

Shifting Semantics of Grandmother in Digital Japanese-Korean Comfort Women Discourses*

KERRY SLUCHINSKI

University of Alberta, East Asian Studies

1 Introduction

When appearing together in discourse, the terms ‘Japan’ and ‘Korea’ often invoke complex and intricate relations frequently approached from the perspective of sentiment analysis. Sentiment analysis is the use of computer science technologies “to systematically detect, extract, and classify the subjective information and affective states expressed in a text, such as opinions, attitudes, and emotions regarding a service, product, person, or topic” (Lei and Liu 2021:1). Public sentiment analysis focusing on Japanese-Korean relations using corpus linguistics (e.g., Doi et al. 2021) and quantitative survey methods (e.g., Cho et al. 2021) have been on the rise. Furthermore, the co-

* This work, *Shifting Semantics of Grandmother in Digital Japanese-Korean Comfort Women Discourses*, is supported in part by funding from the Social Sciences and Humanities Research Council of Canada (752-2021-2541).

Japanese/Korean Linguistics 30

Edited by Sara Williamson, Adeola Aminat Babayode-Lawal, Laurens Bosman, Nicole Chan, Sylvia Cho, Ivan Fong, and Kaye Holubowsky.

Copyright © 2023, CSLI Publications.

occurrence of the two terms 'Japan' and 'Korea' is often tied to historical evocations of 'Comfort Women'.

As outlined by Kim et al. (2019:58), 'Comfort Women' are those who fell victim to "the system of sexual slavery created and controlled by the Imperial Japanese government between 1932 and 1945," the period during WWII. Despite the fact that this system is said to be "the largest case of government-sponsored human trafficking and sexual slavery in modern history," "comfort women discourse is growing increasingly polarized as human rights claims are juxtaposed against populist claims according to a competitive logic" (Eriksson 2023:90). That is, traditionally, much "Comfort Women" research has focused on the historical debate regarding whether the claimed events happened or not, placing the victims and their personal stories under lenses of scrutiny and denying them of their self-proclaimed identities and collective memories (e.g., Ramseyer 2021).

As a result, there is often great tension between Japan and Korea, with continuous shifts in public sentiment among the nationals of both countries towards the nationals of the respective opposing country. A great deal of public sentiment studies mention Comfort Women in some capacity (e.g., Oh and Wha Han 2022; Ahn and Yoon 2020). In addition, Comfort Women Discourses have been receiving more attention from scholars in the fields of international relations and sentiment (e.g., Asaba 2021), as well as collective memory and activism (e.g., Yoon and Chang 2021) and media (e.g., Ogawa and Kobayashi 2021). However, such studies, especially those in the digital space, do not examine the language used by people discussing the topic of Comfort Women as part of their everyday digital engagement. Instead, these studies take narratives of experience as objects of study via ethnography and the products of organized media, such as news articles, as objects of study via large scale corpus linguistics. While these studies are certainly valuable, they do not shed light on the linguistic aspects of Comfort Women Discourses and the language practices of those who make these discourses possible.

Furthermore, it could be argued that the term "Comfort Women" itself devalues the identities of victims. The term is said to be a euphemism coined by the perpetrators (i.e., the Japanese military), with scholars having argued that the term "Comfort Women" downplays the severity of the crime (Kim et al. 2019:58; Ward and Lay 2018:1). However, scholars such as Soh (2008:232) note that the term 'Comfort Woman', in native Korean 위안부 *wianbu*, can be traced back to as early as the King Sejong dynasty in the 15th century. This then questions the legitimacy of claims that 慰安婦 *ianfu* is indeed a coined euphemism by the Japanese military. While the etymology of the term "Comfort Women" itself is worthy of further pursuit, it is beyond the scope of this paper.

Nonetheless, in the Computer Mediated Communication landscape of the 21st Century, 'Comfort Women' are often referred to in the digital discourses

of netizens as ‘Comfort Women Grandmother(s)’ or just ‘Grandmother(s)’. This study, as an extension of Sluchinski (2022), works to further address the research gap surrounding how the identities of Comfort Women are constructed by everyday Korean and Japanese language users through focusing on the discourse of netizens in both language contexts.

More specifically, the quantitative investigation into the language use of Korean and Japanese netizens in the context of digital Comfort Women Discourses by Sluchinski (2022) has shown that the term ‘grandmother’ appears in a plethora of orthographic and morphological variations. Sluchinski’s (2022) survey yielded 14 distinct ‘grandmother’ variants in Korean and 32 distinct variants in Japanese within The Comfort Women Discourses corpus. In addition to identifying the multiple ‘grandmother’ variants used by netizens, Sluchinski (2022) generated a politeness ranking of these variants based on morpho-syntactic criteria pertaining to each lexical item. Amongst the Korean variants, 할머니(들) *halmeoni(deul)* ‘grandmother(s)’ was deemed the most standard, i.e., neutral, while お婆さん(たち)/おばあさん(たち) *obaasan(tachi)* ‘grandmother(s)’ can be considered the most neutral Japanese variant due to lexicalization. Yet, from a pragmatic discourse perspective, the ranking design begs the following question: how neutral can these two grandmother variants be when taken in the specific contexts where these variants appear?

Thus, the analysis is concerned with addressing this question through multiple layers of semantics-focused Corpus Linguistics-based approaches such as (1) Collocates and (2) 5 NGRAMS, for the Korean term 할머니 and the Japanese Kanji term お婆さん and Japanese Hiragana term おばあさん. It must be noted that the results reported here can still be considered as ‘working results,’ the reasons for which are outlined in the Methodology.

2 Data Composition and Theoretical Background

This section explains the data origin, corpus composition, and theoretical background for the study. As the study uses a revised version of the *Digital Comfort Women Discourses Corpus* presented in Sluchinski (2022:245-248), only necessary information as well as information needed to understand the revisions will be presented here. The primary sourcing platform *Kaikaihanno* is briefly introduced in Section 2.1 along with the two primary Korean netizen community sources ILBE and Naver News, including the data allocation in each, as well as the Japanese netizen community and its data allocation; and Section 2.2 introduces the theoretical background and the research focus.

2.1 Corpus Construction and Data Origin

The data for the analyses in this study is drawn from *The Revised Comfort Women Discourses* corpus, constructed in January 2020 from *Kaikaihanno* (a translation blog dedicated to reacting to Korea), using the blog’s ‘Comfort Women’ tag, and consists of a collective of 19,553 comments in Korean, 39,283 comments in Japanese, and 2,729 Korean comments translated into Japanese which comprises some of the content of the 77 Japanese blog entries. The comments are spread across 146 texts (77 Japanese; 69 Korean), and the Korean texts mainly come from ILBE (n=32), a notorious alt-right platform, and Naver News (n=36), an online news portal. In terms of neutral grandmother variants, the data processed thus far consists of 294 おばあさん *obaasan* tokens of 1,149,061 corpus tokens, 154 お婆さん *obaasan* tokens of 1,149,061 corpus tokens, and 1,130 할머니 *halmeoni* tokens of 347,351 corpus tokens. An overview of the corpus is shown in Table 1 below which also reflects the Korean, Japanese, and Translation subcorpora.

Revised Comfort Women Discourses Corpus							
Korean Data						Japanese Data	
Ilbe		Naver		Other		Kaikaihanno	
Articles	Comments	Articles	Comments	Articles	Comments	Entries	Comments
32	6,597	36	12,914	1	42	77	39,283
Total Articles							
						69	
Total Comments							
						19,553	
Translated Comment Data							
1,461		1,236		32		Total Translated Comments	
						2, 729	

Table 1. Revised Digital Comfort Women Discourses Corpus Composition

The translation aspect of this project has been previously documented and explained in Sluchinski (2022) and is still a work in progress. Therefore, it will not be extrapolated further in the current paper.

2.2 Theoretical Background and Research Focus

Adopting a semantics based quantitative Corpus Linguistics approach, this study focusses exclusively on the neutral ‘Grandmother’ variants in context. The study addresses the basic question of whether these seemingly neutral variants may have developed positive or negative associations through particular frequent collocations, i.e., whether they have experienced a semantic shift, thus causing them to be no longer truly neutral as a pragmatic unit. Such

a phenomenon, also known as semantic prosody, has been investigated by scholars such as Sinclair (1991).

3 Software and Methodology

This section outlines the software used and steps taken to process the data through various stages. Specifically, the data was processed into raw txt files (Section 3.1), segmented with part of speech taggers and simple spacing (Section 3.2), was adjusted and fine tuned in bulk (Section 3.3), was subject to a series of Corpus measures under various conditions (Section 3.4), and were cleaned for presentation (Section 3.5). However, as will be discussed in Section 3.6 this multi-step process was cumbersome in several ways and has room for improvement given the appropriate access to better technology.

3.1 Text Files

The data for the study was originally collected by saving the postings, news articles, and their comments in PDF format and storing them in aggregate. These files, for both Japanese data and Korean data, were then converted to word documents to work with the text more easily. In order to be used with the other programs and software universally, these word documents were then converted to .txt files.

After converting to .txt files several modifications were required to ensure minimal interference with collocation results and other frequency based statistical analyses. These modifications included removing hyperlinks and URL tags, non UTF-8 emojis, and instances of phrases such as “this comment has been deleted (e.g., 삭제된 댓글입니다)” from both Japanese and Korean data; removing usernames from the Korean data¹ (not applicable for Japanese data as > 99% of usernames were “anonymous”); and removing translated content/comments from the Japanese data² (not applicable for the Korean data as all the content is originally written in Korean).

¹ ILBE community users tend to use trolling usernames as screen IDs such as “박보검팬티” (Park Bogum’s Panties, where Park Bogum is a famous actor). Programs are unable to distinguish between text in a comment and text in a username, and as thus will count this portion of the username as one of the key words in proximity of other words, creating inaccurate results.

² One key component of the blog posts on *Kaikaihanno* is the translation of native Korean comments into Japanese; many Japanese commentators will quote the translation in their “reply” to the blog post. The continuous quotation of the translated material skews the frequency analyses as the programs do not detect or differentiate repeated text sequences.

3.2 POS Segmenting

After obtaining the .txt files both the Korean and Japanese data were processed with POS tagging software. It is well known that Japanese, like Chinese, does not have spaces and uses a character scripted. In order to be processed by AI and NLP languages like Japanese and Chinese need to have spaces inserted at word boundaries.

For the Japanese data, TagAnt v2.0.4 was used for simple word boundary insertion, taking the text on the left in Figure 1 below and producing the segmented text on the right. However, for the purpose of the current study TagAnt produced what is known as over-segmentation, inserting word boundaries at incorrect and/or excessive junctures as marked by the red arrows. These over-segmented portions required correction which will be addressed in Section 3.3.

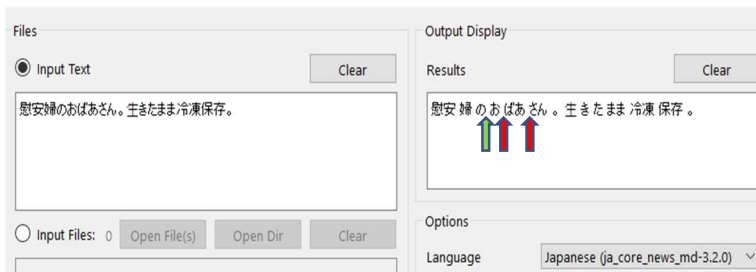


Figure 1. Segmented Text in TagAnt

For the Korean data, POSTAG, released in 2002 by the POSTECH NLP Lab, was used. POSTAG is a python-based text POS tagger that operates based on the Sejong corpus. Unfortunately, POSTAG was the only program which I was able to execute on my data that provided a continuous line output (see Figure 2 below).

Pre-POSTAG	문재인 위안부다 북한에 신경만 쓴다해
Post-POSTAG	문재/NNG 이/VCP L/ETM 위안부/NNG 이/VCP 다/EF 북한/NNP 예/JKB 신경/NNG 만/JX 쓰/VV L 다/EC 하/VX 아/EC

Figure 2. POSTAG Output Sample

That is, while there are more advanced, UI friendly Korean POS tagging programs such as Java-based RHINO 3.8.0, these programs produce isolated, tokenized lists as output which are not compatible with corpus programs such as AntConc that require continuous line output.

POSTAG is not without problems and there is no support offered for users as the creators are unresponsive to inquiries. Some of the cons and bugs discovered when executing POSTAG were: (1) it only takes .txt files; (2) will not execute and generates errors when files contain the following: blank lines, double consonants and single vowels (e.g., ㅈ, ㅊ, ㅌ, ㅍ, ㅍ, ㅍ, etc.), emojis (UTF-8 and non-UTF-8), consecutive clusters (e.g., ㅎㅎ, ㅋㅋㅋ, etc.), consecutive dashes/lines (e.g., -----); (3) the word boundaries are not accurate; and (4) the tags are not correct. As with the Japanese data, the Korean data also required correction.

3.3 Bulk Text Editing

In order to correct both the Japanese and Korean data Notepad ++ v8.4.8. was used. Specifically, it is a bulk text editor which was used to manually manipulate/clean generated word boundaries, tags, and illegible UTF-8 encodings (emojis, etc.). In the Japanese data, this included words immediately relevant to the neutral “grandmother” variant: 慰安婦 (Comfort Woman), お婆さん (Grandmother), おばあさん (Grandmother), 日本人 (Japanese person), 朝鮮人 (Joseon Person), 韓国人 (Korean Person), 加害者 (victim). In the Korean data, there are too many cases to list individually however in general corrections were made for frequently occurring proper names, nouns, and Comfort Women terminology like in the Japanese data (see Table 2 below).








English	Pre-Correction		Post-Correction
Grandmother	할머니/NNG 이/VCP ㅁ/ETN		할머니/NNG
Lee Yongsoo	이용_NNG 수_NNB		이용수_NNG
Lee Yongsoo	이용_NNG 수씨_NNG		이용수씨_NNP
Moon Jaemin	문재_NNG 이_VCP ㄴ_ETM		문재인_NNP
Uniqlo	유니크_NNG ㄹ로_JKB		유니클로_NNP
Sewol Ferry	세월_NNG 호_NNB		세월호_NNG
Comfort Woman	위_NNG 안부_NNG		위안부_NNG

Table 2. Sample Corrections to Korean Data

3.4 Corpus Processing Software

For both the Japanese and Korean data, AntConc v4.1.4., a corpus analysis toolkit for researchers, teachers, and learners, was used. As there is already existing literature with regards to the builds and programming of AntConc,

such matters will not be discussed here. AntConc was used to conduct analyses from two angles: (1) collocates, and (2) NGRAMs. When using AntConc, the POS tags need to be converted from a / format to an _ format (e.g., 할머니/NNG to 할머니_NNG) to be imported and read.

3.4.1 Collocate Parameters

The collocate analysis was conducted at two thresholds, $p < 0.05$ (3.84) and $p < 0.0001$ (15.13), to give a comparative lens. Both thresholds had parameters set with Log-Likelihood as the Likelihood Measure, TSCORE as the Effect Size Measure, with the Threshold at All Values, and a Window Span of 8L + 5R on the Korean data 할머니, the Japanese Kanji お婆さん, and the Japanese Hiraganaおばあさん.

3.4.2 NGRAM Parameters

The NGRAM analysis was conducted at both RAW and Normalized values, with a minimum frequency of 2 for both the Korean and Japanese data. However, as Japanese is a rigidly left-branching language and NGRAM is based on right branching languages it was necessary to use Wildcards (*) to force the detection of characters to the left of the search query lemma (i.e., おばあさん). Thus, when performing NGRAM in Kanji * * * お婆さん was searched and when performing NGRAM in Hiragana * * * おばあさん was searched. The three wildcards in a row tells the engine to search 5NGRAM that includes 3 unknown characters to the left of the search query lemma, include the lemma, and then search 1 character to the right.

3.5 Cleaning and Processing Results

To clean and process the results for presentation, the query results generated by AntConc were exported as .txt files and then imported into excel. Results included a total of two collocate searches for each grandmother variant and one 5NGRAM result for each variant. Cleaning the results consisted of omitting some ranked entries in the collocates which were meaningless and superfluous for examining semantic prosody and statistic results (e.g., 의 (JKG= postpositions (case markers) as a collocate for 할머니).

3.6 Analytical Results Preface

As mentioned previously, the results generated for the Korean data are working results due to limited access to appropriate POS tagging tools. There are too many cons to working with POSTAG, and the workarounds are time consuming and cumbersome. Unfortunately, other Python and/or Java based programs were (for my case) not very accessible/user friendly and did not generate linear output for corpus linguistics (e.g., RHINO, Kakao/Kahiii,

wordseg_kaist_ud, MeCab_kor, hangul-Korean pypi, KoNLPy). I hope that the findings in this paper can generate some sense of the semantic shift in the terms for Grandmother used in Korean and Japanese Comfort Women Discourses, and generate discussion about NLP for better Korean data analysis.

4 Collocates Results and Discussion

This section discusses the collocate results for the Japanese and Korean data, starting with comparing the $p < 0.0001$ Kanji and Hiragana results (Section 4.1.1.), the comparing the $p < 0.05$ Kanji and Hiragana results with highlights notable similarities (Section 4.1.2), and then presenting a comparison of the $p < 0.0001$ vs $p < 0.05$ threshold results in Korean.

4.1 Japanese Results

The $p < 0.0001$ results generated 14 collocates for Hiragana 'grandmother' and 31 collocates for Kanji 'grandmother', while the $p < 0.05$ results generated much more. Consequently, only the top 14 results for $p < 0.0001$ are shown and discussed while only notable unique and shared collocates between Hiragana and Kanji within the top 50 TSCORE effect results are shown and discussed for $p < 0.05$ due to spatial limitations.

4.1.1 Results for $p < 0.0001$

To examine how different ranking parameters may impact the portrayal of semantic shift, I present the results ranked in four different ways: 1) High to Low (H-L) Effect ranking, 2) Left + Right (LR) combined frequency ranking, 3) Frequency Scaled ranking, and 4) TSCORE Likelihood ranking. Due to spatial limitation, Figures are provided periodically for visualization.

Focus is placed on the following collocates and how their rankings shift depending on the ranking parameter: 慰安婦 (Comfort Woman), ハルモニ (The Korean pronunciation of 'Grandmother' *Halmeoni* stylized in katakana), 奴隸 (Slave), バッカス (Bacchus), and かわいそう (Pitiful). 'Bacchus' is an important, culturally charged collocate. Bacchus is well known as the Roman God of fertility, wine, and pleasure. However, in the Korean cultural context 박카스 (Bacchus) is an energy drink launched in 1963. What is more, "Bacchus" is a code for prostitution amongst seniors in Korea with the women offering services known as "Bacchus Ladies" (박카스 할머니) (Yi & Yun, 2022). The film "The Bacchus Lady (죽여주는 여자)" by E J-yong released in 2016 portrays aspects of the Bacchus culture through the story of elderly sex worker So-Young. The collocation of 'Bacchus' with 'Grandmother' in both the Japanese and Korean data indicates an association of 'Grandmother', and by extension 'Comfort Women', with a chosen life of prostitution.

Collocate for おばあさん		Effect		Collocate for お婆さん		Effect
慰安婦 Comfort Woman		8.396	1	慰安婦 Comfort Woman		5.311
たち		6.858	2	の		4.661
の		6.111	3	達		3.682
証言		5.08	4	元		3.342
達		4.417	5	この		3.306
おばあさん Grandmother		3.281	6	バッカス Bacchus		2.977
ハルモニ *Grandmother		3.279	7	かわいそう pitiful		2.954
あの		2.99	8	お婆さん		2.58
歳 age		2.701	9	代		2.45
受け		2.613	10	意		2.046
奴隷 slave		2.613	11	母さん		1.941
亡くなっ		2.558	12	紙幣		1.723
なぜ		2.467	13	ハルモニ		1.722
生き		2.461	14			1.695

Figure 3. Japanese Data Effect Ranking

When sorted by Effect ranking (Figure 3), Comfort Woman is the first collocate for both variants and 'Halmeoni' is ranked seventh for Hiragana and 14th for Kanji. Slave appears as a unique collocate at the 11th rank for Hiragana while Bacchus and Pitiful appear at the sixth and seventh rank respectively for Kanji. Contrarily, when sorted by LR Frequency (Figure 4), Comfort Woman drops to second for both variants and 'Halmeoni' drops to ninth for Hiragana and stays 14th for Kanji. Slave still appears as a unique collocate but at the eighth rank (three increase) for Hiragana while Bacchus and Pitiful remain at the sixth and seventh rank respectively for Kanji.

Collocate for おばあさん		FreqLR		Collocate for お婆さん		FreqLR
の		272	1	の		146
慰安婦 Comfort Woman	↓	113	2	慰安婦 Comfort Woman	↓	50
たち		54	3	この		23
証言		30	4	達		17
達		26	5	元		13
おばあさん		12	6	バッカス Bacchus		9
あの		12	7	かわいそう pitiful		9
奴隷 slave	↑	12	8	歳		8
ハルモニ Halmeoni	↓	11	9	お婆さん		7
歳		11	10	代		5
受け		11	11	意		4
日本		9	12	母さん		3
なぜ		8	13	紙幣		3
生き		8	14	ハルモニ Halmeoni		3

Figure 4. Japanese Data FreqLR Ranking

When sorted by Frequency Scaled, Comfort Woman drops to third for Hiragana and 'Halmeoni' drops to 21st, while for Kanji Comfort Woman remains at second and 'Halmeoni' rises to 12th. Slave still appears as a unique collocate at the sixth rank (two increase) for Hiragana while Bacchus and Pitiful see decreases for Kanji. On the other hand, when sorted by Likelihood (Figure 5), Comfort Woman increases to second for Hiragana and 'Halmeoni' rockets

to fourth while Slave drops to 28th. For Kanji, Comfort Woman remains at second and 'Halmeoni' drops to 13th. Yet both Bacchus and Pitiful see sharp rises for Kanji, coming at first and third respectively.

Collocate		Likelihood		Collocate for お婆さん		Likelihood
たち		192.316	1	バッカス Bacchus		69.929
慰安婦 Comfort Woman	↑	176.16	2	慰安婦 Comfort Woman		64.7
証言		102.104	3	かわいそう pitiful	↑	57.539
ハルモニ	↑	77.247	4	達		45.729
達		59.755	5	元		43.998
の		53.053	6	お婆さん		38.084
おばあさん		47.846	7	の		31.34
亡くなっ		34.241	8	母さん		25.541
可愛		29.209	9	紙幣		24.738
あの		27.067	10	この		22.2
コロコロ		24.15	11	意		20.454
メンバー		23.707	12	歳		18.365
ええ		22.56	13	ハルモニ	↓	17.177
大切		22.31	14	代		15.496

Figure 5. Japanese Data Likelihood Ranking

4.1.2 Results for p< 0.05

Figure 6 below shows a comparison of NOTABLE shared tokens among the top 50 TSCORE Effect for Hiragana and Kanji collocates. The figure reveals that 慰安婦 is most common, direct association between Grandmother and Comfort Women in both contexts. We also see that shared derogatory/negative words rank more frequent in Hiragana while sympathetic words rank more frequent in Kanji.

Token	Semantic Attribute (+ OR – OR =)	Hiragana Stats (Rank/Effect)	Kanji Stats (Rank/Effect)
慰安婦 comfort woman	=	1 8.396	1 5.311
ハルモニ phonetic pronunciation of korean word for grandmother	–	7 3.279	19 1.695
バッカス bacchus, roman god of fertility, wine, pleasure.	–	43 1.656	6 2.977
かわいそう poor; pitiable; pathetic; pitiful	+	33 1.868	7 2.954
嘘つき liar	–	20 2.12	25 1.636
受け to receive; to get; to accept	–	10 2.613	20 1.689
金 money	–	27 1.971	13 2.305

Figure 6. Notable Shared Collocate Tokens

Notable distinct tokens for Hiragana showed that Hiragana has a more negative semantic prosody than Kanji with a liar and derogative identity. The Comfort Women are demanded to show proof of what happened and accused of always changing their words. In addition, the sentiment that they were not victims but sex servant emerges considering “victim” is used with strong sarcasm and occurs in the context of “victim business.” The notable tokens are, with Rank/Effect noted, as follows: 証言 ‘testimony; verbal evidence’ (4/5.08); おばあさん ‘grandmother’ (6/3.281); 奴隷 ‘slave; servant’ (11/2.613); 亡くなっ ‘pass away’(12/2.558); コロコロ ‘changing ones words frequently’ (15/2.322); 性 ‘sex’ (17/2.218); 被害 ‘(suffering) damage; injury; harm; victim’ (19/2.123); 可哀想 ‘poor; pitiable; pathetic; pitiful’ (21/2.078); 可愛 ‘cute’ (25/1.981); 自称 ‘self-proclaimed’ (34/1.831); 証拠 ‘evidence; proof’ (37/1.708); 死ん ‘die’ (41/1.673).

Notable distinct tokens for Kanji showed that Kanji seems to have more positive semantic prosody than Hiragana with a victim identity. The notable tokens are, with Rank/Effect noted, as follows: 元 ‘former’ (4/3.342); お婆さん ‘grandmother’ (8/2.58); 可哀そう ‘poor; pitiable; pathetic; pitiful’ (33/1.394); 貧し ‘poor; needy; lacking’ (35/1.391); 妊娠 ‘pregnancy’ (36/1.389); 偽証 ‘false evidence’ (37/1.381); 苦痛 ‘pain; agony; bitterness’ (39/1.376); 同情 ‘sympathy; compassion; feeling pity for’ (46/1.312).

4.2 Korean Results

The results discussed for the Korean data are based on the top 50 TSCORE Effect ranking, only taking into consideration of the collocates that were tagged as NNP, NNG, NNB, SN, VV, VA, VVS, XSV (purple exclusive to $p < 0.05$), with a focus on the Top 14. Like the Japanese data, one goal of comparing $p < 0.05$ and $p < 0.0001$ under different ranking conditions is to show how different ranking mechanisms produce different results, begging the question as to which mechanism is most reliable. The analysis was conducted from two ranking angles: LR frequency H-L (Figure 7), H-L Effect (Figure 8).

Viewing the rankings in order we can see that the LR Frequency H-L shares a number of collocates between the $p < 0.05$ and $p < 0.0001$ thresholds, notably ‘Comfort Woman,’ ‘receive,’ ‘use,’ ‘politics,’ ‘victim,’ and ‘testimony.’ This ranking mechanism starkly contrast with the H-L Effect ranking, which exhibits identical ranking and collocates between the $p < 0.05$ and $p < 0.0001$ thresholds. Within this parameter, we see that ‘Comfort Woman,’ ‘use,’ ‘testimony,’ and ‘victim’ are prominent collocates in the upper half.

0.0001					Frequency LR H-L					0.05						
Collocate	POS	Freq(Scaled)	FreqLR	Effect		Collocate	POS	Freq(Scaled)	FreqLR	Effect		Collocate	POS	Freq(Scaled)	FreqLR	Effect
위안부 Comfort Woman	NNG	41860	430	17.156	1	위안부 Comfort Woman	NNG	41860	430	17.156		위안부 Comfort Woman	NNG	41860	430	17.156
받 receive	VV	17186	61	3.907	2	받 receive	VV	17186	61	3.907		받 receive	VV	17186	61	3.907
이용 use	NNG	6019	61	6.443	3	이용 use	NNG	6019	61	6.443		이용 use	NNG	6019	61	6.443
할머니 Grandmother	NNG	14248	60	4.483	4	할머니 Grandmother	NNG	14248	60	4.483		할머니 Grandmother	NNG	14248	60	4.483
일본 Japan	NNP	60983	49	-8.452	5	일본 Japan	NNP	60983	49	-8.452		일본 Japan	NNP	60983	49	-8.452
정치 politics	NNG	7241	37	3.971	6	정치 politics	NNG	7241	37	3.971		정치 politics	NNG	7241	37	3.971
피해자 victim	NNG	6487	37	4.191	7	피해자 victim	NNG	6487	37	4.191		피해자 victim	NNG	6487	37	4.191
증언 testimony	NNG	2639	34	5.028	8	증언 testimony	NNG	2639	34	5.028		증언 testimony	NNG	2639	34	5.028
몰아가시 passed away	VV	988	32	5.347	9	몰아가시 passed away	VV	988	32	5.347		몰아가시 passed away	VV	988	32	5.347
위하 for (the purpose of)ums sake)	VV	6110	28	3.243	10	위하 for (the purpose of)ums sake)	VV	6110	28	3.243		위하 for (the purpose of)ums sake)	VV	6110	28	3.243
이용수 Lee Yongsoo	NNP	832	28	-5.013	11	이용수 Lee Yongsoo	NNP	832	28	-5.013		이용수 Lee Yongsoo	NNP	832	28	-5.013
좌파 Left Wing	NNG	4290	24	3.346	12	좌파 Left Wing	NNG	4290	24	3.346		좌파 Left Wing	NNG	4290	24	3.346
김복동 Kim Bokdong	NNP	520	24	4.711	13	김복동 Kim Bokdong	NNP	520	24	4.711		김복동 Kim Bokdong	NNP	520	24	4.711
용서 forgiveness	NNG	2925	23	3.714	14	용서 forgiveness	NNG	2925	23	3.714		용서 forgiveness	NNG	2925	23	3.714

Figure 7. Frequency LR HL Rank

This phenomenon in ranking shift may indicate that among all the ranking mechanisms, Effect by TSCORE is the most accurate/relevant. In addition, we see that the Korean data collocates Follows similar themes as in the Japanese data, but with a more pronounced victim identity of pity, mirroring the Kanji data.

0.0001					TSCORE Effect H-L					0.05						
Collocate	POS	Freq(Scaled)	FreqLR	Effect		Collocate	POS	Freq(Scaled)	FreqLR	Effect		Collocate	POS	Freq(Scaled)	FreqLR	Effect
위안부 Comfort Woman	NNG	41860	430	17.156	1	위안부 Comfort Woman	NNG	41860	430	17.156		위안부 Comfort Woman	NNG	41860	430	17.156
이용 use	NNG	6019	61	6.443	2	이용 use	NNG	6019	61	6.443		이용 use	NNG	6019	61	6.443
몰아가시 passed away	VV	988	32	5.347	3	몰아가시 passed away	VV	988	32	5.347		몰아가시 passed away	VV	988	32	5.347
증언 testimony	NNG	2639	34	5.028	4	증언 testimony	NNG	2639	34	5.028		증언 testimony	NNG	2639	34	5.028
이용수 Lee Yongsoo	NNP	832	28	-5.013	5	이용수 Lee Yongsoo	NNP	832	28	-5.013		이용수 Lee Yongsoo	NNP	832	28	-5.013
김복동 Kim Bokdong	NNP	520	24	4.711	6	김복동 Kim Bokdong	NNP	520	24	4.711		김복동 Kim Bokdong	NNP	520	24	4.711
할머니 Grandmother	NNG	14248	60	4.483	7	할머니 Grandmother	NNG	14248	60	4.483		할머니 Grandmother	NNG	14248	60	4.483
피해자 victim	NNG	6487	37	4.191	8	피해자 victim	NNG	6487	37	4.191		피해자 victim	NNG	6487	37	4.191
정치 politics	NNG	7241	37	3.971	9	정치 politics	NNG	7241	37	3.971		정치 politics	NNG	7241	37	3.971
받 receive	VV	17186	61	3.907	10	받 receive	VV	17186	61	3.907		받 receive	VV	17186	61	3.907
피해 damage	NNG	2873	23	3.733	11	피해 damage	NNG	2873	23	3.733		피해 damage	NNG	2873	23	3.733
용서 forgiveness	NNG	2925	23	3.714	12	용서 forgiveness	NNG	2925	23	3.714		용서 forgiveness	NNG	2925	23	3.714
좌파 Left Wing	NNG	4290	24	3.346	13	좌파 Left Wing	NNG	4290	24	3.346		좌파 Left Wing	NNG	4290	24	3.346
위하 for (the purpose of)ums sake)	VV	6110	28	3.243	14	위하 for (the purpose of)ums sake)	VV	6110	28	3.243		위하 for (the purpose of)ums sake)	VV	6110	28	3.243

Figure 8. TSCORE Effect H-L Rank

5 NGRAM Results and Discussion

The 5NGRAM analysis produced 50 results in the Korean data, 60 in for the Hiragana variant in Japanese, and 10 for the Kanji variant in Japanese. For each variant, I surveyed the range of NormFrequency for cases of each value rounded to the nearest whole number, plotting them in pie charts.

5.1 Korean Results

In the Korean results (Figure 9), we see that more than $\frac{3}{4}$ of the 50 NGRAMS have a Normalized Frequency of 16807, the lowest, while only 2% fall at the highest of 50420. This may indicate that the 5NGRAM results in the Korean data are not significant, potentially because of inadequate POS tagging and segmenting. However, what may be interesting to note qualitatively is the Top 10 NGRAM results in the Korean data (Figure 10) which thematically mirror the collocate results.

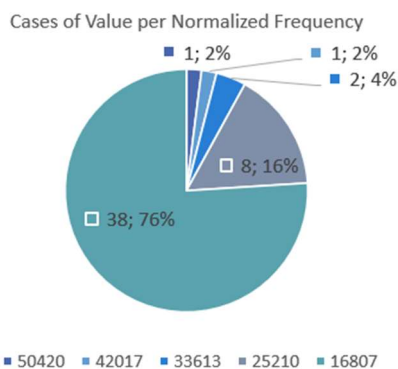


Figure 9. Korean Cases of Value per NormFreq 5NGRAM

Type	Freq
할머니들을 정치 적	6
할머니들의 명예 를	5
할머니들에 대하 ㄴ	4
할머니들이 용서 하	4
할머니들 다 돌아가시 면	3
할머니들이 돈 을	3
할머니들을 위하 아	3
할머니들을 이용 하	3
할머니까지 정치 적 으로	3
할머니들 보상도 필요없다고 하시	3

Figure 10. Top 10 Korean 5NGRAM

5.2 Japanese Results

In the Japanese results we can see a comparison between the Kanji (Figure 11) and the Hiragana (Figure 12). For Kanji, which only produced 10 NGRAMS, we see that the lowest NormFreq is about 90909; This is higher than the majority of NGRAM results for Hiragana which are 15151 (55 cases). The NGRAM results seem to indicate too much diversity.

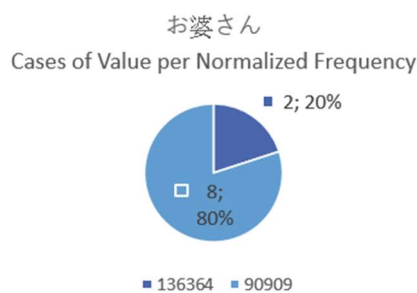


Figure 11. Kanji Cases of Value per NormFreq 5NGRAM

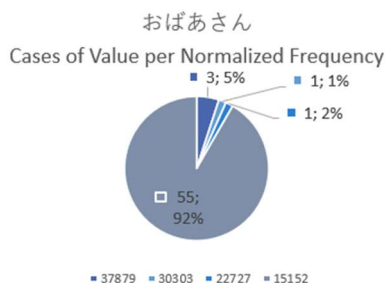


Figure 12. Hiragana Cases of Value per NormFreq 5NGRAM

6 Conclusion

This paper shared working results via: (1) collocates from two threshold levels, and (2) 5NGRAM. Japanese Hiragana Collocates show that both $p < 0.0001$ and $p < 0.05$ construct negativity/hostility, and that Hiragana “Grandmother” has seemed to develop a negative semantic prosody. Japanese Kanji Collocates show that both $p < 0.0001$ and $p < 0.05$ seem to have more positive semantic prosody than Hiragana with a victim identity, and that Kanji “Grandmother” has seemed to develop positive semantic prosody. In Korean Collocates, we saw (1) a similar pattern of language use themes; (2) that semantic prosody seems to remain unchanged but that an association with Comfort Woman is very strong; (3) that “Grandmother” may be considered as a synonym to “Comfort Woman” rather than having experienced semantic prosody; and (4) ‘Grandmother’ is synonymous with a ‘Victim Comfort Woman who was used’. The working results of this study can be greatly improved with more reliable Korean POS tagging software.

Acknowledgments

Warmest thanks to my colleague Dr. Yoshi Ono (U of A) and undergraduate RA Giseo (Peter) Lee for their comments and advice on earlier versions, as well as Dr. Yongtaek Kim (Georgia Tech) for sharing resources.

References

- Asaba, Y. 2021. Bringing People’s Voices Back in Japan–South Korea Relations. In *Japanese Public Sentiment on South Korea*, 1–17. Routledge.
- Ahn, J.H. and E. K. Yoon. 2020. Between Love and Hate: The New Korean Wave, Japanese Female Fans, and Anti-Korean Sentiment in Japan. *Journal of Contemporary Eastern Asia* 19(2): 179–196.
- Anthony, L. 2022. AntConc (Version 4.1.4) | TagAnt (Version 2.0.4) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software>
- Cho, Y.S., J.B. Chung and H.E. Shim. 2021. How do Korean Respond to Japanese Retailers? *Journal of Distribution Science* 19(10): 87–97.
- Doi, S., K. Inamasu, S. Kohama and A. Tago. 2021. South Korea–Japan FCR Crisis and Public Opinion. In *Japanese Public Sentiment on South Korea*, 18–31. Routledge.
- Eriksson, A.K. 2023. Representing the ‘Comfort Women’: Omissions and Denials in Wartime Historiographies in Japan. In *Representing Gender-Based Violence*, 89–109. Springer International Publishing.
- Kim, J., B. Bisland and S. Sun. 2019. Teaching about the Comfort Women during World War II and the Use of Personal Stories of the Victims. *Asian Literature in the Humanities and Social Sciences* 24(3): 58–63.

- Lei, L. and D. Liu. 2021. *Conducting sentiment analysis*. Cambridge University Press.
- Ogawa, Y. and T. Kobayashi. 2021. Semantic Structure of the Comfort Women Issue in Japanese and South Korean Newspapers. In *Japanese Public Sentiment on South Korea*, 87–104. Routledge.
- Oh, D. and M. Wha Han. 2022. Globalization from Above and Below: Rejecting Superficial Multiculturalism and Igniting Anti-Korean Sentiment in Japan. *International Journal of Cultural Studies* 25(1): 51–67.
- Ramseyer, J. M. 2021. ‘Contracting for Sex in the Pacific War’, *International Review of Law and Economics* 65: 1–8.
- Sinclair, J. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sluchinski, K. 2022. Im/Politeness Variations in Digital Japanese-Korean Comfort Women Discourses. *Japanese/Korean Linguistics* 29: 243–257.
- Soh, C. S. 2008. *The Comfort Women: Sexual Violence and Postcolonial Memory in Korea and Japan*. University of Chicago Press.
- Ward, T. and W. Lay. 2018. *The Origins and Implementation of The Comfort Women System*. E-International Relations.
- Yi, H.-Y. and S.-M. Yun. 2022. ‘Bacchus Ladies’ are Everywhere and Nowhere: Reconstruction of Network of Elderly Prostitution in Jongno 3-ga. *Journal of Korean Women’s Studies* 38(3): 179–214.
- Yoon, H. and J. Chang. 2021. The Story "Our Grandmothers" Could Not Tell: Representation of the Comfort Women and the Physical Manifestation of Memory. *Azalea: Journal of Korean Literature & Culture* 14: 311–334.