

# Gender and Usage of Discourse Markers in Spoken Korean

Hyeyoung Kim<sup>1</sup> and Beommo Kang<sup>2</sup>

<sup>1,2</sup>Korea University

simpson15@korea.ac.kr, bmkang@korea.ac.kr

## Abstract

Now that discourse markers frequently appear in spoken languages, it is important to study whether there is a gender difference in the usage of discourse markers. We retrieved 300 ~ 2000 words from each of the 65 speakers (men and women) out of Sejong Spoken Corpus. We have tried to sample it with around 185,000 words so that we can get a more balanced corpus with respect to registers and gender. The result shows that women respond more emotionally than men, and women use more discourse markers than men.

## Keywords

Gender, frequency, keyness, corpus analysis, private speech

## Introduction

One of the factors that influence conversations includes the gender of interlocutors and speech contexts. The role of gender in language has often been used to refer to the research of linguistics for years. This study investigates how the two factors affect the usage of discourse markers by analyzing keywords in the spoken corpus and gender.

## 1 Gender issues

Since Lakoff (1975), a lot of research has been conducted on the difference between men and women's language usage. The study having enormous influence in the differences of gender is Lakoff (1973, 1975). Lakoff maintains women's speech as introspective point of view with her instinct. Lakoff points out that women use far more intensifiers, tag questions, interjections, interruption than men *etc.* However her explanations have problems which are not the empirical methods supporting data. It has reported that women use more tag-questions, exclamatives, or degree adverbs than men, especially in the informal speech context (Graddy 2006, Coates 1998). Now that discourse markers frequently appear in spoken languages, it is important to study whether there is a gender difference in the usage of discourse markers.

## 2 Keywords

Keywords give us to provide a method for identifying in texts. This method identifies items of unusual frequency in comparison with a reference corpus (Scott & Tribble 2006). So, keyness is a quality words may have in a given or target texts, suggesting that they are important, they reflect what the text is really about, avoiding insignificant detail. Finding keywords requires a wordlist of reference corpus which can indicate how often any given word could be expected to occur in the text. It is important to choose the reference corpus for comparing representative characters (Baker 2006). In this case, we will select men's corpus as a target corpus and women's corpus as a reference corpus, and then select women's corpus as a target corpus and men's corpus as a reference corpus. Thus, we will research for comparing men with women's text each other.

The method of making word lists and key word lists will use *antconc* 3.1.

## 3 Using corpus

We retrieved 300 ~ 2000 words from each of the 65 speakers out of Sejong Spoken Corpus, the total of which is around 180,000 words in total. The current study is based on Sejong Spoken Corpus, which contains approximately one million words. We have tried to sample it with around 185,000 words so that we can get a more balanced corpus with respect to registers (private speech) and gender (Han and Kang 2009). The reason we retrieved the sample corpus was to eliminate possibilities of the data being biased to specific speakers, situations or subjects. We have first divided the corpus with register, private speech, and then subdivided it with gender (men vs. female), so that four sets has been made in total. The number of participants is set to be 65 for each set, and 130 speakers have been participated in the whole subcorpus. The details of the balanced subcorpus are shown in table 1.

Table 1: Corpus size

		Words	Morphemes
Private speech	Women	91,677	185,157
	Men	91,559	183,827
Total		183,236	368,984

#### 4 Results

The result is shown in table 2 and 3. It is possible to confirm the differences between men and women’s private speech through the top of wordlist and keyword list.

Table 2: Frequency and keyness in men’s private spoken language

	Word list		Keyword list	
	Word	Frequency	Word	Keyness
1	kwu	1238	hyung	128.79
2	mweh	1110	mwueh	59.03
3	uh	1071	ince	30.18
4	keo	990	anya	29.76
5	ah	780	ceke	27.98
6	kwuntey	672	ccipal	27.76
7	com	540	conna	26.34
8	mak	481	imma	19.78
9	eum	484	com	17.84
10	eung	447	ne	16.12

Table 3: Frequency and keyness in women’s private spoken language

	Word list		Keyword list	
	Word	Frequency	Word	Keyness
1	uh	1930	uwm	396.86
2	hmm	1296	uh	249.41
3	ah	1190	oppa	154.06
4	keo	1007	toykey	132.60
5	mweh	777	ah	85.98
6	kwuntay	773	ney	65.79
7	mak	694	emeo	61.69
8	eung	591	maca	55.60
9	toykey	468	cincca	46.33
10	cincca	448	ca	43.54

The result shows that women use ‘uh’, and men use ‘ku’ most frequently as shown in table 2 and 3. There are 18,657 types in women’s text, and 19,847 types in men’s text. The type/token ratio (men : women) is 0.203508 : 0.216673. So men usually talk to express diversly than women in private speech.

In terms of keyness, discourse markers, such as “hmm” or “uh”, are salient to women’s speech, while exclamatives, such as, “brother!”, “what!”, or “so” are salient to men’s speech. This result shows different results from non-corpus-based studies. Also, it shows women usually use to gain time or do hesitation distinctively as the use of “hmm” or “uh”. Contrary to Im’s research (2004) that hedges

like “mwue” were argued to be keywords in women’s speech, this corpus based-study shows they are characteristics for men’s speech.

Men strongly prefer to use “ince”, while expressions showing women’s eagerness to participate in communications, such as “ney, eme, macca” is also salient. So, it confirms that men prefer to show enthusiastic expressions in an event time. For adverbial discourse markers, women preferred “toykey”, or “cincca”, men preferred “conna”. However, the dispersion value of “conna” was relatively lower than other words(the frequency : 19, the value of dispersion : 0.02). The result shows that women respond or express the status quo more emotionally than men (Scott & Tribble 2006).

#### 5 Conclusion

This research shows that there are difference between men and women’s private speech, especially discourse markers. This result shows different results from non-corpus-based studies before. We used a balanced corpus for researching men and women’s private speech with around 185,000 words. Through making word lists and key word lists, it is possible to extract conspicuous characteristics by gender speech. For example, women use to gain time or do hesitation distinctively as the use of “hmm” or “uh”, while men prefer to use “ince” to participate in communications actively.

#### References

- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Coates, J. (1998). *Gossip Revised: Language in All – Female Groups, Language and Gender*, edited by Coates, J. Blackwell.
- Graddy, D. B.(2006) Gender Saliency and the Use of Linguistic Qualifiers and Intensifiers in Online Course Discussions. *The American Journal of Distance Education* 20/4, 211-229.
- Han, N., and B. Kang. (2009). Sejeong kwue korpuslul yiyonghan sengbyel, sayongyekpyel chai yenkwu. Ms. *Institute of Korean Culture*, Korea University.
- Im, K. (2004). The role of gender in spoken Korean. *Linguistics and language* 83(3), 93-113.
- Lakoff, R. (1975). *Language and women's place*, New York: Harper and Row.
- Scott, M. & C. Tribble. (2006). *Textual Patterns: Key words and corpus analysis in language*. JohnBenjamins Publications