

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE IDIOMAS



Japanese women's language: a corpus-based study

Para obtener el grado de Maestra en Lenguas Modernas

Presenta

Marcela Salgado Mercado

Tijuana, Baja California, 29 de junio 2021

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE IDIOMAS



Para obtener el Grado de Maestro en Lenguas Modernas

Presenta

Marcela Salgado Mercado

Aprobado por

Rosío del Carmen Molina Landeros
Directora

Alfredo Escandón Jiménez
Coodirector

Carlos Alberto Noriega Guzmán
Lector

Tijuana, Baja California, 29 de junio de 2021

Table of Contents	3
List of tables and figures	6
Notes on romanization and abbreviations	7
Summary	8
Chapter I. Introduction	9
1.1 Research Background	9
1.2 Problem Statement.....	13
1.3 Research Objectives	15
1.3.1 Main objective	15
1.3.2 Secondary objectives	15
1.4. Research scope	15
Chapter II. Theoretical Framework	17
2.1 Gender and Cultural Ideology in Japan	17
2.1.1 Gender Expectations	17
2.2 Japanese Women’s Language	18
2.2.1 A brief history	18
2.2.2 Linguistic Characteristics	19
2.2.2.1 Sentence final-particles	20
2.2.2.2 First-person pronouns	21
2.2.2.3 Honorifics and verb endings	21
2.2.2.4 Politeness	22
2.2.2.5 Diaphasic variable	24
2.2.3 Language Change	24
2.2.4 Japanese Language Teaching	25
2.3 Queer Linguistics	26
2.3.1 Trans identity	26
2.3.2 Queer Theory	28
2.3.2.1 Gender	28
2.3.2.2 Performance	29
2.3.2.3 Identity	30

2.3.3 Japanese Queer Terminology	31
2.4 Internet Linguistics	33
2.4.1 General description	33
2.4.2 Twitter characteristics	37
2.4.3 Japanese Cyberculture	40
2.4.4 Orality Online	41
2.4.4.1 Intonation	43
Chapter III: Methodology	45
3.1 Documental research	45
3.2 Corpus Linguistics	47
3.2.1 Corpus Definition	47
3.2.1.1 Corpus-based approach to language	49
3.2.1.2 Software	50
3.2.1.3 Wordlists	51
3.2.1.4 Keywords	51
3.2.1.5 Reference Corpus	53
3.2.2 Corpus Compilation	54
3.2.2.1 Data collection.	55
3.2.2.2 Corpus specifications	59
3.2.2.3 Corpus size and representativeness	60
3.2.2.4 Corpus annotation	60
3.2.2.4.1 Character encoding	61
3.2.2.4.2 Segmentation.....	62
3.3 Variables, categories and indicator	63
3.3.1 Femininity variable	63
3.3.2 Masculinity variable	65
3.3.3 Politeness variable	67
3.3.4 Categories	67
3.3.4.1 Identity	68
3.4 Sample analysis procedure	68

3.5 Design Proposal Description	71
3.5.1 Proposal Criterion	71
Chapter IV.	73
4.1 Keyness analysis results	73
4.1.1 Corpus A initial analysis.	74
4.1.2 Corpus B initial analysis.	78
4.1.3 Crosstabulation among corpora	81
4.2 Frequency comparing among corpora	83
4.3 Concordances	86
4.3.1 Sentence-final particles	86
4.3.1.1 Corpus A results	87
4.3.1.2 Corpus B results	89
4.3.2 Pronouns	93
4.3.3. Verbs	94
4.3.4 Beautification prefixes	96
4.4 Limitations	97
Chapter V. Discussion.	98
References.	102
Appendix A: Japanese Genderlect Lesson Plan.	112
Appendix B: Corpus A top 100 keywords	116
Appendix C: Corpus B top 100 keywords	122

List of Tables and Figures

Tables

1. The first-person pronoun with its immediate collocations in both corpora.
2. Difference in gender presentation with final particles.
3. BIC values and their interpretation
4. Keywords from CA, using the frequency focus approach.
5. Keywords from CB, using the frequency focus approach.
6. Crosstabulation of each of the joseigo variables.
7. Normalized frequencies for CA and CB per 10k
8. Normalized frequencies gap among CA and CB

Figures

1. The first-person pronoun with its immediate collocations in both corpora.
2. Verb ending *-masu* and its clusters in both corpora.
3. Beautification prefix collocates in both corpora

Notes on romanization and abbreviations

Abbreviations for this research are used to provide grammatical information for the linguistic classes. These are as follows: A, adjective; Adv, adverb; AN, adjectival noun; AUX, auxiliary verb; CON, conditional; COP, copula; EM, emoji; (F), female exclusive; GN, genitive marker; HON, honorific form; HUM, humble form or *kenjōgo*; IMP, imperative form; KAO: kaomoji; N, noun; NEG, negative; NOM, nominalizer; (M), male exclusive; OM, direct object marker; PASS, passive form; PFX, prefix; PLN, plain form; POT, potential verb; PRT, particle; PST, past; RSP, respectful form or *sonkeigo*; Q, question; SFP, sentence final particle; SFX. Suffix; TM, subject topic marker; and V, verb.

To transcribe Japanese language into roman letter, we relied on the Hepburn romanization system, as is the most commonly used. This system is accessible to non-Japanese readers as it allows to pronounce Japanese words in a coherent manner. To be more specific, the Modified Hepburn romanization, where *hiragana*'s long vowels are transcribed as ā, ī, ū, ē, ō; in *katakana* however, long vowels are followed by a macron (*chōonpu* or *namidasshu*). In the case of place names, for instance, *Tōkyō* is the correct romanization, but for practical reasons the macrons are omitted.

Summary

This research investigated tweets whose authors are cis-gender and transgender and or transsexual Japanese women, to determine the current usage of Japanese women's language or *joseigo* in the social media setting. The lexical features of *joseigo* became the tool to measure the amount of importance given to the ideology behind such genderlect. Firstly, many authors point out that *joseigo* is an ideological construct voiced mainly through popular media, conversely disseminating stereotypical notions of women and femininity. It represents, thus, a way for women to conform to society's expectations by harnessing *joseigo* language patterns. In that case, this fact may mean that transgender women will wield *joseigo* as a means to further project femininity.

This paper is proposing a description of the usage of lexical features in a social media context, particularly Twitter. The samples are two corpora, compiled from 10 Japanese women Twitter profiles, about 100 tweets from each user. Thus, to analyze *joseigo* usage against general usage, we followed a corpus-based approach to compare the most stereotypical lexical items. In which approximately 16 femininity indexing items were found, however not all of these are considered salient or key. To calculate such saliency, we compared both corpora, cis-gender and transgender, against a reference corpus to obtain keyness values. This value indicated that out of the 16 items, 9 items have positive keyness (saliency) in the cis-gender corpus, and 8 items in the transgender corpus, thus pointing to an informal usage of *joseigo* and femininity indexing resources when using Twitter.

Keywords: *Japanese language; Japanese Women's Language; Joseigo; Sociolinguistics; Corpus Linguistics; LGBT Linguistics.*