com, worldartsme.com)

Titles that are used only for female	Titles that are used only for male	→	Gender-Neutral Job Titles
		→	
		→	
		→	
		→	
		→	
		→	
		→	
		→	
		→	
		→	
		→	
		→	

3. Share the results as a class, and add new words to the table above.

4. Back in pairs, write down gender-neutral terms for each job title in the table.
5. Share the answers as a class. If there are new terms, add them to the table.
6. Discussion Questions:
    1. What do you hear more often, gendered job titles or gender-neutral job titles?
    2. Are there also gendered job titles in your first language?
    3. How would you feel by being called these gendered job titles?
    4. Why are gender-neutral job titles important?

Activity 2: Adjective collocations of *girl/boy* in different registers (60 minutes)

1. Corpus search
    1. Access the Corpus of Contemporary American English https://corpus.byu.edu/coca
    2. Click "**Collocates**", and type in *girl* or *boy*.
    3. In one row below, click **[POS]** to the right of Collocates, and select "**adj.ALL**".
    4. Go one row below, click "**2**" to the left, "**0**" to the right.
    5. Click "**Sections**" and choose the register that you want to look up for Choice 1, choose ignore for Choice 2.
    6. Click "**Find Collocates**" and browse the results.
    7. Write down your answer to the following two questions in the chart below.
        1. Which adjectives are frequently used with *boy/girl*? (Frequent collocations)
        2. Compare the collocations of *girl* and *boy* and write down the differences and similarities between them. Are there any findings that are interesting?

Register you chose: _____		
	boy	**girl**
Frequent collocations		
Interesting findings		

8. Share the findings with the class.
9. Compare the results with a person who chose a different register. Can you see any differences or similarities?
10. Discussion Questions:
    1. Why do you think certain adjectives are used for one gender and not for the other?
    2. If those adjectives are used for the other gender do they sound natural or not?
    3. Why is that?

Activity 3: Verb Collocations of *girl/boy*

1. Fill in the blank in the sentences below with verbs that you think suit them the most.

1. When I saw her, she was _____ in the park.
2. When I saw him, he was _____ in the park.
3. She was _____ all day long.
4. He was _____ all day long.
5. She was scolded by the teacher because she was _____ during the class.
6. He was scolded by the teacher because he was _____ during the class.
7. Her hobby is to _____.
8. His hobby is to _____.
9. She was _____ by a stranger.
10. He was _____ by a stranger.

2. In pair, compare the answer and discuss whether you can find any pattern in the verbs you used for each gender.
3. The following is a list of verbs that occurs commonly with *girl* and *boy* in real-world language. In pairs, discuss the similarities and differences of those verbs.

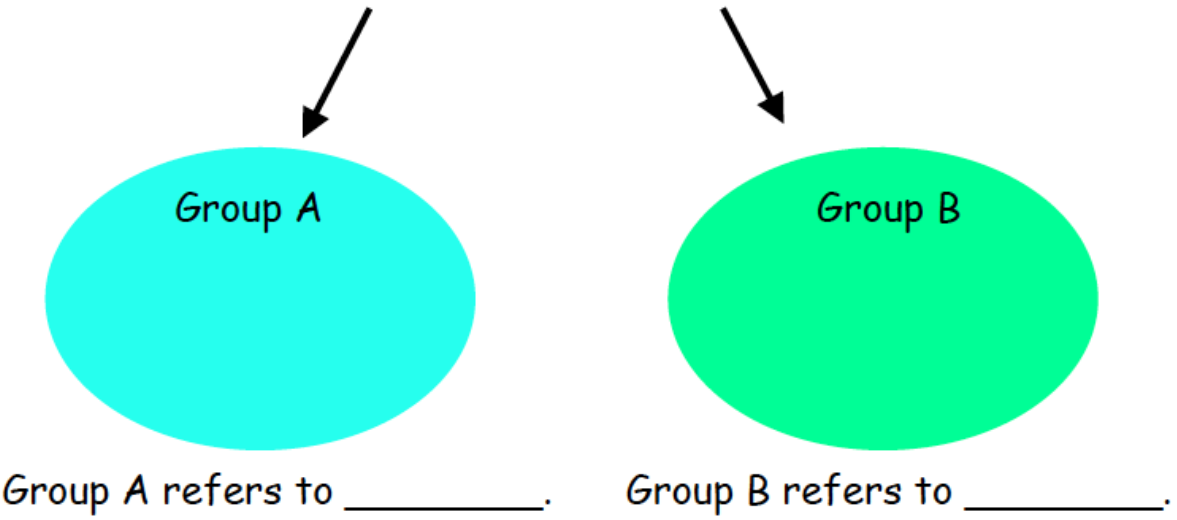
Verbs that collocate commonly with *girl*	**Verbs that collocate commonly with *boy***
marry, kiss	grow
sit, come	run, climb, play
wear, dress	stare

4. In pairs, discuss whether the differences you found from the list applies to the verbs you put in the blanks in the question you answered in the beginning.
5. Discussion Questions:
    1. Do you think there are gender stereotypes of what girls do and what boys do?
    2. Which verbs will you use to describe yourself?
    3. Do they match the tendency that we found above?
    4. What are the social consequences of this gendered language on children? on adults?

Activity 4: Generic Meaning of *man*

1. In pairs, read the following sentences and classify them into two groups, A or B, according the meaning of the word *men*, and write down the sentence number in the oval boxes below it.

1. All men are created equal.
2. there won't be any men in there
3. It is talk between men
4. men are too much trouble altogether
5. President Johnson
   had to send 3,900 soldiers and 100 FBI men
6. he would be one of the most important men in the world
7. He and his men had moved quickly forwards
8. Each man and woman was dressed like a terrible dream.
9. the city's firemen attacked the children with water
10. she did not understand about men



Group A refers to _____. Group B refers to _____.

1. Whom do *man/men* refer to in each group? Discuss the differences and write in down in the blank in the chart.
2. Share the findings with the class.
3. Discussion questions: Do you think there will be any negative effects in using man/men to refer to both man and woman?

Activity 5: Age range of *girls* and *boys*

1. What do you think are the age range of people referred to as *girl* and *boy*? Discuss it with your partner.
2. Read the sentences with the words *girl(s)* and *boy(s)* below. Guess the age of the *girl* and *boy* in each sentences in write it down in the blank.

3. In pairs, compare the answer and explain why you made that guess to each other.
4. Share the answers as a class.
5. Discussion questions:
    1. What are differences between how males and females are described according to the age?
    2. Why are they described that way?
    3. Do you describe males and females differently according to their ages?

Conclusion

In short, the findings of an analysis of the two corpora (MICASE and BASE) were similar to findings of past studies concerning the lemmas *girl* and *boy* in general English corpora. For instance, the usage of *girl* to refer to an adult professional woman that Holmes and Sigley (2002) reported was found in MICASE, and the higher tendency of *girls* to be the

objects of actions than *boys* that was mentioned by Taylor (2013) could be found in the BASE corpus.

One limitation of this study is that MICASE only has a simple search function. In discussing the increase of the uses of *girl(s)* in children's book, MacAlister (2011) suggested that one of the reasons behind it might be the increase of the woman writer. The finding of this study that the majority of the references involving *girl(s)* were made by female speakers supports her opinion. The general teaching approach advocated here has two important facets. First, it is geared toward consciousness raising of gendered terms with the intention that students will learn to recognize gendered language when they encounter it. This aspect of the approach focuses on student's receptive skills. Rather than accept gendered language which supports gender inequality, students are guided to challenge such language and to ask questions, such as "why is this written or spoken this way?" The second aspect ties into the communicative approach taken in the activities. By facilitating the production of language which supports gender equality, students are given the space to practice actively challenging gendered language and the power imbalances which are reflected in the language. My hope is that through these activities both girls and women will learn to take a critical eye to the gendered language they encounter and to take the lead in this aspect of the struggle for gender equality. It is not easy to achieve gender equality in the world, but at least one thing that can be said with confidence: the more women speak or write in public, the more women will be visible.

References

Baker, P. (2010). Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender & Language*, *4*(1).

Baker, P. (2014). *Using corpora to analyze gender*. A&C Black.

Caldas-Coulthard, C. R., & Moon, R. (2010). 'Curvy, hunky, kinky': Using corpora as tools for critical analysis. *Discourse & Society*, *21*(2), 99-133.

Fuertes-Olivera, P. A. (2007). A corpus-based view of lexical gender in written Business English. *English for Specific Purposes*, *26*(2), 219-234.

Gesuato, S. (2003, March). The company women and men keep: what collocations can reveal about culture. In *Proceedings of the corpus linguistics 2003 conference* (pp. 253-262).

Holmes, J., & Sigley, R. (2001). What's a word like girl doing in a place like this? Occupational labels, sexist usages and corpus research. *Language and Computers*, *36*(1), 247-263.

Macalister, J. (2011). Flower-girl and bugler-boy no more: changing gender representation in writing for children. *Corpora*, *6*(1), 25-44.

Pearce, M. (2008). Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine. *Corpora*, *3*(1), 1-29.

Romaine, S. (2001). A corpus-based view of gender in British and American English. *Gender Across Languages: The linguistic representation of women and men*, *1*, 153-175.

Sigley, R., & Holmes, J. (2002). Looking at girls in corpora of English. *Journal of English Linguistics*, *30*(2), 138-157.

Simpson, R. C., S. L. Briggs, J. Ovens, and J. M. Swales. (2002) *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora*, *8*(1), 81-113.

Thompson, P. and Nesi, H. (2001) The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research* 5 (3) 263-264

Thornbury, S. (2010). What can a corpus tell us about discourse. *The Routledge handbook of corpus linguistics*, 270-287.

About the Author:

Reina Hirose Horton received an MA in TESOL (Teaching English to Speakers of Other Languages) from Hawaii Pacific University. She teaches English and Japanese to adult and young learners. Her current interests include motivation in language acquisition and the connection between language and identity.