

**Diffusion of Western Loanwords
in
Contemporary Japanese**

A Sociolinguistic Approach to Lexical Variation

Aimi Kuya

New College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

in Comparative Philology and General Linguistics

Trinity 2016

Acknowledgements

I owe my deepest gratitude to my supervisor, Bjarke Frellesvig, for his plenty of suggestions and feedback throughout planning and writing process of the thesis. I would like to thank my college advisor, Rosalind Temple, for her advice at the early stage of the present research.

I would like to express my deepest appreciation to Makiro Tanaka and Toshinobu Ogiso, who always encouraged me as advisors while I was conducting research in the National Institute for Japanese Language and Linguistics (NINJAL), Japan. I am indebted to Fumio Inoue, Masao Aizawa, Shoichi Yokoyama, Makoto Yamazaki, Eran Kim and Satoshi Nambu for providing me with invaluable scholarly guidance while I was working in NINJAL.

I owe special thanks to Takenori Nakamura for his technical support in data collection from the corpora. Special thanks also go to Anna Rayner-Grignon for proofreading my English.

I also would like to express thanks to any other anonymous predecessors, colleagues and friends who supported me in so many ways on various occasions both in the UK and Japan. Any errors remaining are the responsibility of the author.

Finally, I would not have been able to complete this work without receiving consistent support and assiduous encouragement from my family.

Abstract

Diffusion of Western Loanwords in Contemporary Japanese: A Sociolinguistic Approach to Lexical Variation

Aimi Kuya

New College

D.Phil. Thesis

Trinity 2016

The present research attempts to develop a general model of the diffusion of Western loanwords in contemporary Japanese within the variationist framework. It describes and predicts, based on empirical evidence from apparent- and real-time data, the elaborate process of changes in favor of loanwords as opposed to their existing native equivalents.

First, people's self-reporting shows a consistent tendency for a younger generation to show a stronger preference for loanwords than an elder one. This indicates changes in favor of loanwords are in progress in apparent time (Chapter 4).

Second, the above-mentioned age gradient is attested to by corpus-based data. It also reveals that the occurrence of loanwords is accounted for multi-dimensionally by a wider range of language-external factors such as generation, education, register and style (Chapter 5).

Third, an in-depth study of the individual loanword *keesu* (< *case*) reveals that not only external factors but also internal ones, e.g., usage and collocation of the word, have impacts on its occurrence (Chapter 6).

Fourth, an investigation of the loanword *sapooto* (< *support*) shows that a stylistic variable comes into play in its diffusion in interaction with an educational variable. The loanword is disfavored when the speech setting shifts to formal in particular by the most educated speakers (Chapter 7).

Fifth, a real-time approach to loanword adoption verified that individuals can change their language attitude or behavior throughout their lifetime. It highlights importance of longitudinal observation of the phenomenon in making a more accurate prediction of change (Chapter 8).

The present research confirms that the occurrence of loan variants is bound by various social and linguistic contexts. The above empirical findings contribute to the field of variationist study by opening up the possibility of analyzing linguistic variation in Japanese at the lexical level.

Table of Contents

1. Introduction	1
1.1. Problem.....	1
1.2. Aims.....	5
1.3. Structure.....	8
2. Previous studies of loanwords in Japanese	12
2.1. Backgrounds	12
2.1.1. Lexical strata in Japanese.....	12
2.1.2. History of loanwords in Japanese	16
2.1.3. Classification of loanwords.....	20
2.2. Previous approaches to loanwords.....	24
2.2.1. Quantitative lexicology.....	24
2.2.2. Corpus linguistics.....	27
2.2.3. Sociolinguistics.....	29
2.3. Framework for the present research.....	33
3. Methods	36
3.1. Data.....	36
3.1.1. Interview-based data	36
3.1.2. Corpus-based data: BCCWJ	38
3.1.3. Corpus-based data: CSJ	40
3.1.4. Complementary use of different sources of data	41
3.2. Analytical method.....	45
3.2.1. S-curve model of language change	45
3.2.2. Multiple logistic regression model and independent variables.....	47
3.3. Issues of variationist approach to lexical variation.....	50
4. Diffusion of loanwords in apparent time	56
4.1. Introduction.....	56
4.1.1. Aim: Predicting the diffusion of loanwords based on the apparent-time hypothesis	56
4.1.2. Apparent-time approach to language change.....	56
4.2. Data.....	59
4.2.1. Opinion polls.....	59
4.2.2. Semantic difference between loan and native variants	62
4.2.3. Distribution by sex.....	64
4.2.4. Distribution by age (apparent-time distribution)	65
4.2.4.1. Loanwords showing a monotonically increasing s-shaped curve	65
4.2.4.2. Loanwords showing an s-shaped curve that ends in a drop	67
4.2.4.3. Loanwords showing no systematic pattern across age	68
4.2.5. Delay in loanword diffusion among the youngest age group(s)	69
4.3. Estimating an s-curve model.....	73

4.3.1. Logistic regression analysis.....	73
4.3.2. Results	75
4.4. Predicting change.....	78
4.4.1. Cases in which neither Acquisition nor Sex is significant	79
4.4.2. Cases in which Acquisition is significant.....	80
4.4.3. Cases in which Sex is significant	82
4.5. Discussion and Conclusion.....	84
5. Distribution of loanwords in corpora: A comprehensive approach.....	87
5.1. Introduction.....	87
5.1.1. Aim: Establishing tendencies in synchronic distribution of loanwords.....	87
5.1.2. Research questions.....	88
5.2. Data.....	89
5.2.1. BCCWJ and CSJ as synchronic corpora.....	89
5.2.2. Upward tendency in loanword diffusion in apparent time.....	90
5.3. Selecting potential lexical variables	94
5.3.1. Exploring loanwords by parts of speech and usage.....	94
5.3.2. Identifying loanwords that are high in frequency	96
5.3.3. Identifying loanwords that have native variants	99
5.3.4. Identifying native equivalents.....	100
5.3.5. Scope of analysis.....	102
5.4. Analysis of the written corpus (BCCWJ)	103
5.4.1. Eliciting data	103
5.4.2. Independent variables	104
5.4.3. Year of Birth	105
5.4.4. Sex.....	108
5.4.5. Education	109
5.4.6. Register (Medium Type).....	110
5.5. Analysis of the spoken corpus (CSJ)	114
5.5.1. Eliciting data	114
5.5.2. Independent variables	115
5.5.3. Year of Birth	116
5.5.4. Sex.....	117
5.5.5. Education	118
5.5.6. Register (Speech Type).....	119
5.5.7. Spontaneity	120
5.5.8. Formality.....	121
5.5.9. Audience Size	122
5.5.10. Correlation among independent variables	123
5.6. Discussion and conclusion.....	124
6. A case study of <i>keesu</i> (< case): A contextual approach.....	129
6.1. Introduction.....	129
6.1.1. Aim: Towards a more fine-grained analysis of lexical variation.....	129
6.1.2. Backgrounds: Kim's study (2011) of the loanword <i>keesu</i>	130
6.1.2.1. Shared function of the four variants.....	131

6.1.2.2. Co-occurring predicates	135
6.1.3. Envelope of variation	138
6.2. Eliciting data (BCCWJ)	138
6.3. Internal variables	140
6.3.1. Usage	140
6.3.2. Co-occurring Predicate	145
6.4. External variables	146
6.4.1. Year of Birth	146
6.4.2. Sex	148
6.4.3. Education	149
6.4.4. Register (Medium Type)	150
6.4.5. Genre (PB only)	151
6.5. Multivariate analysis	152
6.5.1. Variables	152
6.5.2. Results	154
6.6. Additional analysis: <i>keesu</i> in CSJ	156
6.6.1. Eliciting data (CSJ)	156
6.6.2. Internal variables	156
6.6.3. External variables	157
6.6.4. Multivariate analysis	160
6.7. Discussion and conclusion	163
7. Stylistic constraints on the use of loanwords: The alternation between <i>sapooto</i> (<i>< support</i>) and its native equivalents as intra-speaker variation	167
7.1. Introduction	167
7.1.1. Aim: Elucidating the effect of style on lexical variation	167
7.1.2. Previous studies on stylistic variation	169
7.1.3. Research questions	171
7.2. Data	172
7.2.1. <i>The National Survey on Attitudes to Loanwords II</i> (NINJAL 2005a)	172
7.2.2. Contributions of the present data to the study of stylistic variation	175
7.2.3. Scope of analysis	176
7.3. Effect of external factors	177
7.3.1. Age	177
7.3.2. Education	179
7.3.3. Sex	181
7.3.4. Style	181
7.4. Multivariate analysis	183
7.4.1. Variables	183
7.4.2. Results	185
7.4.3. Predicting change	187
7.4.4. Summary of quantitative analysis	191
7.5. Qualitative analysis of stylistic variation	192
7.5.1. Distribution of style-shifting	192
7.5.2. Mechanism of switching from LW to non-LW	195
7.5.3. Mechanism of switching from non-LW to LW	197

7.5.4. People's evaluation of the loanword.....	200
7.5.5. Interpretation of style.....	201
7.6. Conclusion	202
8. A real-time approach to the diffusion of loanwords	204
8.1. Introduction.....	204
8.1.1. Aim: Towards more precise predictions of change	204
8.1.2. Background: The need for a real-time approach to lexical variation.....	204
8.1.3. Research questions.....	208
8.2. Data.....	209
8.2.1. Interview-based data	209
8.2.2. Corpus-based data.....	213
8.3. Analysis	214
8.3.1. Examining the variability of the community overall	214
8.3.2. Examining the variability of the individual speakers	221
8.4. Predicting change.....	226
8.4.1. Acceleration of change	226
8.4.2. No acceleration/deceleration of change	228
8.4.3. Deceleration of change	230
8.5. Discussion.....	232
8.5.1. Limitations on the apparent-time approach in making longitudinal predictions of change	232
8.5.2. Mechanism of lifespan change.....	233
8.6. Conclusion	236
9. Conclusions	238
9.1. General summary and discussion	238
9.1.1. Extending variationist methodology to lexical variation.....	238
9.1.2. Structured nature of variation	239
9.1.3. Language attitude and language behavior.....	241
9.1.4. Demonstration of change in real time	242
9.1.5. "Actuation problem" of Western loanwords.....	243
9.1.6. Sociolinguistic status of loanwords	246
9.2. Prospects for further research	247
9.2.1. Exploration of a wider range of lexical variables	247
9.2.2. Search for additional language-internal factors	248
9.2.3. Exploration of other sources of data	250
9.2.4. Application of a multinomial logistic regression analysis	252
9.2.5. Towards cross-linguistic comparison	253
Appendix	255
A: Data set for the apparent-time study, interview-based (Ch. 4).....	255
B1: Frequency lists by usage, BCCWJ (Ch.5).....	259
B2: Distribution of LWs and non-LWs by Year of Birth, BCCWJ (Ch. 5).....	262
B3: Distribution of LWs and non-LWs by Sex, BCCWJ (Ch.5).....	264

B4: Distribution of LWs and non-LWs by Education, BCCWJ (Ch.5).....	265
B5: Distribution of LWs and non-LWs by Register, BCCWJ (Ch.5).....	266
B6: Frequency lists by usage, CSJ (Ch.5).....	267
B7: Distribution of LWs and non-LWs by Year of Birth, CSJ (Ch.5).....	270
B8: Distribution of LWs and non-LWs by Sex, CSJ (Ch.5).....	271
B9: Distribution of LWs and non-LWs by Education, CSJ (Ch.5).....	272
B10: Distribution of LWs and non-LWs by Register, CSJ (Ch.5).....	273
B11: Distribution of LWs and non-LWs by Spontaneity, CSJ (Ch.5).....	274
B12: Distribution of LWs and non-LWs by Formality, CSJ (Ch.5).....	275
B13: Distribution of LWs and non-LWs by Audience Size, CSJ (Ch.5).....	276
C1: Data set for the real-time study, interview-based (Ch. 8).....	277
C2: Data set for the real-time study, corpus-based (Ch. 8).....	279
Bibliography	281
Dictionaries & Corpora	299

List of Tables

Table 1.1	Outline of the thesis	8
Table 2.1	Orthographical distinctions among the three lexical strata	13
Table 2.2	Examples of loanwords by period of borrowing.....	19
Table 2.3	Classification of loanwords.....	23
Table 2.4	Diachronic distribution of lexical strata in <i>Chuuookooron</i> (magazine).....	25
Table 2.5	Like and dislike of the present use of loanwords in general	30
Table 3.1	Examples of questions used in the interview-based studies.....	37
Table 3.2	Strengthes (+) and weaknesses (-) of each source.....	42
Table 4.1	Sources of data	59
Table 4.2	Selection of loanwords and their data source.....	61
Table 4.3	Distribution of loanwords by sex	65
Table 4.4	Data coding for <i>niizu</i>	75
Table 4.5	Results of the regression analyses (17 variables).....	77
Table 5.1	Distribution of the lexical strata in apparent time (BCCWJ).....	92
Table 5.2	Distribution of the lexical strata in apparent time (CSJ).....	93
Table 5.3	Lexical strata by part of speech (type, BCCWJ).....	94
Table 5.4	Criteria for making frequency lists of loanwords	97
Table 5.5	Frequency list (verbal usage)	98
Table 5.6	Selected lexical variables (BCCWJ)	101
Table 5.7	Distribution of LWs and non-LWs (BCCWJ)	104
Table 5.8	Correlation matrix (BCCWJ).....	111
Table 5.9	Selected lexical variables (CSJ).....	115
Table 5.10	Correlation matrix (CSJ).....	123
Table 5.11	Summary of the factor-by-factor analyses	125
Table 6.1	Type of predicate that co-occurs with <i>keesu</i>	136
Table 6.2	Envelope of variation	138
Table 6.3	Tokens of the lexical variants (BCCWJ)	140
Table 6.4	Effect of usage on the occurrence of <i>keesu</i> (BCCWJ).....	144
Table 6.5	Effect of co-occurring predicate on the occurrence of <i>keesu</i> (BCCWJ)....	145
Table 6.6	Effect of year of birth on the occurrence of <i>keesu</i> (BCCWJ)	147
Table 6.7	Effect of sex on the occurrence of <i>keesu</i> (BCCWJ).....	148
Table 6.8	Effect of education on the occurrence of <i>keesu</i> (BCCWJ)	149
Table 6.9	Effect of register on the occurrence of <i>keesu</i> (BCCWJ).....	150
Table 6.10	Effect of genre on the occurrence of <i>keesu</i> (BCCWJ, PB only).....	151
Table 6.11	Correlation matrix: <i>keesu</i> (BCCWJ)	153
Table 6.12	Result of the logistic regression analysis: <i>keesu</i> (BCCWJ).....	155
Table 6.13	Tokens of the lexical variants (CSJ)	156
Table 6.14	Effect of usage on the occurrence of <i>keesu</i> (CSJ).....	157
Table 6.15	Effect of co-occurring predicate on the occurrence of <i>keesu</i> (CSJ).....	157
Table 6.16	Effect of year of birth on the occurrence of <i>keesu</i> (CSJ)	158
Table 6.17	Effect of sex on the occurrence of <i>keesu</i> (CSJ).....	158
Table 6.18	Effect of education on the occurrence of <i>keesu</i> (CSJ)	158
Table 6.19	Effect of register (speech type) on the occurrence of <i>keesu</i> (CSJ)	159
Table 6.20	Effect of spontaneity on the occurrence of <i>keesu</i> (CSJ)	159

Table 6.21	Effect of formality on the occurrence of <i>keesu</i> (CSJ)	159
Table 6.22	Effect of audience size on the occurrence of <i>keesu</i> (CSJ)	160
Table 6.23	Correlation matrix: <i>keesu</i> (CSJ)	160
Table 6.24	Result of the logistic regression analysis: <i>keesu</i> (CSJ)	162
Table 6.25	Classification of co-occurring predicates by their semantic features.....	165
Table 7.1	Lexical choice across different settings	175
Table 7.2	Effect of age on the use/non-use of <i>sapooto</i>	179
Table 7.3	Proportion (%) of LW to non-LW by education and age.....	180
Table 7.4	Effect of education on the use/non-use of <i>sapooto</i>	180
Table 7.5	Effect of sex on the use/non-use of <i>sapooto</i>	181
Table 7.6	Effect of style on the use/non-use of <i>sapooto</i>	182
Table 7.7	Example of data coding (samples of those in 20s).....	185
Table 7.8	Result of the logistic regression analysis: <i>sapooto</i> (interview).....	187
Table 7.9	Patterns of style-shifting	193
Table 7.10	Patterns of shifting by Education (%)	194
Table 7.11	Reason for word choice in Public (A) (LW→LW vs. LW→SJ)	196
Table 7.12	Reason for word choice in Public (B) (LW→LW vs. LW→SJ)	196
Table 7.13	Reason for word choice in Public (A) (LW→LW vs. LW→NJ).....	197
Table 7.14	Reason for word choice in Public (B) (LW→LW vs. LW→NJ).....	197
Table 7.15	Reason for word choice in Public (A) (SJ→SJ vs. SJ→LW)	198
Table 7.16	Reason for word choice in Public (B) (SJ→SJ vs. SJ→LW)	199
Table 7.17	Reason for word choice in Public (A) (NJ→NJ vs. NJ→LW)	200
Table 7.18	Reason for word choice in Public (B) (NJ→NJ vs. NJ→LW)	200
Table 8.1	Types of change defined by (i) age differentiation, (ii) variability of the individuals and (iii) variability of the community	206
Table 8.2	Sources of data on individual lexical items: <i>wain</i> and <i>oopun-suru</i>	210
Table 8.3	Sources of data on like or dislike of the present loanword use.....	211
Table 8.4	Sources of data on like or dislike of a further expansion of loanwords.....	212
Table 8.5	Corpus-based data (1986-2005) (LB, BCCWJ).....	214
Table 8.6	Regression coefficients of the independent variables: (i) Year of Survey and (ii) Age	221
Table 8.7	Regression coefficients of the independent variables: (i) Year of Survey and (ii) Year of Birth.....	226
Table 8.8	Summary of the real-time approach.....	233
Table 8.9	Types of change (revised).....	235
Table 9.1	Number of loanwords in the surrounding context of variants.....	250

List of Figures

Figure 2.1	Historical view of vocabulary strata in Japanese.....	13
Figure 2.2	Advantages of using loanwords.....	15
Figure 2.3	<i>Gairaigo</i> timeline: 1500 – present.....	18
Figure 2.4	S-curve change in influx of loanwords.....	26
Figure 3.1	Scope of comprehensive and contextual approaches.....	54
Figure 4.1	Lexical choice by the speaker's age in Canadian English.....	57
Figure 4.2	Distribution of the LW <i>suutsu</i> (< <i>suit</i>) and <i>sebiro</i>	66
Figure 4.3	Distribution of the LW <i>wain</i> (< <i>wine</i>) and <i>budooshu</i>	67
Figure 4.4	Distribution of the LW <i>risuku</i> (< <i>risk</i>) and <i>kikensei</i>	68
Figure 4.5	Distribution of the LW <i>niizu</i> (< <i>needs</i>) and <i>hitsuyoosei</i>	68
Figure 4.6	Process of standardization over time.....	70
Figure 4.7	Diffusion of <i>niizu</i> (< <i>needs</i>) in real time.....	71
Figure 4.8	Use of <i>niizu</i> (< <i>needs</i>) by age and year of survey.....	72
Figure 4.9	Use of <i>niizu</i> (< <i>needs</i>) by age and year of birth.....	72
Figure 4.10	Adopter categorization on the basis of innovativeness.....	78
Figure 4.11	Prediction of the diffusion of <i>getto-suru</i> (< <i>get</i>).....	79
Figure 4.12	Comparing the goodness of fit of the models for <i>niizu</i> (< <i>needs</i>).....	81
Figure 4.13	Prediction of the diffusion of <i>niizu</i> (< <i>needs</i>).....	82
Figure 4.14	Prediction of the diffusion of <i>bijon</i> (< <i>vision</i>) (men ahead of women)....	83
Figure 4.15	Prediction of the diffusion of <i>oopun-suru</i> (< <i>open</i>) (women ahead of men)	83
Figure 5.1	Distribution of the LW stratum in apparent time (BCCWJ).....	92
Figure 5.2	Distribution of the LW stratum in apparent time (CSJ).....	93
Figure 5.3	Weak complementarity of distribution over time.....	102
Figure 5.4	Effect of Year of Birth on the occurrence of LWs (BCCWJ).....	106
Figure 5.5	Regression analysis of <i>keesu</i> (< <i>case</i>).....	106
Figure 5.6	Regression analysis of <i>shinpuru</i> (< <i>simple</i>).....	107
Figure 5.7	Effect of Sex on the occurrence of LWs (BCCWJ).....	108
Figure 5.8	Effect of Education on the occurrence of LWs (BCCWJ).....	109
Figure 5.9	Effect of Register on the occurrence of LWs (BCCWJ).....	110
Figure 5.10	Phonologically contracted forms by Register.....	114
Figure 5.11	Effect of Year of Birth on the occurrence of LWs (CSJ).....	116
Figure 5.12	Effect of Sex on the occurrence of LWs (CSJ).....	117
Figure 5.13	Effect of Education on the occurrence of LWs (CSJ).....	118
Figure 5.14	Effect of Register on the occurrence of LWs (CSJ).....	119
Figure 5.15	Effect of Spontaneity on the occurrence of LWs (CSJ).....	120
Figure 5.16	Effect of Formality on the occurrence of LWs (CSJ).....	121
Figure 5.17	Effect of Audience Size on the occurrence of LWs (CSJ).....	122
Figure 6.1	Diachronic distribution of <i>keesu</i> , <i>baai</i> , <i>rei</i> and <i>jirei</i> (newspapers).....	131
Figure 6.2	Diachronic distribution of <i>keesu</i> (without a modifier).....	142
Figure 6.3	Diachronic distribution of <i>keesu</i> (with a non-clause modifier).....	142
Figure 6.4	Diachronic distribution of <i>keesu</i> (with a clause modifier).....	142
Figure 6.5	Distribution of the variants by usage (BCCWJ).....	144
Figure 6.6	Distribution of the variants by co-occurring predicate (BCCWJ).....	145

Figure 6.7	Distribution of the variants by year of birth (BCCWJ)	147
Figure 6.8	Distribution of the variants by sex (BCCWJ)	148
Figure 6.9	Distribution of the variants by education (BCCWJ)	149
Figure 6.10	Distribution of the variants by register (BCCWJ)	150
Figure 6.11	Interaction between Register and Spontaneity	163
Figure 7.1	“Water Tank Model”	168
Figure 7.2	Variable /ng/ in Norwich by class and style	171
Figure 7.3	Distribution of <i>sappoto</i> by age and style	177
Figure 7.4	Lexical variation by age	178
Figure 7.5	Lexical variation by education	180
Figure 7.6	Lexical variation by sex	181
Figure 7.7	Distribution of <i>sapooto</i> by age and style	182
Figure 7.8	Distribution of <i>sapooto</i> by education and style	182
Figure 7.9	Probability of <i>sapooto</i> by Education, Style and Year of Birth	189
Figure 7.10	Interaction between Education and Style	190
Figure 7.11	A three dimensional model of the diffusion of <i>sapooto</i>	192
Figure 7.12	Patterns of shifting by Education (simplified)	194
Figure 8.1	Standardization of the variable /u/ in Tsuruoka, Japan	205
Figure 8.2	Variability of the community: <i>wain</i> (< <i>wine</i>)	216
Figure 8.3	Variability of the community: <i>oopun-suru</i> (< <i>open</i>)	216
Figure 8.4	Variability of the community: The present use of LWs in general	217
Figure 8.5	Variability of the community: A further expansion of LWs in future	217
Figure 8.6	Variability of the community: <i>keesu</i> (< <i>case</i>)	218
Figure 8.7	Variability of the community: <i>sapooto-suru</i> (< <i>support</i>)	218
Figure 8.8	Variability of the community: <i>shinpuru</i> (< <i>simple</i>)	219
Figure 8.9	Variability of the community: The LW stratum	219
Figure 8.10	Variability of the individuals: <i>wain</i> (< <i>wine</i>)	222
Figure 8.11	Variability of the individuals: <i>oopun-suru</i> (< <i>open</i>)	222
Figure 8.12	Variability of the individuals: The present use of LWs in general	223
Figure 8.13	Variability of the individuals: A further expansion of LWs in future	223
Figure 8.14	Variability of the individuals: <i>keesu</i> (< <i>case</i>)	224
Figure 8.15	Variability of the individuals: <i>sapooto-suru</i> (< <i>support</i>)	224
Figure 8.16	Variability of the individuals: <i>shinpuru</i> (< <i>simple</i>)	225
Figure 8.17	Variability of the individuals: The LW stratum	225
Figure 8.18	Acceleration of change: <i>wain</i> (< <i>wine</i>)	227
Figure 8.19	Acceleration of change: <i>sapooto-suru</i> (< <i>support</i>)	227
Figure 8.20	No acceleration/deceleration of change: <i>oopun-suru</i> (< <i>open</i>)	228
Figure 8.21	No acceleration/deceleration of change: <i>keesu</i> (< <i>case</i>)	229
Figure 8.22	No acceleration/deceleration of change: <i>shinpuru</i> (< <i>simple</i>)	229
Figure 8.23	No acceleration/deceleration of change: The LW stratum	230
Figure 8.24	Deceleration of change: The present use of LWs in genral	231
Figure 8.25	Deceleration of change: A further expansion of LWs in future	232

Abbreviations

A: adjectival usage
ACC: accusative
APS: Academic Presentation Speech
BBK: *Bunkachoo Bunkabu Kokugoka*, part of the Agency for Cultural Affairs, Japan
BCCWJ: Balanced Corpus of Contemporary Written Japanese
COMP: complementizer
COP: copula
CSJ: Corpus of Spontaneous Japanese
DAT: dative
GEN: genitive
HB: hybrid (word(s))
intr.: intransitive
ISM: Institute of Statistical Mathematics
lit.: literal translation
LW: loanword
N: nominal usage
NEG: negative
NHK: *Nippon Hoso Kyokai*, Japan's Broadcasting Corporation
NICT: National Institute of Information and Communications Technology
NINJAL: National Institute for Japanese Language and Linguistics, formerly NLRI:
National Language Research Institute
NJ: native Japanese (word(s))
NOM: nominative
Non-LW: non-loanword(s), including Sino-Japanese (SJ) words and native-Japanese
(NJ) words
NSDKK: *Naikaku Soori Daijin Kanboo Koohooshitsu*, Public Relations Office of
Cabinet Secretariat
PASS: passive
PAST: past tense
PB: Books in the Publication sub-corpus, BCCWJ
PM: Magazines in the Publication sub-corpus
PN: Newspapers in the Publication sub-corpus
POL: polite form
PROG: progressive
SJ: Sino-Japanese (word(s))
SPS: Simulated Public Speech
TOP: topic marker
tr.: transitive
V: verbal usage

1. Introduction

1.1. Problem

This thesis is, to my knowledge, the first attempt to capture the diffusion of Western loanwords, called *gairaigo*, in contemporary Japanese within the variationist framework. A large influx of lexical borrowings from Western languages, mostly from English, is one of the most noticeable linguistic changes in contemporary Japanese.

In the very early stage of studies of lexical borrowings, as Mougeon & Beniak (1991) implies, the most universal reason for lexical borrowing has been attributed to its linguistic (“gap filling”) function. This type of borrowing, according to Myers-Scotton’s definition (1993: 169, 2006: 212-213), is called “cultural borrowing.” Cultural borrowing occurs when a recipient language needs to fill gaps of its lexicon in order to refer to things or concepts new to the language. Such examples are often found in the fields of science, state-of-the-art technology and other fields of advanced cultural development. One of the most common examples would be computer-related terms such as *computer*, *email*, *software*, *website*, and so on. (Myers-Scotton 2006: 213), all of which have also entered into Japanese as *konpyuutaa*, *iimeeru*, *sofutoueaa*, *uebusaito*, respectively.¹

Myers-Scotton (ibid) proposes the existence of another type of borrowing, i.e. “core borrowing,” which occurs even when a recipient language already has equivalent native

¹ Throughout the thesis Japanese forms are transliterated based on the modified Hepburn System in *Kenkyuusha’s New Japanese-English Dictionary*, 4th Edition (1974: xiii), except proper nouns and the representation of long vowels. Proper nouns are transliterated according to their official name in the origin language where available. A long vowel is generally indicated with double letters: /a:, i:, u:, e:, o:/ are indicated as *aa*, *ii*, *uu*, *ee*, *oo*, respectively. A long vowel /e:/ for Sino-Japanese words is however spelled exceptionally as *ei*, following the convention of this dictionary.

expressions in its lexicon. Hashimoto (2010) and Kim (2011) report that there has been an increasing use of loanwords in Japanese over the 20th century, many of which are abstract words that have existing native equivalents. Compare the following examples taken from the Balanced Corpus of Contemporary Written Japanese (BCCWJ):

(1.1)

a. 持ち家 = “家を所有” することのメリットとは、(PB13_00092)

Mochi-ie = “*ie-o shoyuu-suru*” *koto-no merito-towa*,

One's.own-house = house-ACC possess-do thing-GEN merit-TOP

‘The advantage of “having your own home” is,’

b. このプログラミングモデルの利点は、(PB13_00131)

kono puruguramingu moderu-no riten-wa,

this programing model-GEN merit-TOP

‘The advantage of this programing model is,’

In the above examples, the loanword *meritto* (< *merit*) in (1.1.a) and the existing word *riten* (both meaning ‘advantage’) in (1.1.b) share the same general area of meaning and they are interchangeable without largely changing the meaning of the whole sentence.² Few studies, however, have shed light on the phenomenon where the lexical alternation occurs between core lexical borrowings and existing lexical items.

While cultural borrowing is considered to be necessary to meet the lexical needs of a language, core borrowing is characterized as unnecessary or “gratuitous”

² One might point out that there is a slight difference in nuance or the emotional tinge of a phrase. This is similar to the discussion in Weiner & Labov (1983), as to whether the existence/absence of a passive structure triggers a change in meaning. I will come back to this point in Chapter 3.

(Myers-Scotton 2006: 215, Mougeon & Beniak 1991: 199). A very simple question then arises as to why “borrowing always goes well beyond the actual ‘needs’ of a language” (Haugen 1953: 373).

Corpus linguists may point out that there is a semantic division of labor between borrowed items (as core borrowings) and their existing near-synonyms. This is true in the sense that in principle none of loanwords has exactly the same meaning as its native equivalents do and that many of “established” loanwords had experienced diachronic semantic and structural developments to a certain degree after entering Japanese. It is, however, also true that these near-synonyms always share the same general area of meaning and that alternations between near-synonyms in such semantically overlapped areas cannot be fully accounted for only by linguistic factors.

Language-external reasons, then, need to be considered in the investigation of why people use such “gratuitous” foreign words as opposed to existing native words. Weinreich (1968) points out the possibility that bilingual speakers use core borrowing to display positive social values that the foreign words symbolize:

“Why use the word *belt* when the [Irish] word *crios* exists? Surely here the feeling that English is a superior language, and that an English word confers distinction upon an Irish sentence, plays an important part.” Those American immigrants who borrow as “heavily” as possible to show their advanced state of acculturation, act upon a similar motive. (Weinreich 1968: 60)

On the one hand, in the context of English as a world language, the recent influx of (core) borrowings mainly from English into Japanese demonstrates the prestigious nature of English over Japanese. On the other hand, Japanese is not completely non-resistant to the spread of Western loanwords as the history of these loanwords is

not so long as that of existing vocabulary such as *wago* (native Japanese words) and *kango* (Sino-Japanese words). Considering such historical backgrounds, it is presumed that loanwords have a particular social value and occupy a particular position in the Japanese lexical system/strata, functioning as more than mere near-synonyms of native equivalents.

The importance of the concept of core borrowing is that it enables us to treat loanwords as competing lexical variants of their native equivalents within the variationist framework. Many variationist studies have put emphasis on the importance of integrating language-external dimensions into description of variation, in addition to language-internal ones (Labov 1972, among others). Japanese is not the exception: a correlation between variation and social contexts was repeatedly reported in the study of phonological, syntactic, lexical, and pragmatic variation in Japanese (Hibiya 1988a, Matsuda 1995, Sano 2009, Nambu 2014, Inoue 2000, Yokoyama, Asahi & Sanada 2008, etc.). These studies tell us that factors other than linguistic ones, e.g. social attributes and speech style, are expected to have an effect on the distribution of loanwords.

This thesis takes a variationist approach to lexical variation between loanwords and their native equivalents in order to see in what manner the competition occurs between the two. It is expected that the linguistic heterogeneity as shown in (1.1) exhibits systematic structure of lexical variation in terms not only of language-internal factors but also of language-external ones. These two dimensions will have to be taken into account for a more comprehensive description of the phenomenon.

1.2. Aims

Based on the above discussion, this thesis aims to tackle and answer the following questions.

QUESTIONS

- I. Where does lexical alternation occur between a loanword and non-loanword(s)?
- II. Is lexical variation socially structured? If so, which language-external factors affect lexical choice in favor of loanwords?
- III. Is there a contrast between people's stated attitudes towards loanwords and their actual behavior?
- IV. How does the diffusion of loanwords proceed in real time?

These questions will be examined by means of interview- and/or corpus-based approaches to individual and groups of loanwords.

The first question is a theoretical and methodological argument over how to identify the environments in which lexical variation occurs, i.e., the envelope of variation. This includes the process of identifying competing lexical variants (near-synonyms) as well as locating linguistic contexts where the more than two variants can be used for the "same meaning." Although this procedure is essential for the study of linguistic variation, determining linguistic "equivalence" on a lexical level is not straightforward in practice because words are often polysemous and their meanings entail delicate nuances on a discourse level. In this thesis, the notion of linguistic equivalence is

applied after some modification of the original definition of linguistic variable (Labov 1972).

The second question concerns the problem of social and stylistic dimension of the loanword distribution. The social aspects of loanword diffusion are studied by NINJAL (National Institute for Japanese Language and Linguistics, formerly NLRI: National Language Research Institute) (2004, 2005a, 2007b), Jinnouchi (2007), NHK (*Nippon Hoso Kyokai*, Japan's Broadcasting Corporation) (1980, 1996) and BBK (*Bunkachoo Bunkabu Kokugoka*, part of the Agency for Cultural Affairs) (1995, 1999, 2000, 2003, 2008, 2013, etc.). Their investigations, however, mainly focused on age differentiations, in which younger speakers show a more positive attitude towards the use of loanwords.³ Given that age differentiation is an indication of an ongoing change towards the growing use of loanwords, we need to capture a more elaborate process of the change by investigating effects of a wider range of social variables. One of the main goals of this thesis is, then, to provide a “multivariate” model of word choice in favor of loanwords, with independent variables including age, sex, education and style as well as interactions among them.

The third question is set to overcome the problem of the previous sociolinguistic approaches: That is, they often focus on the dimension of language “attitude” based on interviews. This thesis attempts to undertake, in addition to an interview-based approach, a corpus-based approach to the phenomenon in order to shed light on people's linguistic behavior. It will enable us to inquire whether a particular pattern observed in interview-based data is attested to by evidence from actual language behavior.

³ There are few studies on social factors other than age, e.g. education (Ishiwata 1965), occupation (NINJAL 2004) and register (NINJAL 2005a).

The fourth question requires a shift of methodological paradigm from apparent-time to real-time. Sociolinguistic studies on loanword diffusion, and variationist studies in general, usually rely on apparent-time approaches, considering age differentiation to be a reflection of change in progress in real time. However, Sankoff, G. (2006: 115) points out that, despite the powerfulness of the concept in locating the presence of change, the apparent-time hypothesis often underestimates the rate of change. My purpose here is therefore to make a more accurate prediction of change by making a longitudinal observation of the process of loanword diffusion.

Weinreich, Labov and Herzog (1968: 183-7) raises the following five problems to be solved for the theory of language change: (i) constraints, (ii) transition, (iii) embedding, (iv) evaluation and (v) actuation. These problems will be explored through discussions on the above-mentioned four research questions.

First, the *constraints problem* discovers possible constraints on the form, direction, or structural character of linguistic changes (Labov 1994: 115). This problem is answered through the investigation of Question I.

Second, the *transition problem* inquires how a linguistic change is “transmitted across successive generation” (Labov 2001a: 78). This problem is tackled through the examination of Question II and IV, where age distribution in apparent time and real time is explored in depth.

Third, the *embedding problem* is to observe how a linguistic change is embedded in linguistic and social structures. This problem is discussed through the response to Question II; the investigation of the effect of social, stylistic and linguistic factors on the

on-going change will tell us how the change in question is embedded in the Japanese language as well as in a larger social context of the speech community.

Fourth, the *evaluation problem* deals with the relationship between social awareness and the process of change in behavior. This is to understand the ways in which social awareness is imposed upon the process of change (Weinreich et al. 1968: 186). This problem is discussed by contrasting people’s attitudes and behavior in exploring Question III.

Fifth, the *actuation problem* is the question of why a certain change occurred here and now (Labov 2001a: 466). The answer to this question will be discussed at the end of the thesis (Chapter 9), based on all the findings in the current research.

1.3. Structure

Structure of the thesis is outlined in Table 1.1.

Table 1.1 Outline of the thesis

Preliminaries	Ch1 Introduction		
	Ch2 Literature review		
	Ch3 Method		
Main analyses		Attitude (interview-based)	Behavior (corpus-based)
	Apparent time	Ch4 Analysis I (17 variables) Ch7 Analysis IV (case study)	Ch5 Analysis II (25 variables) Ch6 Analysis III (case study)
	Real time	Ch8 Analysis V (8 variables, interview + corpus)	
Conclusion	Ch9 Conclusion		

Chapter 2 provides a review of the previous approaches to loanwords in Japanese, with introductions to the structure of the Japanese lexicon, a history of Western loanwords in Japanese and classification of loanwords from a variationist point of view. There are three major perspectives on the study of Western loanwords: (i) quantitative

lexicology, (ii) corpus linguistics and (iii) sociolinguistics. Empirical investigations of each field have provided important findings about the loanword diffusion. There seems to be, however, a lack of connection between these different approaches: It seems that more effort will have to be devoted in order to combine their contributions into coherent research. This chapter aims to review important discoveries made in each field of loanword study, to evaluate advantages and disadvantages of each approach and thereby to search for the methodology necessary for integrating these different approaches into one in order to lead to a more comprehensive understanding of the phenomenon.

Chapter 3 outlines the methodological details of the thesis. First, a brief description is given of the research framework, the variationist approach to lexical alternation. It includes discussion on the theoretical issue of how to identify the envelope of variation for lexical variation (QUESTION I). Second, I will give brief introductions to data sources. Two major data sources, interview-based and corpus-based, are examined in this project. As a single database is not sufficient for eliciting all of the independent variables of interest, I will make use of various databases in a complementary manner. Third, an introduction to an analytical model, the logistic regression model, is provided. I will show the set of independent variables under consideration: three social and one stylistic ones.

Chapters 4 – 8 constitute the main body of the thesis, providing five analyses. These chapters are designed to investigate the phenomenon from two different dimensions: *attitude vs. behavior* and *apparent-time vs. real-time*. Chapters 4 and 7 examine people's *attitudes* towards loanwords synchronically based on the data collected from several existing nationwide interviews (opinion polls) conducted between 1999 and the

early 2000s. The aims of these chapters are to answer the question of which language-external factors are the leading causes of the diffusion of loanwords and in what manner they bring about the diffusion (QUESTION II). Chapters 5 and 6 investigate the other dimension, people's actual linguistic *behavior*, based on the data collected from written and spoken corpora. Since these corpora, as with the above-mentioned interviews, concern language produced between 1999 and the early 2000s, it is reasonable to contrast the interview-based data with the corpus-based data. The comparison of *attitude* and *behavior* will provide a useful insight into the validity of arguments made in earlier investigations, which have relied heavily on linguistic attitude (QUESTION III).

Chapter 8, as opposed to Chapters 4 – 7, uses a longitudinal approach to the diffusion of loanwords, for which no systematic attempt has yet been made in sociolinguistic studies of loanwords. The aim of this chapter is to have a look at the diffusion of loanwords in real time by means of observing both interview- and corpus-based data. Although not much data is currently available for systematic real-time analyses, I will attempt several case studies on whether or not people's attitudes and behavior changes over their lifetime. It will be shown that a closer examination of the variability of individuals in *real time*, which the *apparent-time* approach does not presuppose, allows a more accurate prediction of change on a long-term basis (QUESTION IV).

Chapter 9 summarizes and integrates all the findings made in this thesis. Based on the sociolinguistic findings from the earlier chapters, I will discuss what kind of social

values loanwords are given in the Japanese vocabulary. I will also raise some points that remain uncovered, which should be addressed in future research.

2. Previous studies of loanwords in Japanese

2.1. Backgrounds

2.1.1. Lexical strata in Japanese

Lexicon of Japanese is roughly divided into three strata depending on its origin: *wago* (lit. ‘Japanese indigenous words,’ henceforth native Japanese words or NJ), *kango* (lit. ‘Chinese words,’ henceforth Sino-Japanese words or SJ) and *gairaigo* (lit. ‘alien words, words coming from abroad,’ henceforth loanwords or LW).⁴ These strata are differentiated in terms of their (i) history, (ii) orthography as well as (iii) sociolinguistic values in the language.

The history of the three lexical strata is summarized as in Figure 2.1. The native Japanese words (NJ) originate in the ancient Japanese and constitute an integral part of the Japanese lexicon. This is the only stratum contains all parts of speech including the lexical and functional categories (Kageyama & Saito 2016: 17, see also Table 5.3). Sino-Japanese words (SJ) started to enter into Japanese from Chinese in the middle of the 8th century, while loanwords (LW, called “foreign” in Figure 2.1) started to enter in the 16th century. These two lexical strata are similar in a grammatical sense in that they were brought into the language as nouns without containing conjugable predicates (ibid: 21, 26). They can also function as verbs when verbalized with the NJ light verb *-suru* (‘do’) or adjectives when followed by the copula *-da* (‘be’) as shown in Table 2.1 (see also 5.3.1 for more detailed examples).

⁴ There is another stratum called in Japanese as *konshugo/konseigo* (lit. ‘mixed words,’ henceforth hybrid or HB), where a word is formed by the combination of words or morphemes from different strata.

With respect to orthography, NJ is typically written in *hiragana*, *kanji* (‘Chinese characters’) or the blend of *kanji* and *hiragana* syllabaries, whereas SJ and LW are in principle written in *kanji* and *katakana*, respectively as shown in Table 2.1.

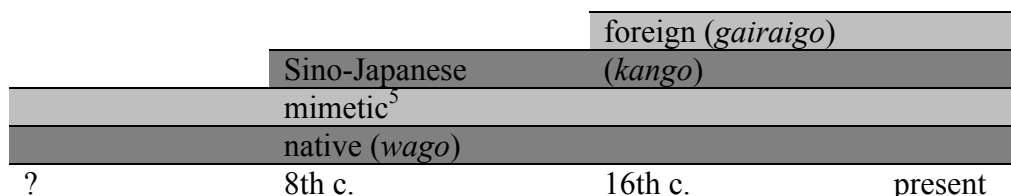


Figure 2.1 Historical view of vocabulary strata in Japanese (Kageyama & Saito 2016: 15, Figure 1)

	writing	examples		meaning
Native Japanese (NJ)	<i>hiragana</i>	が	<i>ga</i>	particle.NOM
		これ	<i>kore</i>	this
	blend	場合	<i>baai</i>	case
		支える	<i>sasaeru</i>	to support
Sino-Japanese (SJ)	<i>kanji</i>	新しい	<i>atarashii</i>	new
		事例	<i>jirei</i>	case
		支援(する)	<i>shien(-suru)</i>	(to) support
		単純(だ)	<i>shinpuru(-da)</i>	simple
Loanwords (LW)	<i>katakana</i>	ケース	<i>keesu</i>	case
		サポート(する)	<i>sapooto(-suru)</i>	(to) support
		シンプル(だ)	<i>shinpuru(-da)</i>	simple

⁵ Note that “mimetic” words are included as an independent stratum in the table since the lexical items in this category often behave differently from those in the other three strata in terms of phonology, morphology and syntax. Kageyama & Saito, however, attract our attention to the fact that most mimetics are considered to have as long history as NJ words do (2016: 14).

These strata are also differentiated with regard to sociolinguistic connotations. NJ is more closely associated with the vernacular, everyday and colloquial style. SJ, in contrast, is associated with more formal and prestigious settings such as public, ceremonial and academic writings. The contrast between NJ and SJ in Japanese seems to correspond with that between the Germanic stratum and the French/Latinate stratum in English. The latter stratum is generally considered more sophisticated in both languages as their donor languages used to be recognized as languages with a more advanced culture. For example, in Japanese, the SJ word *tanjun* ('simple') has a formal connotation whereas the LW *shinpuru* (< *simple*) has a softer and colloquial connotation (Table 2.1). This is just as, for instance, the Germanic vocabulary *buy* sounds more colloquial than its French/Latin donor equivalent *purchase* (Kageyama & Saito 2016: 20).

Loanwords, in contrast with SJ, have been imported relatively recently (this history is further reviewed in 2.1.2). The majority of them are from Western languages including English, French, German, and Dutch (NINJAL 1987: 65, Hashimoto 2010). Loanwords are in general recognized as giving novel, stylish/fashionable and intelligence nuances to one's speech (Figure 2.2).⁶

⁶ It is noted that these three evaluations of loanwords are associated with the functions of core borrowings. In contrast, "expresses a new concept or thing," "reduces negative nuances of existing words" and "rephrases an indecent expression" seem to describe the functions of cultural borrowing, in which a word is borrowed primarily for language-internal reasons. Moreover, "boosts communicative effectiveness" seems to be associated with the function of technical terms (as part of cultural borrowings).

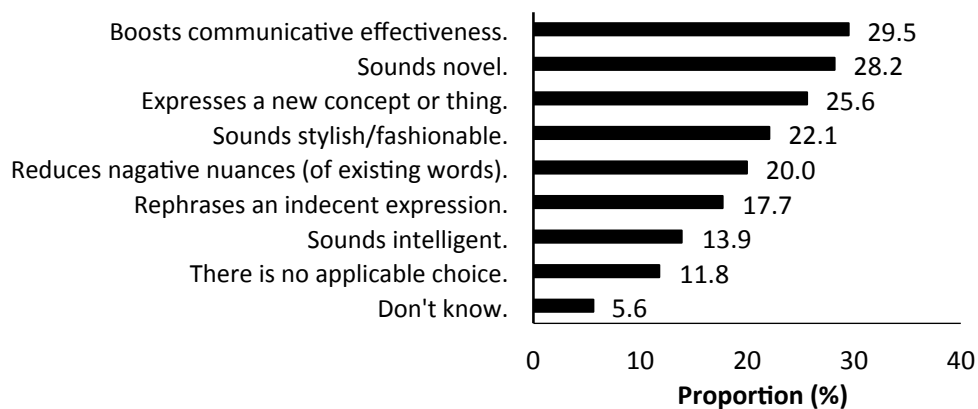


Figure 2.2 Advantages of using loanwords
(NINJAL 2004: 17, Multiple answers up to two [total=174.2%], N=3087)

Both LW and SJ, as opposed to NJ, are classified as borrowings from a historical point of view. However, Japanese speakers in general recognize LW and SJ as distinct vocabulary types for various reasons. Ishiwata (2001: 9) states that SJ vocabulary does not generally fall into the LW stratum in contemporary Japanese for the following reasons: (i) SJ words entered into Japanese much earlier than Western loanwords; (ii) SJ words have developed as a writing system/script for NJ words rather than merely as a class of vocabulary in Japanese, which had not invented its own writing system yet (the *hiragana* and *katakana* syllabaries were created after the adoption of Chinese characters) and (iii) Chinese characters are well integrated into the Japanese vocabulary system because of their high productivity. SJ has thus acquired a special (prestigious) status in Japanese as opposed to that of the LW stratum. In these respects, it would not be an exaggeration to say that SJ tends to be considered part of “native words” by Japanese speakers. As Haspelmath (2009: 38) says, “the status of native words is always relative to what we know about the history of language.” Accordingly, SJ words, as

with NJ words, are henceforth referred to as “native words” or “non-loanwords” as opposed to LW throughout this thesis.

2.1.2. History of loanwords in Japanese

The influx of Western loanwords into Japanese occurred in the following three phases: (i) Iberian period, (ii) Dutch-learning period and (iii) Western period (Irwin 2011: 23).

The first phase that started from the mid-16th century witnessed the influx of lexical borrowings mainly from Portuguese (Po), along with a small number of Latin, Spanish and Dutch words, initiated by Jesuits. These borrowings are divided into two groups: those associated with Christianity and those associated with trade (Ishiwata 2001: 191). The former includes *kirishitan* (< *cristão* (Po) ‘Christian’) and *misa* (< *missa* (Po) ‘mass’) (Irwin 2011: 29, 32) and the latter includes *kasute(e)ra* (< *Castela* (Po) ‘type of sponge cake’), *tabako* (< *tabaco* (Po) ‘tobacco, cigarette’) and so on (ibid: 33, 34, see also Ishiwata 2001: 193-226).

Japanese entered the second phase in the mid-17th century (after *Sakoku*, ‘the Edicts of Seclusion’ or ‘Isolation policy’), where only the Dutch were permitted to trade with Japan. During this time, a number of Dutch (Du) words were borrowed by Japanese highbrows learning European science. This political background promoted lexical borrowings that are associated not only with trading but also with the fields of medicine and science. The former involves *koohee* (< *koffie* (Du) ‘coffee’), *garasu* (< *glas* (Du) ‘(pane of) glass’) and so on. The latter involves *korera* (< *cholera* (Du) ‘cholera’), *pesuto* (< *pest* (Du) ‘plague, Black Death’), and so on, which were only fully understood by a few in the upper class (Irwin 2011: 38, 39-40, see also Ishiwata 2001: 243-245).

The last phase started with the beginning of Meiji Restoration in 1868 (Irwin 2011: Section 2.4). Lexical borrowings in this period were motivated by eagerness to stand as equals with Western countries by means of modernization. These include those borrowed from Russian (Ru), French (Fr), German (Ge) and English. But lexical borrowing became especially oriented to English after WWII with the rise of the economic and political power of the US. Examples from languages other than English include *sobieto* (< совет (Ru) ‘soviet’), *arakaruto* (< *à la carte* (Fr) ‘à la carte’), *bakansu* (< *vacances* (Fr) ‘holiday, vacation’), *meetoru* (< *mètre* (Fr) ‘meter’), *kapuseru* (< *Kapsel* (Ge) ‘capsule’), *arubaito* (*Arbeit* (Ge) ‘part-time job’), *teeze* (< *These* (Ge) ‘thesis’), and so on (ibid: 46, 49-50, 52-53).

Borrowings from English before WWII include those associated with semantic fields of sports (e.g., *riigu* (< *league*)); music (e.g., *jazu* (< *jazz*)); politics (e.g., *suroogan* (< *slogan*)) and fashion (e.g., *suutsu* (< *suit*)) (ibid: 56). Despite the government campaign to ban the use of English donor loanwords during WWII,⁷ the influx of loanwords from English continued after the war due to the re-rise of Anglicism. This includes *sekushii* (< *sexy*), *sutoresu* (< *stress*), *insutanto* (< *instant*), *niizu* (< *needs*), *ruutsu* (< *roots*), and so on (ibid: 57).

Figure 2.3 summarizes a history of loanwords in Japanese roughly between 1500 and 2000. Table 2.2 shows several examples of loanwords borrowed in each phase.

⁷ During the war, the government restricted the use of loanwords. Many of these were rephrased as “native words” including SJ words. For example, *nyuusu* (< *news*), *anaunsa* (< *announcer*) and *sutoraiku* (< *strike* (baseball)) were rephrased as *hoodoo* (報道 ‘report’), *hoosoin* (放送員 ‘broadcaster’) and *yoshi* (よし ‘good’), respectively (Irwin 2011: 57).

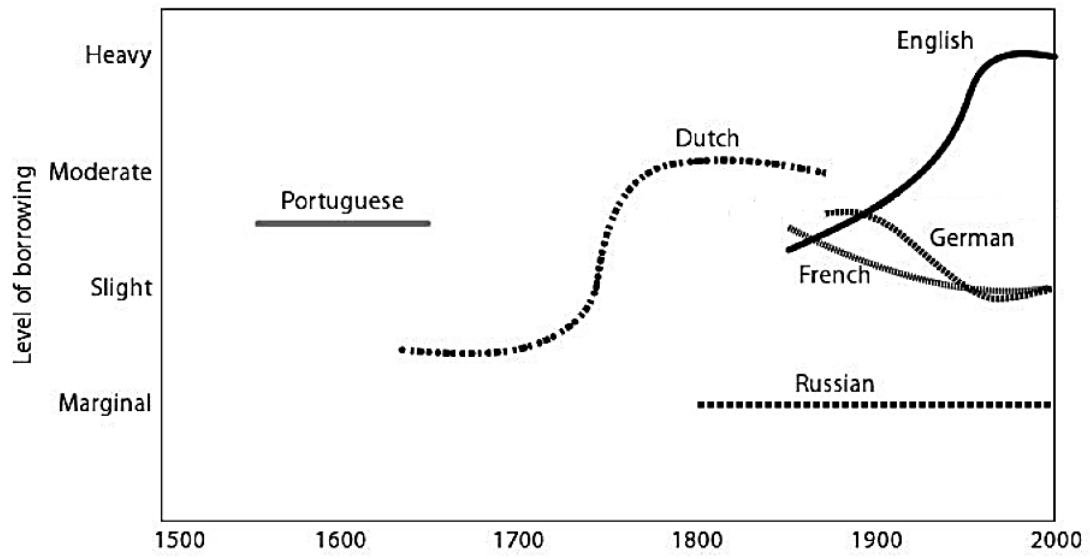


Figure 2.3 *Gairaigo* timeline: 1500 – present
(Irwin 2011: 25, Figure 2.1)

Table 2.2 Examples of loanwords by period of borrowing
(source, Irwin 2011: Sections 2.2 – 2.4)

Iberian period (1549-1639)

Christianity
kirishitan ‘early Japanese Catholic’ [1587]*
pan ‘holy wafer > bread’ [1591]
misa ‘mass’ [1591]

Trade
 1) foodstuffs: *kasute(e)ra* ‘type of sponge cake’ [1625]
 2) cloth & clothing: *kappa* ‘raincoat’ [1608]
 3) drugs & medical goods: *miira* ‘mummy’ [1610], *tabako* ‘cigarette’ [1607]
 4) tools & appliances: *furasuko* ‘decanter’ [1625], *karuta* ‘traditional playing card’ [1596]
 5) place names: *igirisu* ‘England > UK’ [1613]

Dutch-learning period: *Sakoku* (isolation) period (1639-1859)

Medicine & science – only understood by a few in the upper classes
korera ‘cholera’ [1793]
pesuto ‘plague, Black Death’ [1829]
morumotto ‘guinea pig’ [1872]

Trade – penetrated the daily lives
 1) food & drink: *koohii* ‘coffee’ [1615], *biiru* ‘beer’ [1724]
 2) materials: *garasu* ‘pane of glass’ [1763], *gomu* ‘rubber’ [1822]
 3) leisure: *orugooru* ‘music box’ [1803], *dansu* ‘dance’ [1831]
 4) maritime terminology: *dekki* ‘deck’ [1857]
 5) foreign currencies: *pondo* ‘pound sterling’ [1822]
 6) place names: *doitsu* ‘Germany’ [1725]

Western period (1859-present)

Before WWII (Rise of Anglicism)
hoteru ‘hotel’ [1850], *hausu* ‘house’ [1864]
Student slangs: *goorudo* ‘gold’, *mazaa* ‘mother’, *getto* ‘get’, *puree* ‘play’
Baseball: *beesubooru* ‘baseball’ [1889], *auto* ‘out’ [1896], *seefu* ‘safe’ [1906]
Sports: *riigu* ‘league’ [1907], *fan* ‘fan’ [1920]
Music: *jazu* ‘jazz’ [1924]
Politics: *suroogan* ‘slogan’ [1924]
Fashion: *burausu* ‘blouse’ [1925], *suutsu* ‘suit’ [1935]

During WWII (De-Anglicization)
nyuusu ‘news’ > *hoodoo* 報道 ‘report’
anaunsaa ‘announcer’ > *hoosoin* 放送員 ‘broadcaster’
sutoraiku ‘strike (baseball)’ > *yoshi* よし ‘good’

After WWII (Re-rise of Anglicism)
sekushii ‘sexy’ [1956]
sutoresu ‘stress’ [1957]
insutanto ‘instant’ [1960]
niizu ‘needs’ [1975]
ruutsu ‘roots’ [1977]

* [] indicates first written attestation.

Lexical borrowings after the Meiji Restoration in the late 19th century also include loan translations (calques), where many abstract words, mainly academic and technical terms, were translated into SJ words/morphemes. Examples of loan translations involve (i) 社会 *shakai* (< *society*), (ii) 個人 *kojin* (< *individual*), (iii) 哲学 *tetsugaku* (< *philosophy*) (Ishino 1983: 30, Ishiwata 2001: 11, Irwin 2011: 44), (iv) 電子計算機 *denshi-keisanki* (< (*electronic*) *computer*), (v) 写真機 *shashinki* (< *camera*), (vi) 葡萄酒 *budooshu* (< *wine*), and so on. The first three examples are now well integrated into Japanese. As these examples are written in *kanji* (Chinese characters) and do not maintain their donor language pronunciations, ordinary Japanese speakers may use them without being aware that they are borrowed terms. In contrast, the latter three examples failed to obtain established status in contemporary Japanese, having been almost completely replaced by their homophonic translations (i.e., their *katakana* equivalents) such as コンピュータ(一) *compyuuta(a)* (< *computer*), カメラ *kamera* (< *camera*) and ワイン *wain* (< *wine*).

2.1.3. Classification of loanwords

One may notice that early borrowings in Table 2.2 (e.g., those borrowed during the Iberian period), except for Christian terminologies, are primarily related to things (foods, clothes, medical goods, tools, place names, materials, foreign currencies, etc.) that were new to Japanese people. It should be noted that there are no native equivalents in the Japanese lexicon to rephrase them. In contrast, many recent loanwords, particularly abstract words such as *reberu* (< *level*), *ruuru* (< *rule*) and *keesu* (< *case*), are coexistent with existing native equivalents (Hashimoto 2010: 115). For example, *keesu* (< *case*) is

considered equivalent to native words such as *jirei*, *baai* and *rei* (Kim 2011, see also Chapter 6).

This indicates that loanwords in contemporary Japanese are roughly divided into three groups based on their coexistence with equivalent native words as shown in Table 2.3. First, the examples in (I) have no established native equivalents. The first two examples *terebi* テレビ (< *television*) and *botan* ボタン (< *button*) are those that were borrowed into Japanese to introduce new things to the language at a very early phase of Japan's language contact history. More recent loanwords such as *deeta* データ (< *data*), *kurikku* クリック (< *click*) and *intaanetto* インターネット (< *internet*) are also examples of lexical borrowings that introduce new things to the language.

In contrast, loanwords listed in (II) have native equivalents. First, loanwords such as *keesu* ケース (< *case*), *sapooto* サポート (< *support*), *shinpuru* シンプル (< *simple*) and *charenji* チャレンジ (< *challenge*) have well-attested native equivalents, which are frequently used and are in competition with loanword equivalents. Second, loanwords such as *konpyuut(a)* コンピュータ(一) (< *computer*), *kamera* カメラ (< *camera*), *wain* ワイン (< *wine*) and *basu* バス (< *bus*) also have native equivalents via loan translations such as *denshi-keisanki* 電子計算機 (< *computer*), *shashinki* 写真機 (< *camera*), *budooshu* 葡萄酒 (< *wine*) and *norai-jidoosha* 乗合自動車 (< *bus*), respectively. These calques were once in use but have been made obsolete, so they are no longer in active competition with their loanword (*katakana*) equivalents. Third, loanwords such as *aakaibu* アーカイブ (< *archive*) and *aidentiti* アイデンティティ (< *identity*) are the most recent examples of lexical borrowings from English. Words

such as *hazon-kiroku* 保存記録 (< *archive*) and *dokuji-sei* 独自性 (< *identity*) were recently calqued as the native equivalents of these loanwords (NINJAL 2007b). One of the aims of these calques is to avoid communicational difficulties caused by the “over-use” of loanwords, by helping loanword illiterate people, including the elderly and non-professionals, to understand the meaning of unpopular loanwords (Aizawa 2012: 141-3).⁸ The use of these calques, however, does not seem to be gaining in popularity at this point.

Examples in (III) include in-betweens, which have native equivalent words but may not be as freely inter-exchangeable with native words as those in (II). In this category, there appears to be a relatively clear “division of labor” among loanwords and their native equivalents. For example, the loanword *supuun* (< *spoon*) is differentiated from its NJ near-synonym *saji* (‘a Japanese traditional spoon’) in terms of material or use (Jinnouchi 1993). The former typically refers to utensils made of metal and used for eating Western dishes, whereas the latter is generally associated with those made of wood and used for eating Japanese dishes. Jinnouchi reports that such semantic differentiation between the two lexical items has been gradually disappearing in apparent time. That is, *spuun*, as opposed to *saji*, is obtaining an “unmarked” status among younger generations as a result of semantic broadening of the loanword.

⁸ Note that this movement is not a case of language purification like those that took place in France or South Korea. The public pressure for purification of language in Japan is not as strong as in these countries (Kajiki, et al. 1995, Yang 2005).

Table 2.3 Classification of loanwords

	loanwords (homophonic translation)	native equivalents (incl. loan translation)
I	<p>early attestation</p> <p><i>terebi</i> テレビ (< <i>television</i>) <i>botan</i> ボタン (< <i>button</i>)</p> <p>recent attestation</p> <p><i>deeta</i> データ (< <i>data</i>) <i>kurikku</i> クリック (< <i>click</i>) <i>intaanetto</i> インターネット (< <i>internet</i>)</p>	<p>?</p> <p>?</p> <p>?</p> <p>?</p> <p>?</p>
II	<p>well-established</p> <p><i>sapooto</i> サポート (< <i>support</i>) <i>keesu</i> ケース (< <i>case</i>) <i>shinpuru</i> シンプル (< <i>simple</i>) <i>charenji</i> チャレンジ (< <i>challenge</i>)</p> <p>obsoleted</p> <p><i>konpyuut(a)</i> コンピュータ(一) (< <i>computer</i>) <i>kamera</i> カメラ (< <i>camera</i>) <i>wain</i> ワイン (< <i>wine</i>) <i>basu</i> バス (< <i>bus</i>)</p> <p>not in common use</p> <p><i>aakaibu</i> アーカイブ (< <i>archive</i>) <i>aidentiti</i> アイデンティティ (< <i>identity</i>)</p>	<p><i>shien</i> 支援, <i>tedasuke</i> 手助け <i>baai</i> 場合, <i>jirei</i> 事例, <i>rei</i> 例 <i>tanjun</i> 単純, <i>kanso</i> 簡素 <i>choosen</i> 挑戦</p> <p><i>denshi-keisanki</i> 電子計算機 <i>shashinki</i> 写真機 <i>budooshu</i> 葡萄酒 <i>noriai-jidoosha</i> 乗合自動車</p> <p><i>hozon-kiroku</i> 保存記録, <i>kiroku-hozon-kan</i> 記録保存館 <i>dokuji-sei</i> 独自性, <i>jiko-ninshiki</i> 自己認識</p>
III	<p><i>supuun</i> スプーン (< <i>spoon</i>) <i>hoteru</i> ホテル (< <i>hotel</i>)</p>	<p>distinguished from loanwords in meaning</p> <p><i>saji</i> さじ (‘Japanese traditional spoon’) <i>ryokan</i> 旅館 (‘Japanese traditional hotel’)</p>

The focus of the present research is on those LW in (II) (emphasized with a bold frame in Table 2.3); that is, those that coexist with native equivalents. It is worth pointing out that the distinction between (I) and (II), i.e., the presence/absence of native equivalents, does not necessarily correspond to the dividing line between cultural borrowing and core borrowing, because the introduction of new things/concepts occasionally leads to the adoption of loan translations that may later function as “native equivalents” of their homophonic translations (recall the example of *shashinki* vs. *kamera* (< *camera*)).

2.2. Previous approaches to loanwords

This section outlines three recent approaches to loanwords: (i) quantitative lexicology, (ii) corpus linguistics and (iii) sociolinguistics.

2.2.1. Quantitative lexicology

A vocabulary survey is the most basic way of grasping the proportion of the LW stratum out of the whole vocabulary of Japanese. Large-scale vocabulary surveys started in Japan after WWII, initiated by NINJAL. These surveys aim to capture and describe the whole picture of the Japanese lexicon. It is usually done by means of counting the number of word entries (type) and the frequency of each entry (token) in a corpus. The previous surveys by NINJAL include investigations of texts from magazines, newspapers, textbooks, and TV broadcasting.

Although most of them are one-off surveys, there are a few longitudinal observations. One of the oldest surveys on a corpus of magazines conducted between 1956 and 1957 by NINJAL (NINJAL 1962, 1963, 1964) was followed up approximately 40 years later in 1994 (NINJAL 2005b), confirming that the proportion of the LW stratum as opposed to that of SJ, NJ and HB strata is on the increase. NINJAL (1987) also reported a gradual increase of the LW stratum in the magazine *Chuuookoron* between 1906 and 1976. Several follow-up surveys of the same magazine confirmed that there was a further increase in the proportion of loanwords in the following decades: Ishii (1990) for the year 1986; Irie (2010) for the years 1996 and 2006. Table 2.4 summarizes a diachronic distribution of each lexical stratum by type

and token over a century from 1906 to 2006. It shows a relatively consistent increase of the LW stratum in the Japanese vocabulary over the last century.

Table 2.4 Diachronic distribution of lexical strata in *Chuuookoron* (magazine) (sources, NINJAL 1987: 54, Ishii 1990: 4 and Irie 2010: 13)

	1906	1916	1926	1936	1946	1956	1966	1976	1986	1996	2006
Token (%)											
NJ	58.0	60.5	60.7	57.7	53.5	57.5	53.9	54.8	51.8	53.0	50.4
SJ	32.6	30.1	29.1	31.4	34.5	30.9	34.1	30.9	33.8	32.4	33.3
LW	0.9	0.9	1.1	1.5	1.5	2.4	2.4	3.7	3.5	4.3	4.0
HB	8.5	8.5	9.1	9.4	10.5	9.2	9.6	10.6	10.9	10.3	12.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Type (%)											
NJ	35.6	36.3	37.7	34.2	29.7	32.8	28.7	30.8	27.0	28.9	27.8
SJ	48.8	47.4	45.1	47.2	50.6	47.0	50.4	45.4	48.7	46.5	46.0
LW	1.6	1.6	1.9	2.3	2.5	3.7	4.1	5.7	5.9	7.0	6.1
HB	14.0	14.8	15.3	16.4	17.1	16.5	16.8	18.1	18.5	17.6	20.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Hashimoto (2010: Ch10) and Kim (2011: 14) also reported an increase in the use of loanwords in newspapers during the 20th century, especially in the latter half of the century. A report in Hashimoto includes the growing use of loanwords in addresses in the Minutes of the Diet after WWII. Although the scope is limited to a relatively formal register of spoken Japanese, this is one of the few recent investigations of loanwords in a spoken corpus.

The contribution of quantitative lexicology is that it provides empirical evidence for a diachronic increase of the LW stratum in the lexicon of Japanese, both by type and by token. In other words, it has been confirmed that there is a language change towards more use of loanwords as opposed to NJ or SJ words. Several loanwords have consequently obtained status as basic vocabulary in today's Japanese. During the latter

half of the 20th century more and more loanwords have become “basic words” (Kim 2011) that are used frequently and broadly across various registers or genres.

Another contribution of quantitative lexicology is that it describes the manner in which loanwords spread over time. Hashimoto (2010: 247) points out, as shown in Figure 2.4, that the diffusion of loanwords in the 20th century follows an s-shaped curve (“slow-quick-quick-slow pattern”) (Aitchison 2001: 91) with a gradual acceleration of increase at the start, the most rapid increase in the middle, and a gradual deceleration at the end.

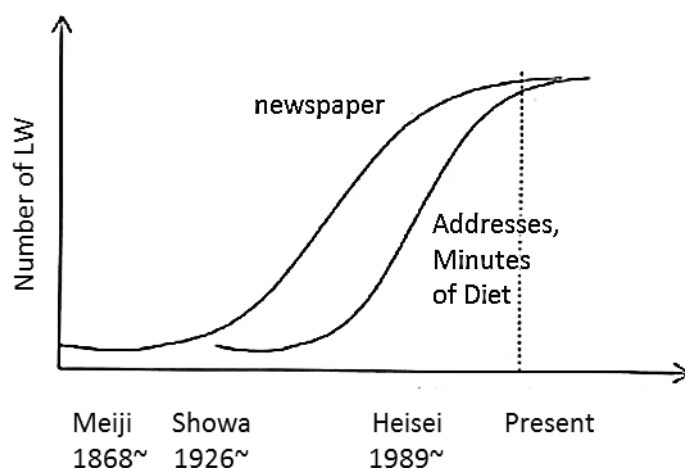


Figure 2.4 S-curve change in influx of loanwords (Hashimoto 2010: 244, Fig.10.11)

The previous studies provide a general picture of the diffusion of loanwords in Japanese, in particular to what degree and in what manner its diffusion has continued over time. In this macro perspective, however, little attention is paid on how individual loanwords have increased their use. Hashimoto states in her future prospectus that an individual loanword should diffuse in an s-shaped curve, as opposed to its native equivalents. She also hypothesizes that the above s-curve pattern should be present

when one looks at changes in a shorter period, since the s-curve pattern she observed is expected to be an accumulation of a series of “overlapping s-curves” (Aitchson 2001: 93). Testing her hypotheses is one of the main subjects in this thesis (Chapters 4, 7 and 8).

Moreover, quantitative lexicology does not distinguish loanwords that do not have native equivalents and those that coexist with their native equivalents, even though the mechanism of diffusion is quite different between the two groups. In the former case, as discussed above, borrowing occurs for a language-internal reason. In the latter group, in contrast, borrowing is motivated by language-external factors too, in which there should be various competitions – not only semantic, but also social and stylistic – among the loanwords and the existing native vocabulary. These aspects are captured only when we conduct a study from a micro perspective.

Another limitation lies in the fact that these surveys are mostly based on written language. Spoken corpus has not been well examined except the study of the language in TV broadcasting in 1989 conducted by NINJAL (1995, 1997, 1999) and political speech by Hashimoto (2010: Ch10).

2.2.2. Corpus linguistics

Corpus linguistics views loanwords from a different perspective, focusing on the use of individual loanwords in context. Their main goal is to provide a detailed description of the meaning and usage of individual loanwords in an empirical manner for the purpose of lexicography, language education, and so on. Such an approach has been attempted by, among others, Mogi (2011, *katto-suru* ‘to cut’; 2015, *maaku-suru* ‘to mark’) and Kim (2013, *chekku-suru* ‘to check’). Semantics of loanwords is sometimes contrasted

with existing native near-synonyms, e.g. Kim (2011: Ch4–6, *keesu* 'case' and *toraburu* 'trouble'); Miyata & Tanaka (2006, *risuku* 'risk'), Miyata (2007, *meritto* 'merit') and Shimooka (2013, *rirakkusu-suru* 'to relax').⁹ Among these studies Kim (2011) attempt diachronic investigations to describe the semantic development of loanwords in relation to their near-synonyms.

Their purpose is to reveal the “division of labor” in semantics between a particular loanword and its near-synonyms. They repeatedly reported that there is always difference in meaning or in collocational patterns between loanwords and their near-synonyms. In some cases of well-attested loanwords, their meaning or usage tends to be more varied than that of existing words. For example, the loanword *meritto* (< *merit* 'merit') can take on more collocational patterns than its equivalent SJ word *riten* (Miyata 2007). Kim (2011) also pointed out that the loan noun *toraburu* (< *trouble* 'trouble') plays an active function as general term/hyponym of its near-synonyms in newspapers. According to Miyata (2007: 408), loanwords generally have not been well established in their meanings or in their collocational pattern compared with existing native lexical items, and that is the reason why they are widely used or tolerant of various collocational patterns.

There is nevertheless a limitation to corpus linguistic approaches. Collocational pattern is only a “tendency” and not something that constrains the occurrence of a particular lexical item “categorically.” In other words, such linguistic factors cannot fully explain the reasons why a particular lexical item is chosen over the others that

⁹ Corpus linguistic approaches to loanwords were accelerated especially after the release of BCCWJ (Balanced Corpus of Contemporary Written Japanese). The studies that use BCCWJ include: Mogi 2011; 2015, Miyata & Tanaka (2006) and Miyata (2007).

share the same/similar meaning. To make a comprehensive explanation of word choice, linguistic external factors, such as social and stylistic ones, will have to be explored.

2.2.3. Sociolinguistics

The previous sociolinguistic studies of loanwords have mainly concerned the diffusion of loanwords in a social context. The arguments against the “flood of Western loanwords” that began to appear in the media in the 1980s (Ishiwata 2001: 27) have been drawing public attention to the use of loanwords in contemporary Japanese.

BBK has been conducting annual opinion polls on public attitudes towards language and language use since 1995, and loanwords have been one of the main topics in the surveys. NHK (1980, 1986, 1996) and NINJAL (2004, 2005a, 2007b) also conducted sampling surveys on loanwords for their own purposes. Some surveys were designed to capture the level of recognition, comprehension and use of selected loanwords (BBK 2003, 2008, 2009; NINJAL 2007b). In a more general context, people were also asked whether they have a positive attitude towards the present use of loanwords (NSDKK: *Naikaku Soori Daijin Kanboo Koohooshitsu*, Public Relations Office of Cabinet Secretariat 1977; BKK 2000, 2003, 2008, 2013) or a further increase of them (NHK 1991, BBK 1995, NINJAL 2004). There are some surveys and analyses from a variationist point of view, in which the focus was on the choice between selected loanwords and their native equivalents (NHK 1980, 1996; BBK 1999, 2000; NINJAL 2004, 2005a; Jinnouchi 1993, 2007, Tanaka 2007).

One of the important findings from the previous sociolinguistic studies is the existence of age differentiation in attitude towards loanwords, with a younger generation more tolerant of the use or increase of loanwords in general than elder one.

For example, in a recent survey conducted by BBK (2013), subjects were examined on their receptiveness to the fact that loanwords are in general use in their everyday life. The result shows a striking tendency for younger informants to be more open to the use of loanwords than the older ones (Table 2.5).¹⁰ Furthermore, it is reported that younger speakers show a higher level of exposure to, understanding of, familiarity with, and use of particular individual loanwords (BBK 1999, 2000; NINJAL 2007b). These surveys indicate that change in favor of loanwords is in progress in apparent time.

Table 2.5 Like and dislike of the present use of loanwords in general (NINJAL 2013: 92)

Year of Survey	Age range	No. of informants	Positive (%)			Negative (%)	
			Agree	Neither	Disagree		
2013	16-19	74	91.9	16.2	75.7	6.8	
2013	20-29	175	83.4	13.1	70.3	16.6	
2013	30-39	291	79.1	12.4	66.7	20.6	
2013	40-49	327	68.8	8.6	60.2	31.2	
2013	50-59	323	62.5	9.6	52.9	36.8	
2013	60-69	446	53.6	7.0	46.6	45.1	
2013	70+	517	48.9	7.5	41.4	47.2	

It has also been reported that there appears to be sexual, educational and occupational differentiation, although these tendencies are not as coherent across surveys as age differentiation. Loveday (1996: Ch6) points out that sex does not appear to be a significant variable on the social reception of contemporary contact with English. Yet on the level of the use of individual loanwords, some loanwords are preferred by female speakers, whereas some are preferred by male speakers (BBK 1999, 2000). With

¹⁰ Note that those choosing “Agree” and those choosing “Neither” are grouped together as those showing a “positive” attitude to the use of loanwords, since these two groups are similar in the age differentiation pattern (see also Ishiwata 2001: 122).

respect to educational/occupational factors, there seems to be a bit clearer tendency: those with higher occupational/educational backgrounds have a more positive attitude towards the use of loanwords in general (NINJAL 2004: 14, 31), or show a stronger tendency to accept/adopt English-based innovations (Loveday 1996: 187, see also Kuya 2013). Ishiwata (1965) and Loveday (1996: 181-2) also point out that more educated people in general recognize loanwords at higher rates.

There are also reports on stylistic differentiation, in which a loanword is selectively chosen from among near-synonyms depending on to whom one is talking or in what setting the utterance is taking place (NINJAL 2004: 32, NINJAL 2005a: 18-33, Jinnouchi 1993: 6). Such stylistic differentiation may be partly induced by the particular social connotations (functions) of that loanword. On the one hand, major advantages of using loanwords include, as seen in Figure 2.2, sounding novel or adding a stylish/fashionable nuance to one's speech. (NINJAL 2004: 17). On the other hand, people are concerned that use of, particularly a heavy use of, loanwords may increase the risk of miscommunication, the devolution of traditional Japanese language, the chance of sounding snobby, and so on. (NINJAL 2004: 20).

The earlier sociolinguistic studies have confirmed that there are social differentiations in the adoption and use of loanwords, according to age, sex, and education as well as style. These are important findings for the variationist study of loanwords, but several points are still overlooked.

First, the effect of linguistic external factors is discussed univariately in the past literature. As linguistic external factors are often interrelated with each other (Labov

1982: 52), a multivariate analysis will be essential in considering the interaction between them.

Second, the previous sociolinguistic approaches to the use of loanwords are mostly based on questionnaire-based interviews. The methodological advantage of a face-to-face interview is that researchers are able to elicit data regarding a targeted phenomenon in a relatively short time. This technique, however, always involves the problem of the “observer’s paradox”¹¹ and may overlook the distribution of language in actual use bound by various social contexts. Further steps will have to be taken to inquire regarding the relation between people’s linguistic attitude and their actual behavior.

Third, no sufficient attempt has ever been made to analyze the phenomenon in real time: almost all of the above-mentioned studies rely on apparent-time approaches. Although an age-graded distribution of an individual loanword is considered an “indication” of change in progress, it does not confirm the existence of language change in real time. Of course, considering the diachronic increase of the whole loanword stratum in corpora (NINJAL 1987, Hashimoto 2010, Kim 2011), it may not be unreasonable to interpret age gradient as a generational change in favor of loanwords (i.e. not age-grading), although no attempt has been made to confirm this about individual loanwords. Moreover, the assumption that one’s linguistic system remains stable over a lifetime after the critical period of language acquisition in one’s early childhood has proved to not always be true (Sankoff, G. 2005). It is noted that the

¹¹ Labov (1972: 209) states that “the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation.”

variability of speakers over their lifetime poses a problem in making a prediction of a given change, since the rate of change is not faithful to what the apparent-time hypothesis predicts. In order to provide a more accurate prediction of the linguistic phenomenon in question, a real-time approach will have to be used where possible.

2.3. Framework for the present research

The discussions above reveal several limitations of the earlier approaches to loanwords in Japanese. The main problem is that a sociolinguistic approach to loanwords has not been well combined with the methodological advantages of quantitative lexicology or corpus linguistics.

First, statistical technique has not yet been employed in the sociolinguistic study of loanwords in earnest. We have seen that a quantitative linguistic approach such as Hashimoto (2010) succeeded in constructing a mathematical model of the loanword diffusion in the Japanese lexicon, demonstrating that an s-curve model is a powerful tool in predicting the phenomenon. In the earlier sociolinguistic approaches, however, the focus has been put on “describing” the phenomenon by means of cross tabulation and little effort has been made to “predict” changes based on a model. Now a sociolinguistic investigation will have to make use of such a statistical model in order to predict the diffusion of loanwords from multiple factors.

Second, sociolinguistic studies of loanwords in natural/spontaneous speech have not yet been well investigated. As shown in the previous section, within a context of lexical variation, questionnaires have been the main means of collecting empirical data. Jinnouchi (1993) and Tanaka (2007) demonstrate that informants’ readiness to employ

loanwords as opposed to their native equivalents systematically increases as informants become younger (although statistical significance was not provided in either case). This hypothesis, however, has not been tested with respect to actual linguistic behavior. One of the reasons why sociolinguists have put little effort into the corpus-based approach seems to be the absence of corpora suitable for the study of lexical variation, in terms of both size and quality. Investigation into an individual set of open-class lexical variants such as, *keesu*, *jirei*, *baai* and *rei* ('case') requires a large collection of spontaneous utterance, because the probability for particular content words to occur is apparently lower than that for functional/grammatical ones (e.g. the nominal particle *ga/no*). To overcome such a difficulty, two large-scale corpora, written and spoken, will be investigated in this project.

Last, but not least, there is the important issue of how the change in question will progress in real time. Although a real-time approach has been often taken in quantitative lexicology (e.g. NINJAL 1987) and corpus linguistics (e.g. Hashimoto 2010, Kim 2011), sociolinguistic studies have often failed to capture changes in real time. In order to predict changes more accurately, it is important to examine the variability of individual speakers over time.

To tackle these problems, it is essential to use an integrated approach, i.e., to introduce advanced techniques employed in quantitative and corpus linguistics to a sociolinguistic approach to loanwords. That is, empirical data that represents people's attitudes based on interviews as well as behavior based on corpora should be analyzed in both apparent- and real-time contexts with quantitative analytical methodologies.

Further details about the methodology of the current research will be described in the following chapter.

3. Methods

3.1. Data

The successful account of a given variation requires a collection of data that is good for the purpose of both quantitative and qualitative analyses. In general, the fieldwork methods of observing language in the community are divided into two types: (i) face-to-face interviews and (ii) participant observations. The former technique, originally developed in Labov's study of phonological variation in New York (1966), is "means of obtaining the volume and quality of recorded speech that is needed for quantitative analysis" (Labov 1984: 29). He, however, draws our attention to the problem of the "observer's paradox" that sociolinguistic interviews are always faced with. The latter technique, meanwhile, attempts to minimize an observer effect in the process of data collection. Still, Labov points out that data collected with this technique is limited in both quality and quantity, concluding that it is important to combine these complementary approaches to reduce the limitations of each (ibid: 30).

Consequently both methods are used complementarily throughout the thesis: (i) interview-based and (ii) corpus-based observation. The former concerns the volume of data required for quantitative analysis, and the latter puts emphasis on the quality of data, i.e., observations with a reduced observer's effect.

3.1.1. Interview-based data

Interview-based data is gathered from several previous nationwide opinion surveys on language and language use in Japan conducted by various institutes including BBK, NINJAL and NHK between 1999 and 2005. Approximately 2,000 – 3,000 people aged

15 – 20 and above were randomly selected from the population of Japan in each survey. The interviewer(s) asked prepared questions to the selected informants in a face-to-face manner. The present thesis focuses on questions that are related to people’s linguistic attitudes towards loanwords. As shown in Table 3.1, questions concern the public attitude to (i) the use of specific loanwords as opposed to their native equivalents (Chapter 4), (ii) the alternation between a loan and non-loan variants across different stylistic settings/registers (Chapter 7) and (iii) the present use or a further expansion of loanwords in contemporary Japanese (Chapter 8). Some questions have been repeatedly asked with intervals of 5 to 20 years. For example, the third question in Table 3.1 has been asked in surveys conducted by NSDKK (1977) and BBK (2000, 2003, 2008 and 2013). Such data is studied diachronically to capture a real-time construct of change in the adoption of loanwords (Chapter 8).

Table 3.1 Examples of questions used in the interview-based studies

Question	Choices
(i) Among a loanword and its native equivalent, which lexical item do you usually use? (BBK 1999: 174-177)	(a) メリット <i>meritto</i> (< <i>merit</i>) (b) 利点 <i>riten</i>
(ii) Among the following loanword and its native equivalents, which lexical item do you use when: a) talking to friends? b) talking in public? c) talking to elderly people? (NINJAL 2005a: 19)	(a) サポート <i>sapooto</i> (< <i>support</i>) (b) 支援 <i>shien</i> (c) 手助け <i>tedasuke</i>
(iii) What do you feel about the present use of loanwords in our daily lives? (BBK 2013: 92)	(a) Agree (b) Somewhat agree (c) Somewhat disagree (d) Disagree

3.1.2. Corpus-based data: BCCWJ

The Balanced Corpus of Contemporary Written Japanese (BCCWJ) is a written corpus of a hundred million words compiled and released by NINJAL in 2011 (NINJAL 2011, Maekawa et al. 2014). It consists of three sub-corpora, called (i) the Publication sub-corpus (35 million words), (ii) the Library sub-corpus (30 million words) and (iii) the Special-purpose sub-corpus (35 million words).

Data in the first two sub-corpora was compiled with a random sampling technique. The population of the Publication sub-corpus is complete written texts from books (PB), magazines (PM) and newspapers (PN) that were published in Japan between 2001 and 2005, so the selected sample is a representative of written texts “published (or produced)” in these media in this specific period of time.¹² The population of the Library sub-corpus, in contrast, is complete written texts from books (LB) that satisfy the following two criteria: (i) they were published during the period 1986 – 2005 and (ii) were stocked in the collections of more than 13 metropolitan libraries in Tokyo as of 2007. The selected sample is therefore considered to be a representative of written texts that have good “exposure” with the public, i.e., relatively popular texts. The Special-purpose sub-corpus, although it was not based on random sampling, consists of corpora of a wide range of written texts, including governmental text from White Papers and Minutes of the Diet, textbooks, online text from *Yahoo! Answers* and *Yahoo! Blog*, and so on.

The first two sub-corpora, the Publication and Library sub-corpora, are selected for the purpose of the present sociolinguistic investigation. First, both of them are good

¹² The population of each sub-corpus is calculated in terms of the number of characters (not words) contained in the publications in question.

representatives of a given population of written language. Second, they are large enough to collect tokens of particular lexical items. Third, roughly half of the samples contain sociolinguistically useful information on the author, including year of birth, sex and educational background.¹³ Such information is not usually available in other corpora of written Japanese, including the Special-purpose sub-corpus, which does not include such information about the authors.

These two sub-corpora are used differently in the current study: the Publication sub-corpus is used as a synchronic corpus to observe language change in apparent time at a certain point of time: the beginning of the 21st century (2001-2005) (Chapters 5 and 6), while the Library sub-corpus is used as a diachronic corpus to observe language change in real time over a certain period of time: the 20 years from 1986 to 2005 (Chapter 8).

In these sub-corpora, different types of text, fixed-length and variable-length, are elicited. In the fixed-length data, the length of a sample is fixed at 1,000 characters, which corresponds to approximately 590 words (NINJAL 2011: 17), starting at a set datum point. In the variable-length data, although containing a starting datum point just as a fixed-length sample does, the length of a sample varies (up to a maximum of 10,000 characters) because a sample is extracted as a complete series of sentences or paragraphs on a certain topic.

The current research uses both types of data depending on the purpose of the investigation. Since every sample is controlled for length, the fixed-length data is more

¹³ Educational history of the author is not included in the official corpus DVD and therefore has to be identified by data from the questionnaire survey of authors conducted separately by NINJAL between 2006 and 2010. I obtained special permission from NINJAL to use this additional information.

suitable for a more rigorous quantitative analysis, e.g., making a frequency list to elicit basic vocabulary (NINJAL 2011:17, 23) (see 5.3.2). Meanwhile, the variable-length data is more suitable for a qualitative analysis, e.g., elucidating structure, discursive context and stylistics of the text. Since the size of variable-length data (32.2 million words) is four times as large as that of the fixed-length data (8.5 million words), it is more useful for an in-depth study of individual lexical variables, which are relatively rare in frequency compared with phonological/syntactic variables (see 5.4, 5.5 and Chapter 6).

3.1.3. Corpus-based data: CSJ

The Corpus of Spontaneous Japanese (CSJ) is a collection of spontaneous speech that was released in 2004 by NINJAL and NICT (the National Institute of Information and Communications Technology) (NINJAL 2006). It consists of recordings of approximately 660 hours, 7.5 million words, from 1,417 speakers under NINJAL and NICT's project conducted between 1999 and 2003.

The majority (approximately 90 %) of the recorded speech belongs to the category of monologue speech, which consists of Academic Presentation Speech (APS) and Simulated Public Speech (SPS). The former is a series of recordings of academic presentations given in various fields. The latter includes recordings of public talks given by selected ordinary speakers (but not limited exclusively to non-academics). Speakers are asked to give a 10 – 12 minutes-long talk to a small audience on everyday topics such as “the most joyful moment of my life,” “the hardest moment of my life,” “the town where I live,” and so on.

In order to maximize a stylistic contrast between APS and SPS, the recording staff members tried to make the speech setting as relaxed as possible during the recording of SPS by having a chat with the informant before recording. They also kept trying to show a certain reaction (e.g. nodding) proactively while recording was taking place. The two types of speech are thus often contrasted in terms of formality (or what they call “spontaneity”). That is, SPS has more “relaxed” or more informal speech style than APS (NINJAL 2006: 11). Of course, a register difference may not be a direct indicator of formality. Samples in the same register may exhibit various degrees of formality in speech. On such an assumption, samples in APS and SPS are rated individually in terms of various “impressionistic” aspects of speech style, including “Spontaneity,” “Speaking Rate,” “Clearness of Speech Articulation,” “Formality,” and so on. Such an evaluation (what they call “impressionistic rating”) of these features was done by one of the recording staff members according to a five-point scale. In addition to register distinction, the rating of “Formality” and “Spontaneity” will serve as useful factors in the study of linguistic variation (Maekawa 2003).

The use of CSJ enables us to explore spontaneous spoken language. Although it is smaller in sample size than BCCWJ, CSJ is, for the time being, largest of released spoken corpora in Japan, with rich annotation for the speaker’s sociolinguistic information.

3.1.4. Complementary use of different sources of data

As mentioned above, any data collection method has its strengths and weaknesses. It is thus important to use different sources of data complementarily. Table 3.2 compares advantages and disadvantages of using the above-mentioned three data sources.

Table 3.2 Strengths (+) and weaknesses (-) of each source

	Interviews	Corpora	
		BCCWJ	CSJ
Sample size	+(large)	+(large)	– (relatively small)
Representativeness	+	+	–
Annotation for social attributes	+	not all authors	+
Annotation for style/register	+(limited items)	register distinction (PB vs. PM vs. PN in the Publication sub-corpus)	register distinction (APS vs. SPS) & rating for stylistic factors
Spontaneity (observer's effect)	– (more)	+(less)	+(less)
Real-time approach	+(limited items)	+(the Library sub-corpus)	–

The advantage of using the interview-based data is, first of all, that it makes a *large-scale* sample immediately available for quantitative analysis. Second, a random sampling technique enables us to elicit a sample that is a good *representative* of the speakers of Japanese in Japan. *Social attributes* of speakers, e.g., age, sex, or education, can therefore expect to be well balanced. Third, questions like (ii) in Table 3.1 encourage us to investigate *stylistic* constraints on the use of loanwords (Chapter 7). Fourth, a *real-time* approach is also possible for limited items (Chapter 8).

Nevertheless there are limitations in using the interview-based data. The major limitation is that people's self-reports do not necessarily reveal their actual behaviors, because an interview method could maximize the *observer's effect* on informants' responses. In an interview situation where informants are formally asked about their linguistic behavior, they are likely to become more conscious about the social value of given variants than usual. It should be kept in mind that the questions used here take the

form of a direct question, where the interviewer asks informants directly about their lexical choices among prepared sets of lexical items as shown in Table 3.1. Therefore, the results of interview-based study will have to be re-examined by studying linguistic behavior with corpora.

The interview method, on the other hand, is an effective way of eliciting large-scale tokens of targeted lexical variables in a limited time. One can gather data in a more indirect way, employing a traditional dialectological technique in which an interviewer asks informants the name of a certain item by showing a picture. However, this technique may not be so suitable for eliciting the use of lexical items referring to abstract concepts. Kim (2011) points out that the majority of loanwords that show diachronic increases in their occurrence over the latter half of the 20th century are those that belong to abstract nouns such as *keesu* (< *case*), *toraburu* (< *trouble*), and so on. It would then be difficult for us to lead informants immediately to utter our targeted lexical items through this technique.

The strength of BCCWJ is, first, that it provides a *large* and sufficient amount of tokens of individual lexical variables to pursue my research quantitatively. Second, a sample of each sub-corpus is a good *representative* of a targeted set of written text, which is why this corpus is called a “balanced” one. It is, however, noted that *social attributes* such as age, sex and education are not annotated for all the authors because such information may not always be traceable. Moreover, the corpus is not well-balanced in terms of social attributes of the authors. Male, well-educated or middle-/old-aged authors are predominant in this corpus. Third, it would be useful to make a contrast between different *registers* such as books (PB) vs. magazines (PM). It

also needs to be noted, however, that although this corpus provides register distinction among books, magazines, and so on, this does not explain fully the mechanism of stylistic difference. Fourth, as BCCWJ is a collection of *spontaneous* writings, it is expected that the *observer's effect* in this corpus is reduced compared to formal face-to-face interviews, although writing convention still restricts the author's use of language in a different way. It is still important that the corpus of this scale, with information on social attributes of authors, is available for the sociolinguistic study of infrequent variables like lexical items. Lastly, its sub-corpus, i.e., the Library sub-corpus, enables us to observe the phenomenon in *real time* over 20 years between 1986 and 2005.

With regard to CSJ, in contrast to the other two data sources, its *sample size* may not be enough for us to analyze individual lexical items quantitatively. Second, the corpus is not *representative* of a particular population as random sampling was not applied in the collection of the speakers. Speaker attributes are, however, more balanced in terms of age and sex in SPS, particularly when compared with BCCWJ (although young male speakers are predominant in APS). Third, a *real-time* approach is not possible.

Despite these weaknesses, CSJ is more useful than the rest of the two sources in several aspects. First, the corpus is rich in annotation for *social attributes* of the speakers, e.g., age, sex, education, and so on. Second, the contrast between the two registers (APS vs. SPS) as well as indicators of style such as formality and spontaneity will contribute to elicitation of *stylistic* effects on the phenomenon. Third, as the corpus consists of recordings of *spontaneous* speech, the *observer's effect* is expected to be weaker than in interviews. It is nevertheless noted that the relatively formal nature of

the two registers of public speaking would restricts the speaker's manner of speech in some way.

Taking all these advantages and limitations into account, these data sources are used complementarily in the following analyses. Each data source is worth examining to elucidate different aspects of distribution of loanwords, and it is expected that the synthetic use of these sources will lead us to gain a more comprehensive understanding of the linguistic phenomenon in question.

3.2. Analytical method

3.2.1. S-curve model of language change

It is said that language change follows an s-curve pattern (Weinreich, Labov & Herzog 1968: 113, Chambers & Trudgill 1980: 177-180), or a “slow-quick-quick-slow” pattern (Aitchison 2001: 91), in which a given newer linguistic variant increases slowly at the beginning, has an acceleration in the middle and then a deceleration towards the end. Previous variationist studies have described and interpreted mainly phonological and syntactic variation with an s-curve model (Labov 1972, Sankoff and Larberge 1978, Hibiya 1988b, Matsuda 1993, Nambu 2007, Sano 2008, Sano & Kawahara 2013). Studies such as Chambers (1990) and ISM (the Institute of Statistical Mathematics) & NINJAL (2015) showed that lexical alternation also follows an s-shaped pattern.

The purposes of employing an s-curve model to the analysis of language change are twofold. First, in the context of variationist studies, the main purpose is to “identify” statistically significant variables that affect the language change in question. A logistic regression model (see the following section) is widely used for this. In the present

research, the s-curve model is used for this purpose, i.e., to recognize significant internal and external variables on the occurrence of loanwords (Chapters 5 and 6).

Second, quantitative linguistics uses this mathematical model to make a “prediction” of change. Examples of studies for this purpose include an estimation of duration of phonological change in Yamazoe, Japan (Inoue 2000); an estimation of the period at which the rate of growth of Meiji era “scholarly *kango*” became sluggish (Sanada 2002); a prediction of diachronic process of dialect standardization in Tsuruoka, Japan (Yokoyama & Sanada 2010); a prediction of diachronic process of the use of honorifics in Okazaki, Japan (Yokoyama et al. 2008), and so on.

Yokoyama & Sanada (2007) and Sanada (2008) put an emphasis on the usefulness of an s-curve model, not only for the first purpose but also for the second. In other words, they suggest that we make the best of an s-curve model in predicting language change, i.e., deductive (hypothesis-testing) studies, in contrast to inductive (descriptive) studies.

With regard to the diffusion of lexical items, Altman et al. (1983: 112) points out that the process of acceptance of words in real time is describable by an s-curve. Hashimoto (2010), in her study of the process of diachronic increase of the loanword stratum in newspapers in the 20th century, demonstrated that an s-curve model fit the data well. Based on these studies, the present research presupposes that loanwords, as opposed to their existing native equivalents, diffuse in an s-shaped manner in both apparent and real time. In fact, a number of previous surveys, including BBK (1999, 2000), NINJAL (2004, 2005a) and Tanaka (2007), show that there is an increase in the adoption of loanwords in apparent time. These studies, however, do not exceed a purely

descriptive level. This thesis hence attempts to employ an s-curve model to the existing data outlined in the previous section, so that the resultant statistical model can be used for the prediction of the change in question (Chapters 4, 7 and 8).

3.2.2. Multiple logistic regression model and independent variables

The multiple logistic regression model¹⁴ is a statistical model used to estimate the probability of a binary (i.e., categorical: 0 or 1) dependent variable based on more than one independent variable. The formula of the model is describable by [Formula 1]. [Formula 1] can be rearranged and yields [Formula 1'] to give a predicted probability p .

$$\log[p/(1-p)] = a_1x_1 + a_2x_2 + \dots + a_nx_n + b \quad \text{[Formula 1]}$$

$$p = 1/[1+\exp(-1)(a_1x_1 + a_2x_2 + \dots + a_nx_n + b)] \quad \text{[Formula 1']}$$

In the above formulae, p represents the probability of the occurrence of a given variant, while x_n is an independent variable, a_n is a parameter value of x_n and b is an intercept. It is important to notice that unlike ordinary linear regression, the output is between 0 and 1, following an s-shaped curve.

In this thesis the dependent variable is binary, i.e., the occurrence or non-occurrence of loanwords. p is the probability of the occurrence of a given loanword as opposed to that of its native equivalents. Independent variables x_n include (i) Year of Birth (Age), (ii) Education, (iii) Register/Style, (iv) Sex and (v) Year of Survey:

¹⁴ Variable rule program (Cedergren & Sankoff 1974) and SPSS are available for a multivariate logistic analysis. Examples of analysis of language change in Japanese using the former include Hibiya (1988b) and Matsuda (1993), and those using the latter include Nambu (2007) and Sano (2008, 2011).

p = Probability of the occurrence of a given loanword as opposed to its native equivalent(s)

x_1 = Year of Birth (Age)

x_2 = Education

x_3 = Register/Style

x_4 = Sex

x_5 = Year of Survey

Year of Birth (Age) is an essential factor for the study of variation and change in apparent time. The question of whether the distribution of a loanword is graded according to the age of the speakers will be examined throughout the thesis. The following chapter (Chapter 4) puts a primary focus on the structure of age differentiation in the adoption of loanwords. We will see that age differentiation does not show a monotonous gradient, but that it involves a delay in the acquisition process of the new variants among the youngest age group. This may sometimes complicate the description and prediction of process of change.

The effect of education as one of indicators of social stratification has proved to play an important role in language change in Japanese. For example, more educated people show a stronger orientation toward the preservation of older forms (i.e. Tsuruoka dialect) than less educated people (NINJAL 2007a: 159-160). It has been shown, however, that they show a higher degree of using newer forms (standardized variety of Japanese) in their attitudes when it comes to individual phonological, grammatical and lexical items (ibid: 117-157). An educational factor has been examined by a very limited number of researchers in the context of loanword studies. Ishiwata (1965) points out that more educated people exhibit higher levels of recognition of loanwords. The

question then arises as to whether more educated people uses loanwords more frequently. This query will be examined across several chapters (Chapters 5, 6 and 7), revealing that the effect of education is modified by register/style.

Register/Style, besides age and social stratification such as education, is expected to be another important factor that determines the sociolinguistic contexts of utterances. The question of whether or not loanwords are favored in formal registers or settings will be examined (Chapters 5, 6 and 7). It will also be shown that stylistic differentiation is more prominent among more educated people (Chapter 7). These observations are useful to evaluate the sociolinguistic status of loanwords as opposed to their equivalent existing native forms.

With regard to Sex, its effect on the adoption of loan variants is not consistent across words. This is because adoption is specific to the context of the words in question. The current research will show that those that are well attested in the domain of business tend to be favored by men, whereas those that are well attested in the domain of everyday life may be favored by women (Chapter 4). On the other hand, the sex effect will not prove to be significant on the adoption of the well-attested loanwords *keesu* (< *case*) and *sapooto* (< *support*) (Chapters 6 and 7). These results seem to show that sex is not a bona-fide variable per se.

Year of Survey is an important factor for the study of language change in real time. In particular, the combination of this factor and Year of Birth allows us to examine the variability of the community as a whole over time, and, at the same time, the variability of individuals over their lifespan. The examination of these variables will reveal that the

process of language change cannot be described as simply as the apparent-time hypothesis presupposes (Chapter 8).

3.3. Issues of variationist approach to lexical variation

Before looking at the distribution of any set of lexical items, we need to discuss the general problem of the study of lexical variation, i.e., the question of in what manner lexical items are treated within the framework of variationist study.

As Labov states, “the definition of a linguistic variable is the first and also the last step in the analysis of variation” (2004: 7). In its original sense, a linguistic variable is defined as “two or more (alternative) ways of saying the same thing” (Labov 1972, Weiner & Labov 1983: 31). In other words, an alternation between competing variants should not involve a change in meaning. For instance, the presence or absence of [r] in *fourth floor* does not affect their meaning per se. In this sense, phonological variables are considered to be most useful (Lavandera 1978: 176).

Variationists are nevertheless aware that “variable elements are found at all levels of linguistic structure” (Weiner & Labov 1983: 31). The question then arises as to whether or not the original notion of *semantic equivalence* developed in the study of phonological variation could be applied in a strict sense to the analysis of non-phonological variation. This is because, although phonological alternations are meaning-preserving, alternations of different forms at non-phonological level are not necessarily so. With regard to lexical variation, the application of the traditional notion of semantic equivalence might become even more problematic because “there are no true synonyms in an absolute sense” (Labov 1978: 8). It seems that the notion of

semantic equivalence potentially poses a problem in studying variables “above and beyond phonology” (Sankoff, G. 1973).

These discussions show that the notion of semantic equivalence is not applicable to the identification of non-phonological variables (Lavandera 1978), i.e., the definition of a linguistic variable needs revising. Weiner & Labov (1983: 31), in their analysis of variation between active and passive forms, points out that, “there is no reason to confine the study of variation to alternative ways of saying the same thing, although this kind of variable has been the chief focus in the development of variable rules.” Labov also admits the existence of stylistic demands that force substitution of one word for another as a stylistic variant (1978: 8). Weiner & Labov (1983: 30) states:

But it seems to us that this concession to an idealistic semantics is needlessly unrealistic. If we isolate words from their use, we can show that there is no such thing as a precise synonym, since all words have slightly different privileges of occurrence when we consider every possible context. But in practice, the need for stylistic variation leads all speakers and writers of English to substitute one word for the other with the expectation that any differences that might arise in other contexts will not affect interpretation in that one.

Lavandera (1978: 181) and Sankoff, D. (1982) suggest that, for non-phonological variables, the condition of semantic equivalence be “relaxed” and that *functional equivalence* or *discourse equivalence* be employed instead. In the study of Montreal French by Sankoff & Thibault (1980: 312), they identify environment where *avoir* (‘to have’) and *être* (‘to be’) share the same function as auxiliaries by eliminating contexts where they have other functions, e.g. the use of *être* as a copula.

These studies suggest that non-phonological variables are worth examining, although we should note that “we obviously have a much better chance of getting intersubjective agreement in identifying formal variants than semantic variants,” or,

“the possibility of accurate measurement is less immediate with semantic variation” (Weiner & Labov 1983: 31). If we accept replacing the original notion of semantic equivalence with the notion of functional or discourse equivalence, it would seem that the analysis of lexical variation is not unreasonable.

With the above discussion, I am attempting to apply the notion of functional equivalence to the analysis of lexical variation among the loanword *keesu* (< *case* as in *there is a case where...*) and its native equivalents *jirei*, *baai* and *rei* (Chapter 6). I will identify carefully which context should be counted as variable environments and which should not be, by means of having a closer look at the functions of the proposed lexical variants depending on context. It is noted that most lexical items are polysemous and have their own unique functions and any irrelevant contexts are excluded from the analysis. For example, the loanword *keesu* can refer to ‘a container’ as a concrete noun as opposed to its (principal) usage as a formal noun, which is shared by the four lexical variants. Furthermore, it is reported that the native form *baai* can be used to form a hypothetical subordinate clause to mean ‘if/when’ whereas the loanword *keesu* and the rest of the native forms cannot be used in this function (Kim 2011). Categorical (or non-overlapped) environments like these should not fall within the scope of variationist analysis (Lavandera 1978: 178). Note that here I only focus on function, not on meaning, of a word, i.e., I disregard difference in nuance or tinge of each expression that might be recognized differently by “idiosyncratic interpretation” (Tagliamonte 2006: 74).¹⁵

¹⁵ Tagliamonte (2006: 73-4) points out that there are two different levels of meaning. In regards to meaning, there is, on the one hand, “comprehensive meaning, which takes into consideration every possible inference,” and which is “subject to idiosyncratic interpretation and an infinite range of potential meanings.” On the other hand, there is “meaning as it is used in the speech community,” which is “by definition a consensus that is shared and relatively constant.” It should be noted that meaning in the latter sense should represent a narrower interpretation.

The above procedure, however, is a painstaking task, since the analyst is required to identify all usages, or functions, of sets of given lexical variants by examining all tokens that occur in a corpus. Unfortunately, in-depth investigations of individual loanwords have not been sufficient (Mogi 2011). The primary limitation of this approach then is that the analyst cannot observe enough loanwords at a time as to draw a general conclusion about the distribution of loanwords.

To overcome this problem, Chapter 5 will undertake a more comprehensive approach. This approach takes into account all the environments in which potential lexical variants occur. The limitation of this approach is, of course, that irrelevant contexts cannot be completely excluded from the analysis. Nevertheless, as shown later, we will find the general tendency for loan variants to replace their native equivalents in apparent time. Moreover, we will also find that loan variants are distributed differently along social dimensions such as education and style. The result shows that there is “weak complementarity of distribution,” in which competing forms are distributed differently across different segments of the speech community (Sankoff, D. 1982: 683-4). This can usually be considered supporting evidence that different forms serve an equivalent function in discourse for different members of the community (ibid: 684).

It seems that the notion of weak complementarity puts more emphasis on a resultant distribution, or the “distributional facts” in Tagliamonte’s terminology (2006: 74), of the forms rather than the “definition” of (semantic) equivalence among variants.¹⁶ Sankoff, D. (1982: 683) says, “what is important here is that whatever distinctions the

¹⁶ See also D’Arcy (2006) and Schwenter & Torres Cacoullos (2010) for the application of this notion.

analyst may wish to draw between two forms showing weak complementarity, these distinctions can have little consequence for the speakers, for the communicative use of language within the community”.

In summary, this thesis employs two approaches in identifying the scope of investigation: (i) contextual approach and (ii) comprehensive approach. The first approach, i.e., contextual approach, counts only the restrictive area in which proposed synonyms overlap in function. This scope is conceptually represented by the shaded areas of overlap in Figure 3.1. This approach is introduced in a corpus-based case study of the loanword *keesu* (< *case*) in Chapter 6.

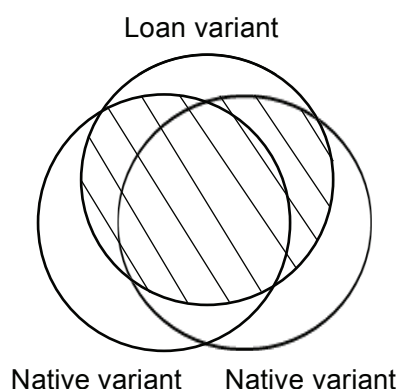


Figure 3.1 Scope of comprehensive and contextual approaches

The second approach, i.e., comprehensive approach, takes into account all the environments that proposed synonyms occur (Chapters 5 and 8). The scope corresponds to both the shaded and unshaded areas in Figure 3.1. This approach may be criticized for not excluding “noise” in the data, but it enables us to observe a larger amount of linguistic variables, with the result that the analyst can avoid the risk of drawing a general conclusion from a too small pool of data. In fact, as will be shown in Chapter 5,

this approach nevertheless succeeds to some extent in eliciting complementary distributions of several lexical variables across generation or other social dimensions, indicating that the proposed variants are in competition in the community.

Both approaches are expected to contribute to investigation of the phenomenon. The contextual approach gives a more fine-grained analysis with a small amount of data on the one hand, whereas the comprehensive approach gives an analysis of a larger amount of data that is useful for generalization of the phenomenon; however, it is less in-depth than the former approach.

The above discussion of circumscribing the variable context is regarded as a matter relevant only to the study of natural speech with a discursive context (Chapters 5 and 6). Interview-based studies (Chapters 4 and 7) primarily concern the speaker's choice between conceptually synonymous forms A and B, with no discursive context attached. In the latter cases, we are only looking at the status of a given lexical item in one's mental lexicon, and thus its distribution in natural speech is disregarded in principle.

4. Diffusion of loanwords in apparent time

4.1. Introduction

4.1.1. Aim: Predicting the diffusion of loanwords based on the apparent-time hypothesis

The purpose of this chapter is to describe and predict the diffusion of loanwords as opposed to that of their native equivalents based on their distribution in apparent time.¹⁷

Many loanwords have been increasingly used despite the fact that their native equivalents already exist in Japanese, replacing their existing words over time. Here, loanwords are considered to be new lexical variants in rivalry with existing native lexical items. The investigation of loanword diffusion is thus important to the study of language variation and change in Japanese.

This chapter, using 17 loanwords as examples, examines how these loanwords spread over generations. Based on several nationwide opinion polls conducted between 1999 and 2004, we will see that synchronic distributions of loanwords verify the presence of change in apparent time, i.e., a change toward more adoption of loanwords. I will also show that their distributions are predictable with an s-curve model.

4.1.2. Apparent-time approach to language change

The apparent-time approach aims to predict the existence or direction of change from age differentiation as observed in a single synchronic study. Figure 4.1 shows that lexical variation in apparent time in Canadian English (Chambers 1990). The figure

¹⁷ Chapter 4 is a revised version of the paper published as: Kuya, A. (2016) Mikakejoo no jikan o riyoo shita gairai shiyoo ishiki no tsuuji henka yosoku. *Nihongo no Kenkyuu*, 12(4).

demonstrates that younger people tend to use the American variation *couch* as opposed to the Canadian variation *chesterfield*.

The age differentiation predicts two things here. First, age differentiation is an indication of a change in progress, i.e., it is possible to predict that *couch* will spread further, replacing *chesterfield* according to the alternation of generations. Second, the age gradient in apparent time can be considered the rate of change in real time. These predictions are based on the well-known hypothesis that one's linguistic system, acquired during the critical period of language acquisition, is invariable throughout one's lifetime (Apparent-time Hypothesis).

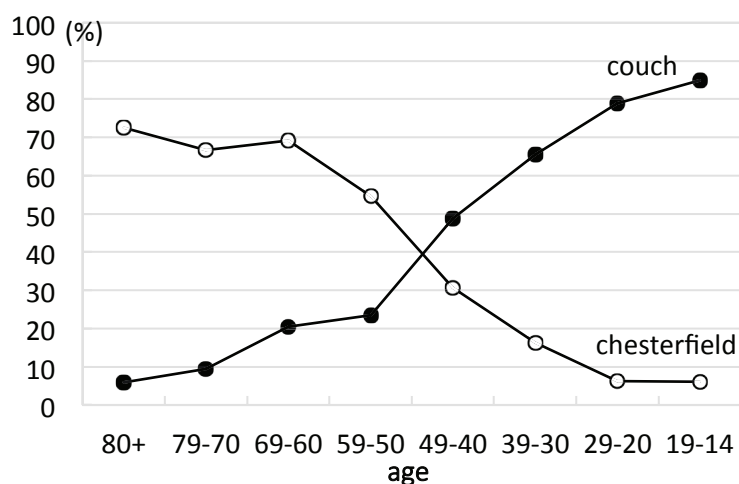


Figure 4.1 Lexical choice by the speaker's age in Canadian English (Chambers 1990: 162, Figure 1)

The apparent-time approach is a popular way of studying language variation and change (Labov 1963, 1972, Trudgill 1974, among others) since one can predict the existence and direction of change to a certain degree even when real-time data is not immediately available. Japanese sociolinguists, such as Shibata (1978), Matsuda (1993),

Hibiya (1995), Nambu (2007), Sano (2008), have also mainly used the apparent-time approach in their studies of variation and change.

There are, however, several limitations in this approach. First, age differentiation per se is not necessarily evidence of a change in progress: it could be the reflection of age-grading. Age-grading, or age-graded variation, occurs when one changes one's linguistic behavior at different stages in one's lifetime but it does not lead to change of the community as a whole (Labov 1994: 83). Therefore, in order to conclude that a given age differentiation is indeed change in progress, it is necessary to make certain that change occurs at the communal level in real time.

Second, a prediction based on the apparent-time hypothesis might lead to a miscalculation of the rate of change (Sankoff, G. 2006). It has been shown that the apparent-time hypothesis often fails to capture more complicated cases of language change, in which individual speakers change their linguistic behavior throughout their lifetimes either in the direction of a change in progress in the community (Sankoff, G. 2005) or in the opposite direction (Yokoyama & Sanada 2010). Yokoyama & Sanada pointed out that in such cases the apparent-time hypothesis might underestimate or overestimate the rate of change. (For a more detailed discussion see Chapter 8.)

The previous studies suggest that we should interpret age differentiation carefully when an apparent-time distribution is the only thing we can make use of.

4.2. Data

4.2.1. Opinion polls

We shall examine the diffusion of 17 loanwords based on self-reports collected from four nationwide opinion polls conducted around 2000. Table 4.1 gives a summary of the four surveys. Each survey targeted the whole Japanese-speaking population in Japan aged 15 or 16 and above. The samples are 3,000 to 4,500 people randomly selected from the population. The number of valid response ranges from 2,000 to 3,000. In a face-to-face interview, an informant was asked to choose from among a loanword and its native equivalent(s), one lexical item that s/he usually uses or feels more familiar with.

Table 4.1 Sources of data

Survey ID	Survey name (source)	Month/Year of survey	Targeted population	No. of valid response (%)
A	<i>Opinion Poll on the National Language: Heisei 10 (1998)</i> (BBK 1999: 174-177)	Jan/1999	those aged ≥ 16	2200 (73.3%)
B	<i>Opinion Poll on the National Language: Heisei 11(1999)</i> (BBK 2000: 18)	Jan/2000		2196 (73.2%)
C	<i>The National Survey on Attitudes to Loanwords</i> (NINJAL 2004: 38, 43)	Oct-Nov/2003	those aged ≥ 15	3087 (68.6%)
D	<i>The National Survey on Attitudes to Loanwords II</i> (NINJAL 2005a: 19) ¹⁸	Oct-Nov/2004		3090 (68.7%)

The proportion of those who selected “loanwords” as opposed to the rest of the choices, i.e., “native equivalents,” “use both,” “do not use either of them” and/or “not

¹⁸ Proportions by sex are not available in the report of this survey. The age-specific proportions shown in Appendix A were calculated by the author based on raw data provided by NINJAL.

sure,” is considered to be the level of diffusion of the loanword in question. The overall level of diffusion of each loanword is provided as “LW (%)” in Table 4.2. The distribution of each item across age and sex is provided in Appendix A.

Table 4.2 Selection of loanwords and their data source

ID	Loanwords (LW)			Native equivalents		LW (%)	Survey ID-Question No.
	lemma_JP	lemma	Etymology	lemma_JP	lemma		
1	ニーズ	<i>niizu</i>	needs	必要性	<i>hitsuyoosei</i>	36.4	A-Q21: Which of the two do you feel more familiar with (provided that they are used in the same sense)?
2	イベント	<i>ibento</i>	event	催し	<i>moyooshi</i>	58.2	
3	メリット	<i>meritto</i>	merit	利点	<i>riten</i>	57.9	
4	コンセンサス	<i>konsensasu</i>	consensus	合意	<i>gooi</i>	9.0	
5	スキーム	<i>sukiimu</i>	scheme	計画	<i>keikaku</i>	3.5	
6	リスク	<i>risuku</i>	risk	危険性	<i>kikensei</i>	35.5	
7	ビジョン	<i>bijon</i>	vision	展望	<i>tenboo</i>	30.7	
8	アカウント ビリティ	<i>akauntabiri ti</i>	accountability	説明責任	<i>setsumeisekinin</i>	3.6	B-Q4: Which of the two do you usually use (provided that they are used in the same sense)?
9	ワイン	<i>wain</i>	wine	ぶどう酒	<i>budooshu</i>	75.6	
10	スーツ	<i>suutsu</i>	suit	背広	<i>sebiro</i>	52.8	
11	キッチン	<i>kitchin</i>	kitchen	台所	<i>daidokoro</i>	12.1	
12	オープンする	<i>oopun-suru</i> ¹⁹	open	開店する	<i>kaiten-suru</i>	34.6	
13	ゲットする	<i>getto-suru</i>	get	獲得する	<i>kakutoku-suru</i>	14.0	
14	トータルで	<i>tootaru-de</i>	total	合計で	<i>gookei-de</i>	13.3	
15	キャンセル	<i>kyanseru</i>	cancel	解約 取り消し	<i>kaiyaku torikeshi</i>	57.5	C-Q12-13: Which word do you want to use most (provided that they are used in the same sense)?
16	ハッピー	<i>happii</i>	happy	幸福 幸せ	<i>koofuku shiawase</i>	17.7	
17	サポート	<i>sapooto</i>	support	支援 手助け	<i>shien tedasuke</i>	33.4	D-Q2: Which word do you use when talking to friend? (context provided ²⁰)

¹⁹ *-suru* ('do') is a light verb that is attached to a verbal noun to form a verb (see 5.3.1 for more detail).

²⁰ The provided context is the following:

*Atarashiku noogyoo-o hajimeru tame-ni-wa
chiiki-no []-ga hitsuyoo-desu.*

newly farming.business-ACC to.start in.order-to-TOP
community-GEN []-NOM necessary-COP.POL

'You need the [support] of community when starting a new farming business.'

4.2.2. Semantic difference between loan and native variants

We should pay attention to semantic difference between a loanword and its corresponding native word because such a difference may affect the distribution of paired lexical items as I discuss later in 4.4.3. I conducted keyword search with the written corpus BCCWJ for some of those listed in Table 4.2.²¹ There appear to be two types of semantic difference.

First, there is a case in which we can find a relatively clear division of labor between a loan variant and its native variant. For example, the loanword *getto-suru* (< to get) tends to take as its object a concrete noun that can be obtained relatively easily (e.g. clothing, a book, a car and so on) (4.1a); In contrast, the corresponding SJ word *kakutoku-suru* shows a stronger association with abstract things, that is, something that cannot be simply exchanged for money (e.g. right, status, prize, prize money, knowledge and so on) and gives the impression that one is required to exert much effort to obtain it (4-1b).

(4.1)

- a. 人気雑誌を確実に GET するには？ (PM21_00604)

Ninki zasshi-o kakujitsuni getto-suru ni-wa

Popular magazine-ACC certainly get-do in.order.to-TOP

‘How to get a popular magazine certainly/successfully’

- b. 平和と自由を獲得する (PB34_00305)

Heiwa to jiyuu-o kakutoku-suru

²¹ The search was limited to the following context in the Publication sub-corpus.

Nouns: (a word ends in hiragana writing + key word + particle/auxiliary)

Verbs: (a word ends in hiragana writing + key word + *suru* (‘do’)/*dekiru* (‘be able to’))

Peace and freedom-ACC get-do

‘Gain peace and freedom’

Second, there is a case in which a loan variant covers a more general area of meaning and thus is used more widely than its native existing variant. For example, the loanword *suutsu* (< *suit*) is used to refer to clothing for both sexes in the corpus (4.2a-b), while its native equivalent *sebiro* is usually used to refer to men’s clothing (4.2c). Another example shows that the loanword *oopun-suru* (< *to open*) takes a more diverse things as its subject; it co-occurs not only with shops or restaurants (4-3a) but also with other things such as galleries, institutions, facilities and so on (4.3b). In contrast, the SJ word *kaiten-suru* (‘open’) is usually associated with the place where goods/foods are provided in exchange for money as its morpheme *ten* (店 ‘shop’) literally indicates (4.3c).

(4.2)

- a. 上品なスーツを着こなしたリッチそうな中年男性である (PB13_00524)

Joohin-na suutsu-o kikonashi-ta ricchi-souna

chuunen dansei dearu

elegant-COP suits-ACC dressed.well-PAST rich-seems.to.be

middle-aged man be

‘He is a middle-aged man dressed well in an elegant suit’

- b. (母親は) 子供が洋服の場合は、子供に合わせてスーツやワンピースが適切である (PB13_00726)

(Hahaoya-wa) kodomo-ga yoofuku-no baai-wa,

kodomo-ni awasete suutsu ya wanpiisu-ga tekisetsu dearu

(mother-TOP) child-NOM Western.clothes-GEN case-TOP,
 child-to dressed.alike suit or one-piece.dress-NOM appropriate be
 ‘(Mother) should wear a suit or a one-piece dress if her child is in Western clothes’

- c. 男達の二人は背広に短靴といういでたちで (PB27_00160)

Otokotachi no futari-wa sebiro ni tangutsu toiu idetachi-de
men of two.people-TOP suit with low.shoes of attire-in
 ‘Two of those men were in their suits with low shoes’

(4.3)

- a. 近所にアイスクリーム屋がオープンした (PB39_00017)

Kinjo ni aisukuriiimu ya-ga oopun-shita
 neighborhood in ice.cream shop-NOM open-do.PAST
 ‘A (new) ice cream shop has been opened in the neighborhood’

- b. 新しい保育所がオープンした (PN1a_00005)

Atarashii hoikusho-ga oopun-shita
 new nurcery.school-NOM open-do.PAST
 ‘A new nursery school has been opened’

- c. シャネルブティック第1号店が開店したカンボン通り (PM51_00101)

Shaneru butikku daiichigoo-ten-ga kaiten-shita kanbon doori
Chanel boutique first-shop-NOM open-do.PAST Cambon street
 ‘Cambon Street, where the first Chanel boutique was opened’

4.2.3. Distribution by sex

Table 4.3 shows the average level of diffusion of each loanword by sex. The differences between males and females are shown in bold when they are 5% or above. Male

speakers appear to have higher preference for: *konsensasu* (< *concensus*), *bijon* (< *vision*) and *risuku* (< *risk*), for example (see also Figure 4.4 for *risuku*). Female speakers, meanwhile, appear to have higher preference for: *suutsu* (< *suit*), *oopun-suru* (< *open*) and *kitchin* (< *kitchen*), and so on (see also Figure 4.2 for *suutsu*).

Table 4.3 Distribution of loanwords by sex

ID	loanwords	etymology	Male	Female	Difference (M-F)
1	<i>niizu</i>	needs	37.9	35.1	2.8
2	<i>ibento</i>	event	58.3	58.1	0.2
3	<i>meritto</i>	merit	58.2	57.6	0.6
4	<i>konsensasu</i>	consensus	12.9	5.6	7.3
5	<i>sukiimu</i>	scheme	4.2	2.8	1.4
6	<i>risuku</i>	risk	38.3	33.2	5.1
7	<i>bijon</i>	vision	34.4	27.4	7.0
8	<i>akauntabiriti</i>	accountability	4.3	3	1.3
9	<i>wain</i>	wine	74.2	76.7	-2.5
10	<i>suutsu</i>	suit	46.3	58.1	-11.8
11	<i>kitchin</i>	kitchen	8.4	15.1	-6.7
12	<i>oopun-suru</i>	open	28.7	39.3	-10.6
13	<i>getto-suru</i>	get	11.7	15.8	-4.1
14	<i>tootaru-de</i>	total	14	12.7	1.3
15	<i>kyanseru</i>	cancel	55.4	59.1	-3.7
16	<i>happii</i>	happy	16.8	18.4	-1.6
17	<i>sapooto</i>	support	32.5	34.2	-1.7

4.2.4. Distribution by age (apparent-time distribution)

Self-reports show that loanwords are divided into three groups according to distribution by age.

4.2.4.1. Loanwords showing a monotonically increasing s-shaped curve

The first group shows a typical s-shaped curve, with the diffusion of loanwords monotonically increasing from older to younger generations. It includes the loanwords such as *wain* (< *wine*), *suutsu* (< *suit*), *getto-suru* (< *get*), *ibento* (< *event*) and *happii* (<

happy). For example, the distribution of *suutsu* (< *suit*) among females shows a typical s-shaped curve, in which the diffusion is accelerated at the age group 50-59 and starts to decelerate at the age group 20-29 (Figure 4.2). *Wain* (< *wine*) appears to be reaching the final stage of an s-shaped growth, where the diffusion starts decelerating at the age group 40-49 and then begins to level out at almost 100% among the age groups 20-29 and below (Figure 4.3).

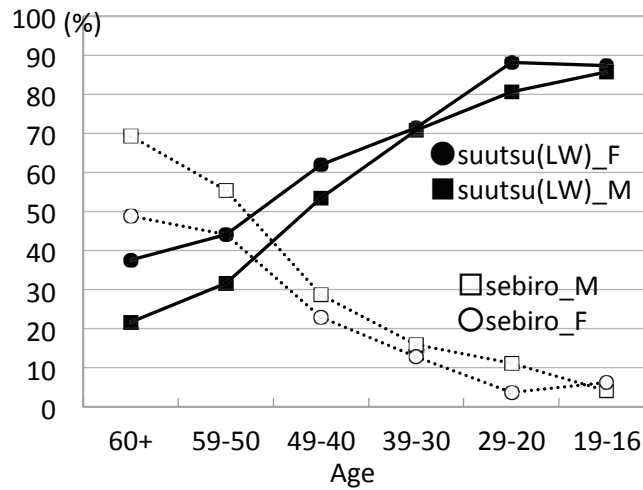


Figure 4.2 Distribution of the LW *suutsu* (< *suit*) and *sebiro*

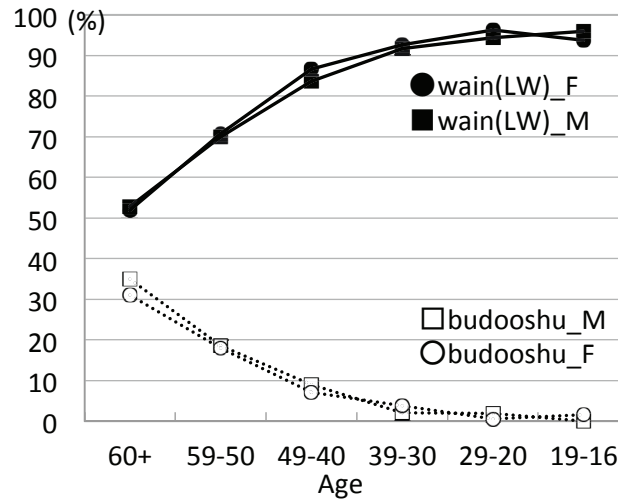


Figure 4.3 Distribution of the LW *wain* (< *wine*) and *budooshu*

4.2.4.2. Loanwords showing an s-shaped curve that ends in a drop

The second group, on the other hand, forms an s-shaped curve that ends in a drop in the diffusion of loanwords among those in their late teens and/or twenties. For example, Figure 4.4 shows that the diffusion of *risuku* (< *risk*) increases as the speakers become younger but there is a drop by approximately 10% between those aged 20-29 and those aged 16-19. A drop among those in their teens is also observed in loanwords such as *sapooto* (< *support*), *kyanseru* (< *cancel*), *oopun-suru* (< *open*) and *bijon* (< *vision*, male speakers only). A drop also occurs among those in their twenties. *Niizu* (< *needs*), for example, peaks at the 30-39 age group and then begin to drift downward at the age group 30-39 (Figure 4.5). A similar pattern is also observed in *meritto* (< *merit*, male speakers only).

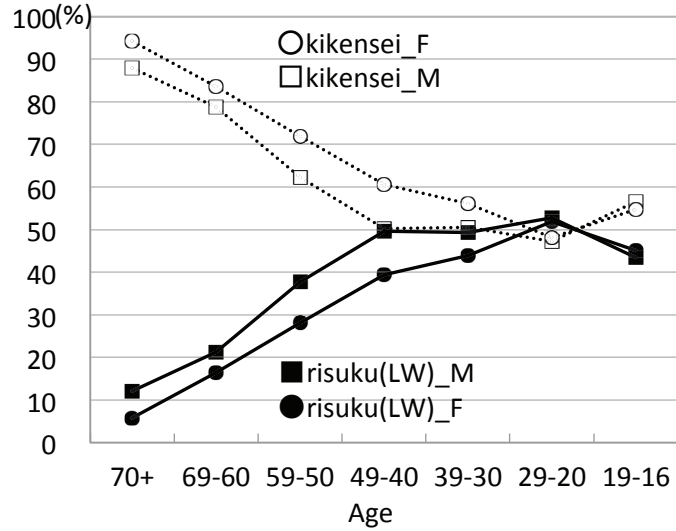


Figure 4.4 Distribution of the LW *risuku* (< risk) and *kikensei*

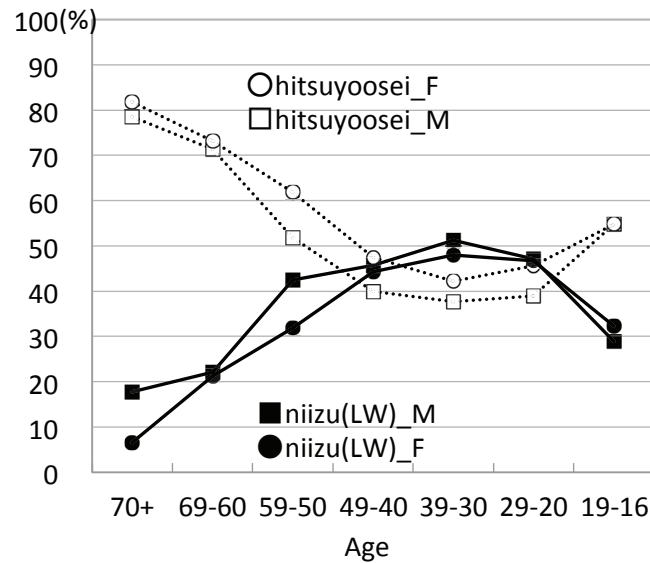


Figure 4.5 Distribution of the LW *niizu* (< needs) and *hitsuyoosei*

4.2.4.3. Loanwords showing no systematic pattern across age

There is no clear age differentiation observed in the loanwords such as *konsensasu* (< *concensus*), *sukiimu* (< *scheme*) and *akauntabiriti* (< *accountability*). We can see in Table 4.2 that the average level of diffusion of these loanwords is less than 10%.

4.2.5. Delay in loanword diffusion among the youngest age group(s)

Why does a delay in the diffusion of loanwords often occur among the youngest age group(s)? A possible reason is that they have not yet fully acquired the usage of the loanwords in question. We notice that such a delay often occurs in loanwords that are well attested in business environments, e.g. *niizu* (< *needs*) and *risuku* (< *risk*). These words appear to be less familiar to those in their late teens or early twenties, who either have not entered or are still new to the job market.²²

In fact, a similar age distribution pattern has often been attested to in the process of dialect standardization in Japan. The studies in Shirakawa and Tsuruoka demonstrated that there is a delay in standardization of the dialects among those aged 20-24 and below (Shibata 1978: 236, see also Kunihiro, Inoue & Long (eds.) 1999: Ch11). He points out that the youngest age groups tend to be delayed in acquiring standard forms because in general they have fewer social opportunities in which the use of standardized language is required than the older age groups do. He predicts from a diachronic point of view that a drop due to acquisitional delay will become gentle or disappear after a while.

Inoue (2000) shows such a process in a diagram (Figure 4.6). The figure demonstrates that the synchronic distribution in which the diffusion of an innovative variant drops at the youngest age group appears in the middle of the whole s-curve

²² BBK (2014: 82-83) examined ten pairs of lexical items consisting of a loanword vs. its native equivalent in an annual opinion poll in 2014. For all the word-pairs, the proportion of those who know a native expression only, i.e., those who does not know the meaning of a loanword, gets lower as they become younger. The 16-19 age group, however, is the only exception: This age group almost always falls behind in acquisition of a loanword compared to those in their 20s (in 9 out of the 10 word-pairs under investigation). These include *niizu* (< *needs*) vs. *hitsuyoosei*; *kyanseru* (< *cancel*) vs. *torikeshi*, and *risuku* (< *risk*) vs. *kikensei*, which are all classified as those showing a delay in acquisition among the youngest age groups. This implies that those in their teens are still in the process of acquiring these loanwords.

growth as shown in (b). It is then turned to a monotonically increasing s-curve at later stages of language change as shown in (c) and (d).

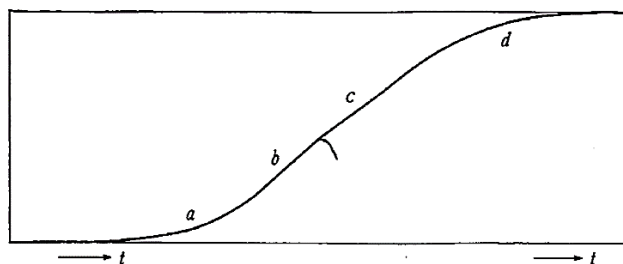


Figure 4.6 Process of standardization over time
(Inoue 2000: 530, Figure 26-4)

Let us turn to the distribution of *niizu* (< *needs*), which shows the most striking acquisitional delay among the youngest speakers as shown in Figure 4.5. Figure 4.7 shows diachronic shifts in the diffusion of *niizu* in terms of the level of recognition, comprehension and the actual use of the lexical item.²³ The result is based on the national survey conducted by NHK in 1995 (NHK 1995: 70) and the opinion polls conducted by BBK in 2002 and 2009 (BBK 2003: 189, 2009: 244). The upward trend over the almost 15 years from 1995 to 2009 confirms that the diffusion of the loanword is in progress in the community overall.

²³ The figures in Figures 4.7 – 4.9 simply represent the ratio of those who recognize, comprehend or use the loanword in contrast to those who do not. Note that it is different from the term “the proportion of those who prefer the loanword over its native equivalents” used throughout this chapter.

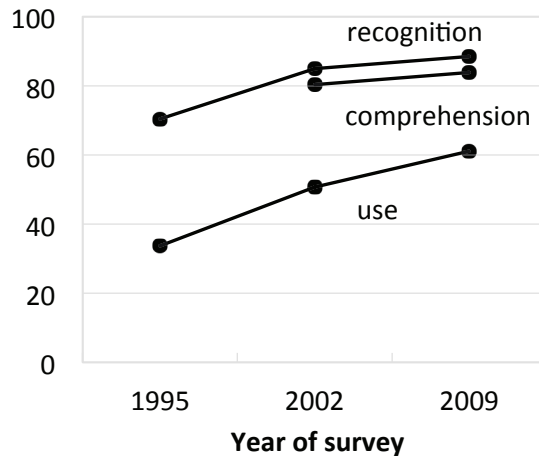


Figure 4.7 Diffusion of *niizu* (< *needs*) in real time
(sources: NHK 1995: 70, BBK 2003: 189, BBK 2009: 244)

The diffusion, however, does not progress in the same way for all age groups. Figure 4.8 compares the proportion of those who stated that they actually use *niizu* by speakers' age and year of survey.²⁴ The figure shows that the use of *niizu* increased in seven years from 2002 to 2009 among those aged 50-59 and above, meanwhile it remains at almost the same level for the age groups of 40-49 and below (except the age group 30-39). This implies that the change in favor of *niizu* is in progress in the former age groups. When the distribution in Figure 4.8 is sorted by speakers' year of birth as shown in Figure 4.9, we notice that the use of *niizu* between 2002 and 2009 had increased among those born in the 1960s and later. This implies that the acquisitional delay in the youngest generations at certain point of time changes as the speakers acquire the word as they age.

²⁴ Age-specific proportions are not available in the report of the 1995 survey (NHK 1995).

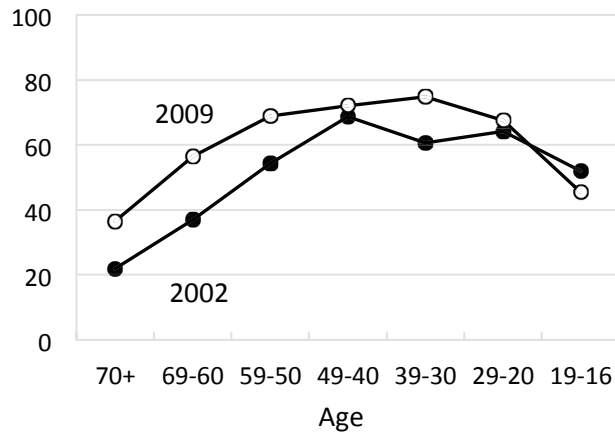


Figure 4.8 Use of *niizu* (< needs) by age and year of survey

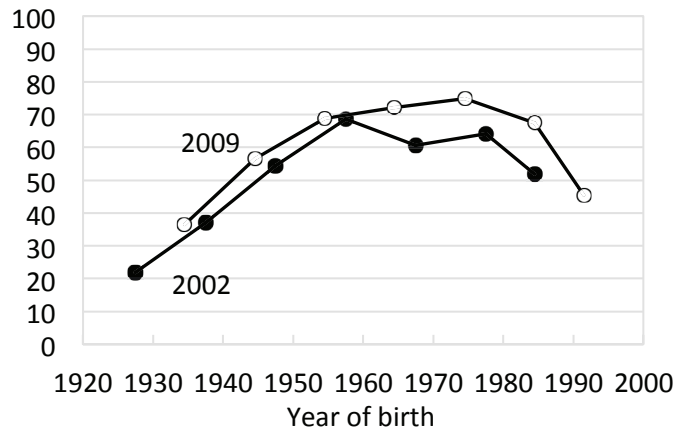


Figure 4.9 Use of *niizu* (< needs) by age and year of birth

The above discussion leads us to conclude that this is a change in progress in the community as a whole, but, at the same time, involves age-grading among the youngest age groups. It seems that the diffusion of the loanword *niizu* is in the middle of an s-curve growth, i.e., located in (b) in Figure 4.6. We can therefore predict that synchronic distribution of *niizu* will become a monotonically increasing s-curve pattern as shown in (c) or (d) in Figure 4.6 as the diffusion process reaches a more advanced stage.

It is then clear that the second group of loanwords, like *niizu*, should be predictable by a monotonically increasing s-curve model when viewed from a long-range perspective of language change, as is the case with the first group. This raises the question of how the influence of the acquisitional delay should be treated in an s-curve model. This point will be discussed in the following section.

4.3. Estimating an s-curve model

4.3.1. Logistic regression analysis

A logistic regression analysis is employed to model the diffusion of the 17 loanwords. Before conducting analysis, we have to consider how to treat the influence of the acquisitional delay in the model. A monotonically increasing logistic curve does not fit well to the data if the loanword distribution ends in a drop among the youngest age groups, like *niizu* does. In such a case, a possible way of tackling the problem would be to exclude data from the youngest generations in which a drop occurs. Exclusion of data, however, may not be the best choice in this case as they are not completely random outliers: The observed age differentiation (acquisitional delay) is well attested across various linguistic variables in the language (Shibata 1978). Another possibility then is the inclusion of such a disruptive effect as an independent variable. Yokoyama & Sanada (2007) suggest that a disruptive factor can be separated from independent variables of capital interest to an analyst by including it as one of the independent variables. The disruptive factor called “Acquisition” is therefore included in the model.

The general model of the loanword diffusion is shown in [Formula 2], with (i) Year of Birth (converted from age), (ii) Sex (M=0, F=1) and (iii) Acquisition included as

independent variables. p is the probability of the diffusion of (or preference for) a loanword as opposed to its native equivalent(s) and a_1 - a_3 represent parameter values of the independent variables x_1 - x_3 .

$$\log[p/(1-p)] = a_1 (x_1: \text{Year of Birth}) + a_2 (x_2: \text{Sex}) + a_3 (x_3: \text{Acquisition}) + b \text{ [Formula 2]}$$

Note that “Acquisition” was coded as a numeral variable. The value represents a distance from an age group with the highest level of loanword adoption: the higher the value is, the more striking the delay in the loanword acquisition is. For instance, if a drop in diffusion occurs between those aged 20-29 and those aged 16-19, those aged 20-29 and older are coded as “0” whereas those aged 16-19 are coded as “1.” In the case of a drop occurring between the age group 30-39 and the age group of 16-19, those aged 30-39 and older are coded as “0,” those aged 20-29 are coded as “1” and those aged 16-19 are coded as “2.”²⁵ Table 4.4 shows how samples are coded for as an example (*niizu*).²⁶

²⁵ In fact, most of those classified as “loanwords showing a monotonically increasing s-shaped curve” in 4.2.4, numerically speaking, involve a slight drop among the youngest age group. Although it seems very small, such a drop is marked as “acquisitional delay.” (See Appendix A for detail.)

²⁶ To avoid the problem of multicollinearity, correlations between independent variables were evaluated for each data set. It showed that the independent variables are not highly correlated with each other (the highest correlation coefficient remains less than 0.4) and that VIF, which is one of the indicators of the risk of multicollinearity, is less than 4 in each case.

Table 4.4 Data coding for *niizu*

Sex	Age	Year of Birth (median)	Acquisition	LW diffusion (%)	
M	0	16-19	1981.5	2	29.0
M	0	20-29	1974.5	1	47.2
M	0	30-39	1964.5	0	51.3
M	0	40-49	1954.5	0	45.7
M	0	50-59	1944.5	0	42.5
M	0	60-69	1934.5	0	22.2
M	0	70+	1924.5	0	17.8
F	1	16-19	1981.5	2	32.3
F	1	20-29	1974.5	1	46.8
F	1	30-39	1964.5	0	48.0
F	1	40-49	1954.5	0	44.3
F	1	50-59	1944.5	0	31.9
F	1	60-69	1934.5	0	21.3
F	1	70+	1924.5	0	6.6

4.3.2. Results

The logistic regression analyses (stepwise) of the 17 variables were conducted with SPSS. The results are listed in Table 4.5 in order of the overall level of diffusion (“Diffusion (%)”). “Regression coefficient (B)” in the table corresponds to a_n in [Formula 2] and represents how much effect an independent variable (x_n) has on the probability (p). A positive coefficient means that the independent variable has a positive effect on p , i.e., a higher level of diffusion of a loanword, whereas a negative coefficient refers to a negative effect of an independent variable on p . “0” refers to “no significance” of the effect of a variable on the outcome. “NA” means that the factor in question is irrelevant and thus was not incorporated into the model. “Constant” in the table corresponds to b in [Formula 2]. R^2 is a coefficient of determination (contribution ratio), i.e., the square of a coefficient of correlation between observed values and predicted values (Yokoyama & Wada 2006). It ranges from 0 to 1 (1 represents the maximum contribution).

As for loanwords whose diffusion involves a delay in acquisition among the youngest generation, two types of analyses were conducted: (i) one that includes the variable “Acquisition” in the model (Model A) and (ii) the other that does not include it (Model B). Table 4.5 shows the values of R^2 for Model B where relevant. The positive values in the difference in R^2 between Model A and B ($R^2 [A-B]$) show that Model A better fits to the data in most cases. (Note that the values of “Regression Coefficient (B)” and “Constant” for Model B are omitted from the table for this reason.)

Since R^2 usually increases when more independent variables are added to a model, AIC is also calculated in order to evaluate the contribution of the inclusion of a new variable. Table 4.5 compares the value of AIC for Model A with the value of AIC for Model B. It also shows that Model A gives better fits to the data since it always has the smaller AIC.

Table 4.5 Results of the regression analyses (17 variables)

LW	etymology	Diffusion (%)	B				Constant (b)	R ²			AIC		
			Year of Birth (a ₁)	Sex (a ₂)	Acquisition (a ₃)	Model		Model					
						A		B	A-B	A	B	A-B	
<i>wain</i>	wine	75.6	0.0746	0	-0.9367	-144.2208	0.99	0.99	0	59.58	59.58	0	
<i>ibento</i>	event	58.2	0.0590	0	-1.2462	-114.6673	0.96	0.91	0.05	88.14	106.81	-18.67	
<i>meritto</i>	merit	57.9	0.0497	-0.2466	-0.8044	-96.5074	0.91	0.75	0.16	94.90	125.77	-30.87	
<i>kyanseru</i>	cancel	57.5	0.0320	0	-0.8180	-62.1183	0.90	0.64	0.26	71.13	83.81	-12.68	
<i>suutsu</i>	suit	52.8	0.0632	0.4099	0	-123.4208	0.98	0.98	0	69.10	69.10	0	
<i>niizu</i>	needs	36.4	0.0484	0	-0.8939	-94.8470	0.89	0.39	0.50	89.86	143.61	-53.75	
<i>risuku</i>	risk	35.5	0.0430	-0.2468	-0.8558	-84.4045	0.89	0.76	0.13	93.99	111.84	-17.85	
<i>oopun-suru</i>	to open	34.6	0.0148	0.4192	0	-29.7063	0.86	0.86	0	69.90	69.90	0	
<i>sapooto</i>	support	33.4	0.0411	0	-0.9895	-80.9564	0.93	0.72	0.21	85.28	109.62	-24.33	
<i>bijon</i>	vision	30.7	0.0335	-0.4330	-0.7988	-66.0486	0.93	0.83	0.10	82.89	91.58	-8.69	
<i>happii</i>	happy	17.7	0.0252	0	-0.5711	-50.7461	0.71	0.57	0.14	70.37	72.79	-2.41	
<i>getto-suru</i>	to get	14.0	0.0786	0	NA	-155.9317	0.97	NA	NA	60.70	NA	NA	
<i>tootaru-de</i>	in total	13.3	0.0193	0	0	-39.6040	0.65	0.65	0	62.74	62.74	0	
<i>kitchin</i>	kitchen	12.1	0	0.6065	NA	-2.3075	0.68	NA	NA	60.01	NA	NA	
<i>konsensasu</i>	consensus	9.0	0	-0.8833	NA	-2.0198	0.41	NA	NA	82.91	NA	NA	
<i>akauntabiriti</i>	accountability	3.6	0	0	NA	-3.3621	<0.01	NA	NA	64.78	NA	NA	
<i>sukiimu</i>	scheme	3.5	0	0	NA	-3.4305	0.02	NA	NA	58.88	NA	NA	

Year of Birth is shown to be significant for all but the four items at the bottom, i.e. *kitchin*, *konsensasu*, *akauntabiriti* and *sukiimu*. Note that the level of diffusion of these four loanwords remains very low (3.5 – 12.1%). Given that the process of linguistic change progresses in a similar manner to that of “diffusion of innovations” (Rogers 2003) as Jinnouchi (2007) points out, these four loanwords appears to have been at a very early stage of diffusion, where only a small number of people called “innovators” or “early adopters” adopt an innovation as shown in Figure 4.10. According to the figure, the diffusion of a given innovation rapidly grows after being adopted by 16% of the population, where it then starts to be adopted by the “early majority.” It is thus possible that age differentiation is not significant in the case that the level of diffusion does not greatly exceed 16%.

Meanwhile, Year of Birth is shown to be a significant variable for the rest of 13 loanwords, including those of which diffusion largely exceeds 16% as well as those of which diffusion is more or less 16% (i.e. *happii*, *getto-suru*, *tootaru-de*). A positive coefficient indicates that the younger the speakers are, the higher the probability of the diffusion of loanwords becomes. The result allows us to interpret that the diffusion of these 13 loanwords is in progress in apparent time, indicating that there is an ongoing change toward more frequent use of these loanwords. Sex and Acquisition are not always significant. The effects of these factors are interpreted in the following section.

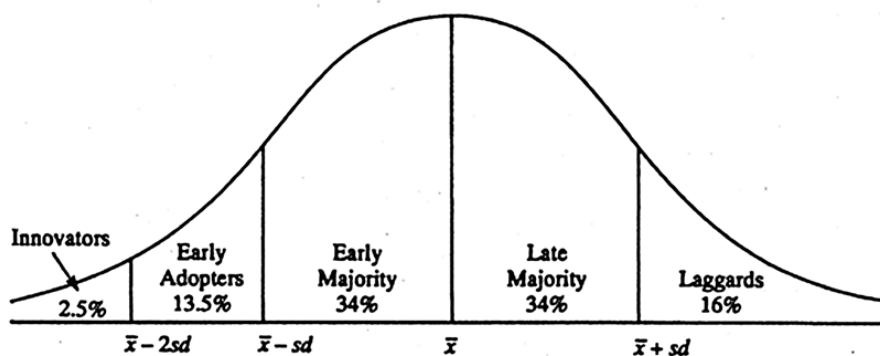


Figure 4.10 Adopter categorization on the basis of innovativeness (Rogers 2003: 281, Figure 7-3)

4.4. Predicting change

This section attempts to make a prediction of the probability of loanword diffusion based on the results of the logistic regression analyses shown in Table 4.5. The probability p is calculated with [Formula 2'], which is derived from [Formula 2]. For example, the probability of the diffusion of *niizu* (< *needs*) among female speakers aged 16-19 is predictable with the equation in [Formula 2'']. The prediction was made for 100

years of generations including the observed age groups (50 to 60 years of generations) as well as the future generations following them.

$$p = 1/[1+\exp(-1)(a_1(\text{Year of Birth}) + a_2(\text{Sex}) + a_3(\text{Acquisition}) + b)] \quad [\text{Formula 2'}]$$

$$p = 1/[1+\exp(-1)(0.0484(1981.5) + 0(1) + (-0.8939)(2) + b)] \quad [\text{Formula 2''}]$$

4.4.1. Cases in which neither Acquisition nor Sex is significant

This category includes *getto-suru* (< *get*) and *tootaru-de* (< *total*). In these cases, the diffusion of the loanwords is predictable from Year of Birth alone. Figure 4.11 shows the predicted probability of the preference for *getto-suru* over its native equivalent *kakutoku-suru* ('to get'). It is predicted that its diffusion will proceed without a delay among the youngest age group(s).

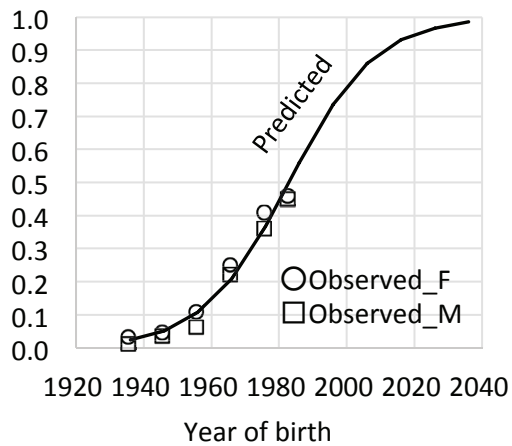


Figure 4.11 Prediction of the diffusion of *getto-suru* (< *get*)

4.4.2. Cases in which Acquisition is significant

Acquisition is shown to be significant for all lexical items but *suutsu* (< *suit*), *oopun-suru* (< *open*) and *tootaru-de* (< *total*). In all the cases this factor has a negative effect on the loanword diffusion. The inclusion of this variable, however, does not always greatly improve the model's goodness of fit to the data. Those loanwords whose coefficient of determination (R^2) increase by 0.1 and above include *meritto* (< *merit*), *kyanseru* (< *cancel*), *niizu* (< *needs*), *risuku* (< *risk*), *sapooto* (< *support*), *bijon* (< *vision*) and *happii* (< *happy*) as shown in bold in Table 4.5.

R^2 for *niizu* (< *needs*), for example, reaches 0.89 when Acquisition is incorporated into the model, whereas it decreases to 0.39 when this variable is not included. Figure 4.12 compares the goodness of fit of the models with each other (in the case of male speakers). The figure demonstrates that Model A fits the observed values much better than the Model B (the same holds true for female speakers). This implies that the inclusion of the variable "Acquisition" in the model contributes to a better fit to the data.

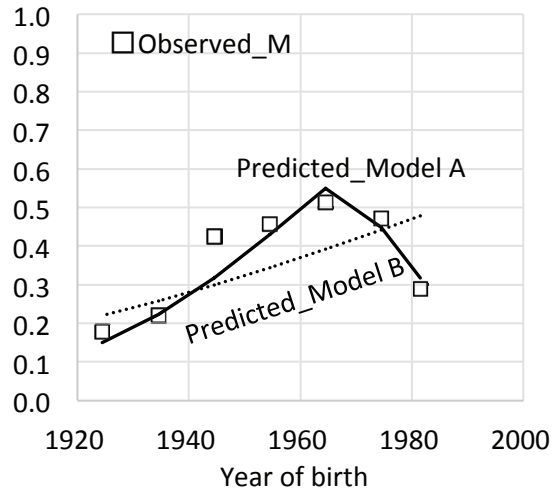


Figure 4.12 Comparing the goodness of fit of the models for *niizu* (< *needs*)

Then we turn to Figure 4.13, where the loanword diffusion is predicted with Model A, with the effect of Acquisition disregarded (eliminated).²⁷ By doing this we can visualize the effect of speakers' Year of Birth alone. Figure 4.13 clearly shows that the rate of change has a steeper rise when predicted with Model A than when predicted with Model B. In other words, Model B underestimates the rate of change.

²⁷ The effect of Acquisition is eliminated by substituting 0 for a_3 in [Formula 2'].

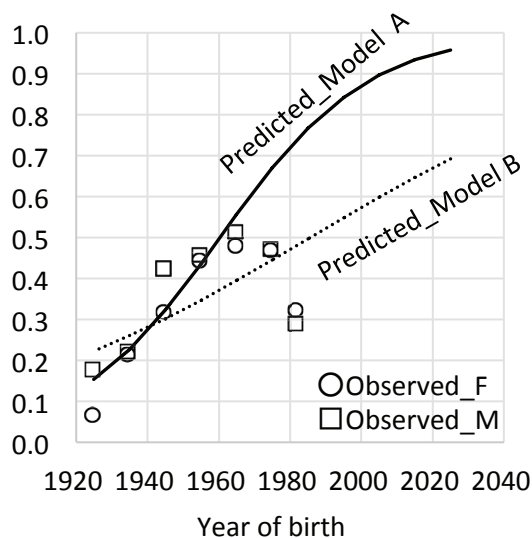


Figure 4.13 Prediction of the diffusion of *niizu* (< needs)

4.4.3. Cases in which Sex is significant

Sex differentiation is significant for only 7 items. A negative coefficient indicates that male speakers tend to use the loanword in question more frequently than female speakers, and a positive coefficient suggests the opposite. The former group includes *meritto* (< merit), *risuku* (< risk), *bijon* (< vision) and *konsensasu* (< consensus) and the latter includes *suutsu* (< suit), *oopun-susu* (< open) and *kitchin* (< kitchen). We can see that male speakers are ahead of female speakers in the diffusion of *bijon* (Figure 4.14) whereas female speakers are ahead of male speakers in the adoption of *oopun-suru* (Figure 4.15).

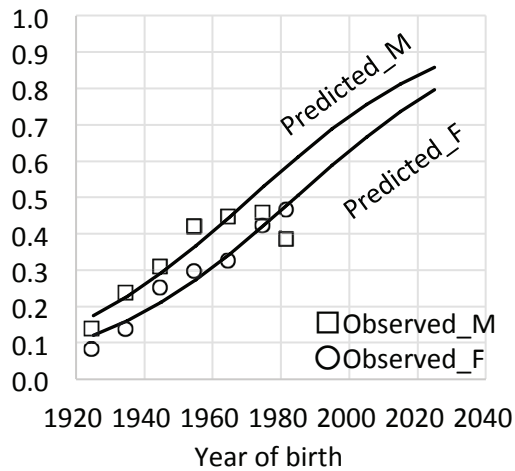


Figure 4.14 Prediction of the diffusion of *bijon* (< *vision*) (men ahead of women)

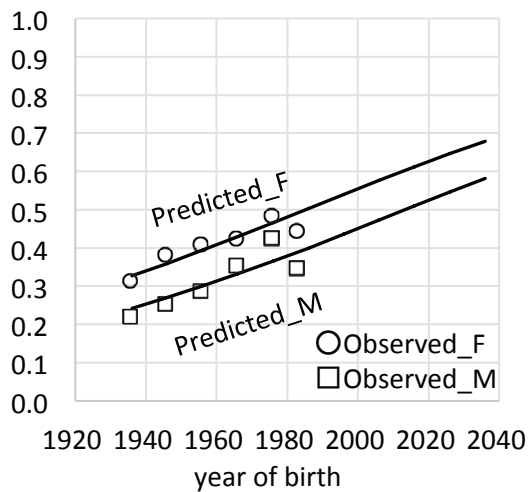


Figure 4.15 Prediction of the diffusion of *oopun-suru* (< *open*) (women ahead of men)

It is assumed that sex differentiation is caused by the context in which the loanwords in question are used: *Meritto* (< *merit*), *risuku* (< *risk*), *bijon* (< *vision*) and *konsensasu* (< *consensus*) are well attested in the domain of business and not heavily used in everyday life, whereas *suutsu* (< *suit*), *oopun-susu* (< *open*) and *kitchin* (< *kitchen*) are attested more frequently in everyday life. Since male employment rate is

much higher than female employment rate,²⁸ i.e., men are more strongly associated with the business environment than women, they are likely to use business-related terms more often than women. It should be noted that different parameters, such as occupation, could come into play if the social advancement of women is achieved further in future.

Furthermore, it is important to pay attention to the fact that sex differentiation is induced by the semantic difference between competing variants. For example, the loanword *suutsu* (< *suit*) refers to a set of clothes to be worn by both men and women, whereas its native equivalent *sebiro* often refers only to men's clothing (*Nihon Kokugo Daijiten* 2nd edition). Given that some informants assumed that the reference was to their own clothing during their interviews, women would have had only one choice, i.e., the loanword *suutsu*. This could lead to the predominance of the female speakers in selecting the loanword.

4.5. Discussion and Conclusion

Self-reports in the previous nationwide surveys have shown that there are two systematic patterns of loanword diffusion according to distribution by age. One shows a monotonically increasing s-shaped curve, and the other shows an s-shaped curve that ends in a drop among those aged 16-19 and/or 20-29. The delay in the acquisition of the loanwords in the latter group arguably reflects the age-specific context of use of the loanwords in question. That is, the domain- or context-specific use of loanwords induced the acquisitional delay among the youngest age groups, who have not been in the

²⁸ According to the result of the census conducted in 2000 (Soomushoo Tookeikyoku (Statistics Bureau, Ministry of Internal Affairs and Communications, Japan)), the employment rate is 74.8% for males and 48.2% for females (for those aged 15 and above). (http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000000030587&cycleCode=0&requestSender=search, retrieved on 29/3/2016)

job market for long or at all. We should note that “age is not merely the biological age, but includes concomitant social conditions” (Kunihiro et al. (eds.) 1999: 211).

We saw that those in the youngest age groups acquire the loanwords in the latter group, such as *niizu*, as they age (Figure 4.9). Consequently, I concluded that both types of loanword diffusion should be predictable with a monotonically increasing s-curve model. This raises the question as to how the influence of the “acquisitional delay” could be separated from the speakers’ birth-year effect without excluding the data of the youngest generations. The problem is shown to be solvable by incorporating an “Acquisition” term into the model as one of the independent variables. The inclusion of this new variable, as a result, has advantages as a more accurate prediction of change: It enables us to elicit the birth-year effect alone, and in turn, corrects underestimations of the rate of change.

There are nevertheless some problems left unanswered. The first question is whether age differentiation can be simply interpreted as an ongoing change. A more careful interpretation is required when it comes to lexical items of which use is specific to a certain domain, e.g., business related terms like *niizu* (< *needs*), as age differentiation could be a reflection of age-grading in such cases. In this chapter the real-time approach has shown that the diffusion of the business-related term *niizu* is indeed a change in progress in the community as a whole (although age-grading occurs among the youngest age groups). Ideally speaking, verification by real-time observations would be needed for all other lexical items.

Second, the predicted rate of change should be treated carefully. The application of the apparent-time hypothesis, i.e., the assumption that one’s linguistic system is

invariable throughout one's lifetime, should be done with care, especially for lexical variation. In general, "people can easily add to their vocabulary throughout their lives" (Meyerhoff 2011: 160). It is often the case that, in particular, people have a good chance of acquiring new loanwords, vogue words or buzzwords after their critical period.

The above-mentioned problems will be solved by conducting a real-time study of the phenomenon. I will discuss how to overcome the limitations of an apparent-time approach in Chapter 8.

5. Distribution of loanwords in corpora: A comprehensive approach

5.1. Introduction

5.1.1. Aim: Establishing tendencies in synchronic distribution of loanwords

This chapter aims to look at the general tendency in synchronic distribution of loanwords in corpora by means of “comprehensive approach” (See 5.3.5) to lexical variation.

The last chapter has shown that the public attitude exhibits an upward tendency in loanword diffusion along apparent time. It demonstrated that in some cases informants’ readiness to use loanwords as opposed to their native equivalents systematically increases as informants become younger. This chapter further investigates whether a similar tendency is attested to by their actual linguistic behavior using corpora as synchronic collections of real language data.

First, we see that the loanword stratum as a whole is on the increase in apparent time in both written and spoken Japanese. Second, after selecting approximately 25 loanwords as potential candidates for new sociolinguistic lexical variants as opposed to their native equivalents, I describe in what manner the loan variants are distributed according to various extra-linguistic contexts. These include not only age but also education, register and style. Certain tendencies in synchronic distribution of the loanwords are elicited despite the methodological limitations of the comprehensive approach in the study of variation, i.e., difficulty in identifying their native variants and insufficiency of identification of the envelope of variation.

5.1.2. Research questions

This and the next chapters are designed to answer the following research question: What is the general tendency in synchronic distribution of loanwords as opposed to their native equivalents, depending on one's social attributes and style?

To answer the above question in the context of variationist study, we need to tackle the following methodological problems:

- (i) Which loanwords are potential candidates for newer linguistic variants?
- (ii) Which items are native equivalents of a given loanword?
- (iii) What is considered to be the envelope of variation?

We will first inquire which loanwords are used frequently in a corpus to tackle the first issue (5.3.2). We will then attempt to identify native equivalents of the selected loan variants to satisfy the second inquiry (5.3.4). We will explore the last question in earnest in Chapter 6 by carefully identifying a functional overlap among variants by context under the name of “contextual approach.” To begin with, this chapter will examine the phenomenon without identifying the envelope of variation in a strict sense in order to gain a quick overview of a *general* trend in the synchronic distribution of loanwords (“comprehensive approach”) (5.3.5).

5.2. Data

5.2.1. BCCWJ and CSJ as synchronic corpora

The Publication sub-corpus in BCCWJ and the monologue components in CSJ will be investigated as synchronic corpora. Both of the corpora were collected in the early 2000s: the former concerns written texts published between 2001 and 2005 and the latter concerns spoken data recorded between 1999 and 2003. These corpora are then considered synchronic corpora that allow us to access collections of language produced roughly in the beginning of the 21st century.

The Publication sub-corpus in BCCWJ includes information on the year of birth, sex and educational background of the authors. Besides, the sub-corpora consists of three types of media: books (PB), magazines (PM) and newspapers (PN), so it allows us to make a comparison of the probabilities of loanwords that appear across different registers (note that the Library sub-corpus contains books only). Among these, however, PN was excluded from the analysis since it is very small in size compared with the other two registers.²⁹

In CSJ, as already mentioned in 3.1.3, the majority of recordings are monologues uttered in two relatively formal settings: Academic Presentation Speech (APS) and Simulated Public Speech (SPS). CSJ provides the information on the author's year of birth, sex, educational background and register (APS vs. SPS). It also contains information on evaluations of stylistic aspects such as the formality, spontaneity and audience size, of each sample. Although the register difference in CSJ is considered to some extent a reflection of difference in speech style, i.e., APS has a more formal style

²⁹ PB, PM and PN consist of approximately 28.6, 4.4 and 1.4 million words, respectively.

than SPS, it will be useful for us to have a look at these factors as additional indicators of style.

5.2.2. Upward tendency in loanword diffusion in apparent time

Table 5.1 and Table 5.2 show the distribution of loanwords (LW), Sino-Japanese words (SJ), native Japanese words (NJ) and hybrids (HB) in BCCWJ (fixed-length data) as well as that in CSJ, by the author's/speaker's year of birth.³⁰ Figure 5.1 shows the proportion (%) of LW as opposed to the sum of all lexical strata (LW + SJ + NJ + HB) for each birth-year group based on Table 5.1. For example, Figure 5.1 shows that, by token, the proportion of LW used by those born in the 1920s is 2.3% ($=4,409/189,892*100$). Note that this figure (%) is not equal to the proportion (%) shown in the "Total" column in Table 5.1, as the latter figure (%) represents the proportion of each age group out of the total samples, i.e. the sample size of those born in the 1920s accounts for 8.5% of all ($=189,892/2,230,782*100$). The same holds true for Table 5.2 and Figure 5.2.

In Table 5.1 the average proportion of LW in BCCWJ is 4.3% by token and 10.3% by type. It should be noted that the size (i.e. the word counts) of samples collected from texts written by those born in the 1910s and before, and by those born in the 1970s and after are extremely small (roughly, less than 5% of the total) (see the column "Total (in Token)" in Table 5.1). Mori (2013: 279) points out that when a small-scale birth-year

³⁰ Particles and auxiliaries are excluded. Fillers, which are particularly frequent in spoken language, are also excluded. In BCCWJ, the investigation is limited to the fixed-length samples written by a single author whose year of birth is available. In CSJ, the investigation is limited to the samples uttered by a speaker whose year of birth is available.

group shows an irregular distribution, it should be interpreted with care.³¹ If we ignore such small-scale age-groups and focus on those born between the 1920s and 1960s in Figure 5.1, we can see a clear tendency for the proportion of LW to increase monotonically from younger generations to older ones in both token and type. For example, the difference between those born in the 1920s and those born in the 1960s are approximately 4% by token and 7% by type.

Likewise, we find a relatively consistent trend toward an increase of LW in both token and type along apparent time in CSJ (Figure 5.2) especially when focusing on those born between 1940-44 and 1975-79, in which the size of sample of each birth-year group is relatively stable (more than 5% of the whole sample).³² For example, the difference between those born in 1940-44 and those born in 1975-79 are approximately 2% by token and 4% by type. The average proportion of LW in CSJ is 5.0% by token and 15.1% by type.

These figures indicate that the vocabulary in the LW stratum as a whole is on the increase in the Japanese lexicon in both written and spoken language.

³¹ Mori points out that the social distributions of authors in BCCWJ are imbalanced. He shows that birth year of the author is available only in 71.4% of the whole sample in the Library sub-corpus. Among those, the majority of authors were those born between the 1920s and 1960s and the proportion of those born before or after that period account for only a small portion. Some attention is required when treating data related to these birth-year groups. He also points out that samples collected from those born in the 1910s and before are likely to be of reprinted books, since those born in the 1910s should be in their 90s in 2001-2005. It appears that the same holds true for the Publication sub-corpus.

³² Note that size (i.e. the word counts) of samples collected from those born in 1935-39 and before and those born in 1980-84 are extremely small (less than 5%, see Table 5.2) and show occasional irregularities.

Table 5.1 Distribution of the lexical strata in apparent time (BCCWJ³³)

Year of birth	Token						Type					
	LW	SJ	NJ	HB	Total		LW	SJ	NJ	HB	Total	
1830		203	203	13	419	0.0%		115	110	8	233	0.1%
1840		198	402	10	610	0.0%		124	136	6	266	0.2%
1850		175	221	23	419	0.0%		127	122	15	264	0.2%
1860	27	899	1453	51	2430	0.1%	14	515	543	36	1108	0.7%
1870	45	1640	3064	103	4852	0.2%	24	877	1069	64	2034	1.2%
1880	47	2266	3499	112	5924	0.3%	36	1120	1191	75	2422	1.4%
1890	147	4085	6734	185	11151	0.5%	85	1718	1563	95	3461	2.0%
1900	333	6906	11942	346	19527	0.9%	193	2652	2172	149	5166	3.0%
1910	648	19078	24835	815	45376	2.0%	322	5017	3385	279	9003	5.3%
1920	4409	82618	99565	3300	189892	8.5%	1405	11164	6929	666	20164	11.9%
1930	12319	174353	194151	6620	387443	17.4%	2657	14321	8859	840	26677	15.7%
1940	21784	251525	268852	9723	551884	24.7%	3623	15466	9803	976	29868	17.6%
1950	24978	230390	266055	8994	530417	23.8%	3779	14451	9638	947	28815	17.0%
1960	23154	151127	188154	6484	368919	16.5%	3421	11556	8310	774	24061	14.2%
1970	6894	35296	56725	1773	100688	4.5%	1736	5854	4996	370	12956	7.6%
1980	470	2831	6987	212	10500	0.5%	237	1243	1609	90	3179	1.9%
1990	31	78	213	9	331	0.0%	24	69	148	9	250	0.1%
Total	95286	963668	1133055	38773	2230782	100.0%	17556	86389	60583	5399	169927	100.0%
	4.3%	43.2%	50.8%	1.7%	100.0%		10.3%	50.8%	35.7%	3.2%	100.0%	

LW: Loanword, SJ: Sino-Japanese, NJ: Native Japanese, HB: Hybrid

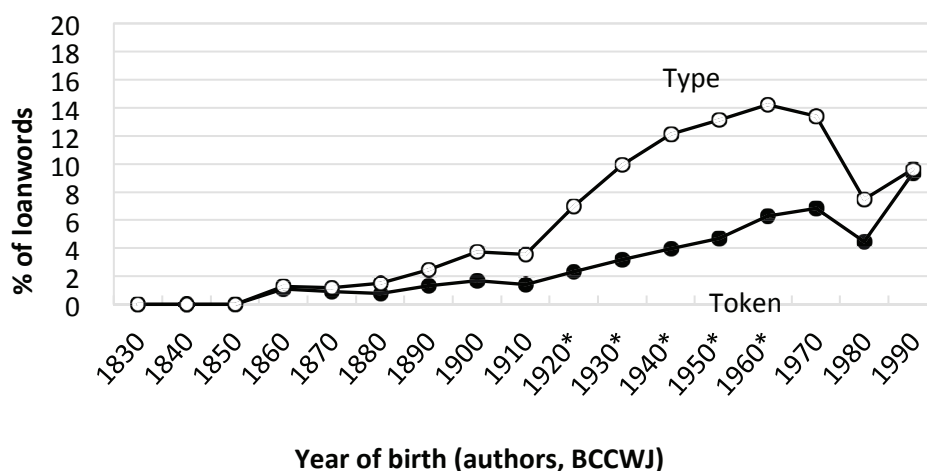


Figure 5.1 Distribution of the LW stratum in apparent time (BCCWJ)

³³ The BCCWJ data is analyzed with UniDic, an electronic dictionary that annotates Japanese texts with morphological information. Here, SUW (short unit word) is used for counting words. For more detail about morphological analyses in BCCWJ, see NINJAL (2011: Ch4).

Table 5.2 Distribution of the lexical strata in apparent time (CSJ³⁴)

Year of birth	Token					Type						
	LW	SJ	NJ	HB	Total	LW	SJ	NJ	HB	Total		
1915-19	278	2389	5265	60	7992	0.2%	82	670	599	26	1377	1.0%
1920-24	78	978	1259	35	2350	0.1%	13	185	226	11	435	0.3%
1925-29	81	1161	1939	56	3237	0.1%	33	335	326	18	712	0.5%
1930-34	3815	41366	84248	2777	132206	3.9%	977	5137	3731	294	10139	7.5%
1935-39	5013	45777	101828	2204	154822	4.6%	1138	4993	3970	306	10407	7.7%
1940-44	6907	56785	115222	2729	181643	5.4%	1431	5216	3937	312	10896	8.0%
1945-49	10918	78546	163165	3962	256591	7.6%	1919	5978	4645	364	12906	9.5%
1950-54	7684	63026	117897	2827	191434	5.7%	1502	5203	3756	319	10780	7.9%
1955-59	17761	123837	221036	5716	368350	10.9%	2345	6596	4911	395	14247	10.5%
1960-64	19855	135345	226960	5906	388066	11.5%	2423	6390	4701	453	13967	10.3%
1965-69	26185	159319	270183	6903	462590	13.7%	2675	6886	5030	424	15015	11.1%
1970-74	33636	221341	350625	9272	614874	18.2%	2884	7351	5607	538	16380	12.1%
1975-79	36122	216930	320087	8194	581333	17.2%	2562	6637	5055	441	14695	10.8%
1980-84	1604	9980	22195	499	34278	1.0%	493	1695	1617	103	3908	2.9%
Total	169937	1156780	2001909	51140	3379766	100.0%	20477	63272	48111	4004	135864	100.0%
	5.0%	34.2%	59.2%	1.5%	100.0%		15.1%	46.6%	35.4%	2.9%	100.0%	

LW: Loanword, SJ: Sino-Japanese, NJ: Native Japanese, HB: Hybrid

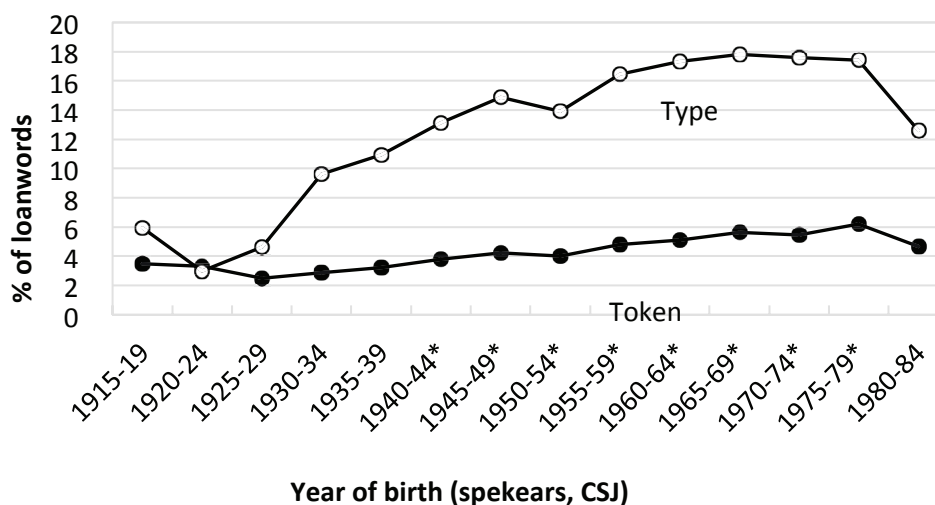


Figure 5.2 Distribution of the LW stratum in apparent time (CSJ)

³⁴ I used the UniDic version of CSJ (SUW) (not the original version). It is noted that, since I used the available in-progress data as of November 2014 (with permission from NINJAL), the results shown here will be different from what will be the finalized version. The finalized version was made available to corpus DVD holders from 1st September 2016. For more detail about the UniDic version of CSJ, see Watanabe, Tanaka & Koiso (2015).

5.3. Selecting potential lexical variables

5.3.1. Exploring loanwords by parts of speech and usage

Table 5.3 shows the distribution of each lexical stratum by parts of speech in the fixed-length component of PB and PM in the Publication sub-corpus, BCCWJ. Particles and auxiliaries are excluded.

Table 5.3 Lexical strata by part of speech (type, BCCWJ)

	LW	SJ	NJ	HB	Total	Example	Etymology
noun							
(i) noun	9801	21075	11479	2026	44381	<i>hoteru</i>	hotel
(ii) verbal noun	527	6393	510	64	7494	<i>kyanseru</i>	cancel
(iii) adjectival noun	209	787	115	44	1155	<i>orijinaru</i>	original
(iv) verbal/adjectival noun	9	51	7	1	68	<i>oopun</i>	open
(v) adverbial noun		236	143	18	397		
(vi) counter noun	148	124	37	11	320	<i>paasento</i>	percent
(vii) numeral noun	1	59	17	13	90	<i>zero</i>	zero
(viii) other noun			1		1		
(ix) adjectival noun	215	573	391	48	1227	<i>shinpuru</i>	simple
pronoun		13	74	6	93		
verb			5221	309	5530		
adjective			549	31	580		
adverb		105	1579	27	1711		
adnominal			38	8	46		
interjection	2	1	185	4	192	<i>bonjoruno</i>	Buongiorno!
conjunction		2	29		31		
prefix		206	25		231		
suffix	4	491	192		687	<i>-chikku</i>	-tic
Total	10916	30116	20592	2610	64234		

Note that the majority of loanwords are nouns. Nouns are subdivided into several categories ((i)-(viii)) depending on their possible usages.

First, the most frequent category is (i) noun (9,801 types). A noun such as *hoteru* (< *hotel*) is used only as a noun followed by a postpositional particle or a copula as in (5.1) and (5.2).

The second most frequent category is (ii) verbal noun (527 types). A verbal noun such as *kyanseru* (< *cancel*) is used as a noun when followed by a particle/copula as in (5.2) while it serves as a verb when compounded with a light verb such as *suru* ('do') or *dekiru* ('be able to') as in (5.1).

The third most frequent group is adjectival nouns, including (iii) adjectival noun (as a sub-category of "noun") (209 types) and (ix) adjectival noun (as an independent category) (215 types). An adjectival noun such as *shinpuru* (< *simple*) is used as an adjective taking an auxiliary such as the copula *-na* as in (5.1). This category also serves as an adverb when followed by the copula *-ni* as *summuzu* (< *smooth*) shown in (5.2). It also serves as a noun when followed by a particle as *orijinaru* (< *original*) shown in (5.3).

Those in (iv) verbal/adjectival noun (9 types) are used either as a noun, verb or adjective just as outlined above.

(5.1) 最もシンプルな方法は、ホテルの予約をキャンセルすることだ

*Mottomo shinpuru-na hoohoo-wa hotel-no yoyaku-o
kyanseru-suru koto-da.*

most simple-COP way-TOP hotel-GEN reservation-ACC
cancel-do thing-COP

'The simplest solution is to cancel your hotel reservation.'

(5.2) ホテルの予約のキャンセルがスムーズにできない

Hoteru-no yoyaku-no kyanseru-ga sumuuzu-ni deki-nai

hotel-GEN reservation-GEN cancel-NOM smooth-COP be.able.to-NEG

'I cannot cancel my hotel reservation smoothly (I have difficulty in canceling my hotel reservation).'

(5.3) オリジナルの曲

original-no *kyoku*

original-GEN song

‘The original song’

In summary, the major usages of loanwords are divided into three groups: nominal, verbal and adjectival usages. In the following section, then, the focus is put on these three usages.³⁵

5.3.2. Identifying loanwords that are high in frequency

As Labov says, one of the criteria for a useful variable is being high in frequency (1966: 49). The first step is therefore to identify loanwords that are high in frequency. Word frequency lists were made based on the fixed-length component of PB and PM in the Publication sub-corpus, BCCWJ.

Based on the usages of loanwords outlined in 5.3.1, three frequency lists were made by usage: nominal, verbal and adjectival usage.³⁶ The criteria for extracting tokens of each usage are shown in Table 5.4. The most frequent loanwords in each usage are listed in Appendices B1a-c: Appendix B1a includes those with a raw frequency of 100 and more in nominal usage (92 words); Appendix B1b includes those with a frequency of 30 and more in verbal usage (33 words), and; Appendix B1c includes those with a

³⁵ Note that Sino-Japanese words (SJ) are integrated into Japanese in the same way as the Western loanwords as outlined above. Although LW and SJ do not have lexical items categorized as “verbs” nor “adjectives” in Table 5.3, they do serve as verbs and adjectives when relevant derivational suffixes are attached, competing with NJ verbs/adjectives that are similar in meaning.

³⁶ Note that some loanwords have different native alternatives depending on its usage. For example, *imeeji* (< *image*) is a near-synonym of *inshoo*, *kanji* or *gazoo* (‘impression, feeling, image’) in its nominal usage while its verbal usage *imeeji-suru* is synonymous with the verb *omoiukaberu* (‘to imagine’). That is why it is important to make frequency lists by usage.

frequency of 30 and more in adjectival usage (12 words).³⁷ Appendix B1b is repeated here as Table 5.5 as an example.

Table 5.4 Criteria for making frequency lists of loanwords

Usage	preceding context	key	following context
Nominal	a word that ends in <i>hiragana</i> writing ³⁸	(i) noun (ii) verbal noun (iii) adjectival noun (in “noun”) (iv) verbal/adjectival noun (vi) counter noun (vii) numeral noun (ix) adjectival noun	-particle -auxiliary (copula in most cases)
Verbal		(ii) verbal noun (iv) verbal/adjectival noun	- <i>suru</i> (‘do’) - <i>dekiru</i> (‘be able to’)
Adjectival		(iii) adjectival noun (in “noun”) (iv) verbal/adjectival noun (ix) adjectival noun	-auxiliary (copula in most cases) ³⁹

The numbers “(i)~(ix)” correspond to those in Table 5.3

³⁷ Note that verbal and adjectival usages are relatively low in frequency compared with those of nouns.

³⁸ The scope of investigation is limited to the case where a preceding word ends with hiragana. This is to effectively eliminate the cases where the loanwords in question are used as part of compounds or nonce words (*rinji ichigo*). Note that words ending with *kanji* or *katakana* in writing are in most cases nouns and form compounds or nonce words with the loanwords following them, e.g., 脳死ケース *nooshi-keesu* (brain.death-case, ‘the case of a brain death’).

³⁹ Note that samples extracted with these criteria are not completely exclusive to each other. For example, a context like “[adjectival noun] + [auxiliary]” does not distinguish adjectival usages (e.g. *orijinaru-na/da* ‘original-COP’) from nominal usages (e.g. *orijinaru-no* ‘original-COP’).

Table 5.5 Frequency list (verbal usage)

	lemmaID	lemma_JP	lemma	meaning	frequency	Is there native single-word equivalents? (Yes=1)
1	10494	クリック	<i>kurikku</i>	click	898	
2	23405	チェック	<i>chekku</i>	check	267	
3	19233	スタート	<i>sutaato</i>	start	145	1
4	2503	イメージ	<i>imeeji</i>	image	130	1
5	13232	コントロール	<i>kontoororu</i>	control	114	
6	6986	カバー	<i>kabaa</i>	cover	92	
7	116693	ドラッグ	<i>doraggu</i>	drag	89	
8	795	アップ	<i>appu</i>	up	85	
9	458	アクセス	<i>akusesu</i>	access	84	
10	20229	セット	<i>setto</i>	set	80	
11	25641	デザイン	<i>dezain</i>	design	73	
12	939	アピール	<i>apiiru</i>	appeal	71	
13	14371	サポート	<i>sapooto</i>	support	69	1
14	6756	カット	<i>katto</i>	cut	66	
15	5554	オープン	<i>oopun</i>	open	62	1
16	10471	クリア	<i>kuria</i>	clear	58	
17	40345	リラックス	<i>rirakkusu</i>	relax	58	1
18	2758	インストール	<i>insutooru</i>	install	57	
19	33702	プラス	<i>purasu</i>	plus	55	
20	40443	リード	<i>riido</i>	lead	48	
21	12939	コピー	<i>kopii</i>	copy	47	
22	23716	チャレンジ	<i>charenji</i>	challenge	47	1
23	33749	プレゼント	<i>purezento</i>	present	47	
24	33757	プレー	<i>puree</i>	play	44	
25	35712	マスター	<i>masutaa</i>	master	44	1
26	23101	ダウンロード	<i>daunroodo</i>	download	42	
27	9743	キープ	<i>kiipu</i>	keep	40	1
28	15683	シフト	<i>shifuto</i>	shift	40	1
29	8884	キャッチ	<i>kyatchi</i>	catch	39	
30	40352	リリース	<i>riiisu</i>	release	39	
31	35818	マッチ	<i>matchi</i>	match	36	
32	1284	アレンジ	<i>arenji</i>	arrange	35	
33	878	アドバイス	<i>adobaisu</i>	advice	30	1

5.3.3. Identifying loanwords that have native variants

In the next step, among those loanwords listed in Appendices B1a-c, those which seem to have clear, interchangeable native *single-word* equivalents have been selected by the intuition of the author, a native speaker of Japanese.⁴⁰

First, cultural borrowings do not fall under the scope of the present study. For example, *kurikku-suru* (< *click* ‘to push/press a button on a computer mouse’) is a textbook example of cultural borrowing, since it was borrowed to fill a lexical gap for computer-related terms. *Akusesu-suru* (< *access* ‘to gain access to information held in a computer, to approach something or a place’) is another example of cultural borrowing, as it is specifically related to computer technology. The definitions of these loanwords are usually given in the form of a *sentence* in a dictionary and so it is generally difficult to rephrase them with existing native *single words*.

Second, abstract words are mainly selected in the case of nominal usage. Kim (2011) points out that loan nouns that are on the increase in the latter half of the 20th century belong mainly to abstract nouns.

Consequently, 29 loanwords, including 16 in nominal usage, 10 in verbal usage and 3 in adjectival usage, have been selected as possible loan variants that seem to be competing with their native *single-word* equivalents. They correspond to those marked with “1” in Appendices B1a-c and Table 5.5.

⁴⁰ Although using one’s intuition is subjective, this procedure is essential to effectively reduce linguistic variables to a reasonable number for further investigations. The final decision on native alternatives of the selected loanwords will be made by a more objective and rigorous method in 5.3.4.

5.3.4. Identifying native equivalents

The last step is to identify native variants of the selected 29 loanwords. Synonymy among the lexical variants is determined based on dictionaries. I consulted eight dictionaries (listed below) to find definitions of these loanwords. Only those definitions that are given in the form of single-word expressions and appear in at least four out of the eight dictionaries are considered native variants of a given loanword in this chapter. This is because it is likely that single-word expressions that consistently appear in different dictionaries could potentially serve as native variants of a given loanword.

If a given loanword has even one native variant that is defined by a phrase or sentence, it is excluded from further analysis. That is, only those of which *all* their native variants are single-word expressions are further investigated. As a result, 4 out of the 29 loanwords have been excluded and the 25 loanwords listed in Table 5.6 have been selected for further investigation.

Dictionaries consulted with:

1. *Gendai Yogo-no Kiso-chishiki Katakana Gairaigo/Ruigo Jiten* (4th Edition, 1st Printing) (2011)
2. *Sanseido Kokugo Jiten* (7th Edition, 1st Printing) (2014)
3. *Shinmeikai Kokugo Jiten* (7th Edition) (2011) (web)⁴¹
4. *Concise Katanago Jiten* (4th Edition) (2010) (web)
5. *Reikai Shin Kokugo Jiten* (8th Edition, 1st Printing) (2012)
6. *Kojien* (6th Edition) (2008)
7. *Nihon Kokugo Daijiten* (2nd Edition) (2000-2002)
8. *Iwanami Kokugo Jiten* (7th Edition, 1st Printing) (2009)

⁴¹ Web dictionaries (3 and 4) are available at: <<http://www.sanseido.net/>> (accessed between 4-16 May 2015 and 7-17 June 2016).

Table 5.6 Selected lexical variables (BCCWJ)

ID	lemma_JP	lemma	meaning (etymology)	usage	native equivalents defined by the dictionary-based approach
1	アドバイス	<i>adobaisu</i>	advice	V	<i>jogen, chuukoku, kankoku</i>
2	バランス	<i>baransu</i>	balance	N	<i>kinkoo, tsuriai</i>
3	ケース	<i>keesu</i>	case	N	<i>baai, jirei</i>
4	チャレンジ	<i>charenji</i>	challenge	V	<i>choosen</i>
5	チャンス	<i>chansu</i>	chance	N	<i>kikai, kooki</i>
6	イベント	<i>ibento</i>	event	N	<i>gyooji, moyooshimono, dekgoto</i>
7	イメージ	<i>imeeji</i>	image	V	<i>omoiukaberu, omoiukabu</i> ^{*42}
8	キープ	<i>kiipu</i>	keep	V	<i>iji, kakuho</i>
9	レベル	<i>reberu</i>	level	N	<i>suijun, hyoojun</i>
10	マスター	<i>masutaa</i>	master	V	<i>jukutatsu</i>
11	メリット	<i>meritto</i>	merit	N	<i>riten, choosho, kooseki</i>
12	ニーズ	<i>niizu</i>	needs	N	<i>yookyu, juyoo</i>
13	オープン	<i>oopun</i>	open	V	<i>kaiten, kaigyo, hiraku, hirakeru, *aku, akeru*</i>
14	プロセス	<i>purosesu</i>	process	N	<i>katei, keika, tejun, hoohoo</i>
15	リラックス	<i>rirakkusu</i>	relax	V	<i>kutsurogu</i>
16	リスク	<i>risuku</i>	risk	N	<i>kiken</i>
17	ルール	<i>ruuru</i>	rule	N	<i>kisoku</i>
18	シフト	<i>shifuto</i>	shift	V	<i>ikoo</i>
19	シンプル	<i>shinpuru</i>	simple	A	<i>tanjun, kanso</i>
20	スムーズ	<i>sumuuzu</i>	smooth	A	<i>nameraka</i>
21	スタート	<i>sutaato</i>	start	V	<i>syuppatsu, kaishi, hajimeru, hajimaru</i>
22	サポート	<i>sapooto</i>	support	V	<i>shien, sasaeru, shiji</i>
23	テーマ	<i>teema</i>	theme	N	<i>shudai, daimoku</i>
24	タイプ	<i>taipu</i>	type	N	<i>ruikei, kata</i>
25	ユニーク	<i>yuniiku</i>	unique	A	<i>dokutoku, dokuji</i>

Alphabetically ordered by “meaning (etymology)”

⁴² Note that a LW/SJ verbal noun like *imeeji* (‘(to) image’) serves as both a *transitive* verb taking *suru* (‘do’) and an *intransitive* one taking *dekiru* (‘be able to’) or *sa-reru* (‘do-PASS’), whereas the NJ verbs have different inventories for each usage: *omoiukaberu* (tr)/*omoiukabu* (intr). Therefore, in some cases, when a dictionary shows only an NJ transitive verb, i.e., *omoiukaberu*, as a definition of a given loanword, the intransitive NJ counterpart, i.e., *omoiukabu* is automatically added to the list of native equivalents and are marked with an asterisk (*) in the table.

5.3.5. Scope of analysis

As outlined in 3.3, this chapter will undertake a large-scale “comprehensive approach,” in which we shall consider all the contexts where the linguistic variants in question occur. Although this approach is less fine-grained, it gives an analysis of large amount of data, which will be useful for us to gain a quick overview of a general trend in the synchronic distribution of loanwords.

In this approach, *weak complementarity* in the resultant synchronic distribution (Sankoff, D. 1982: 683-4, Tagliamonte 2006: 74), rather than the *meaning* per se, is regarded as evidence of “equivalence” among competing forms. In other words, competing forms are considered to be equivalent if they are distributed complementarily along social dimensions (e.g. age) as conceptualized in Figure 5.3.

Extending this approach to the current data, it is hypothesized that there should be complementary distributions across social dimensions if the selected loanwords and their native equivalents are in competition.

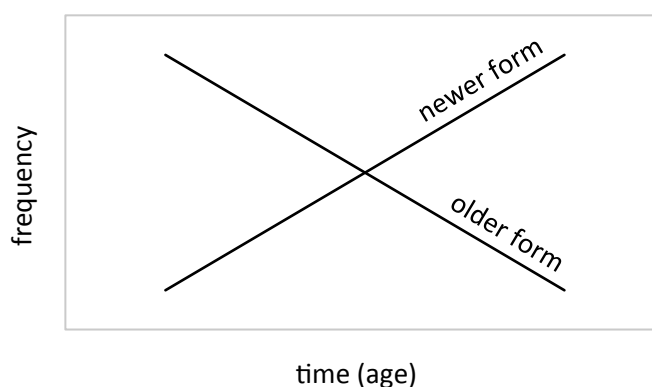


Figure 5.3 Weak complementarity of distribution over time

5.4. Analysis of the written corpus (BCCWJ)

5.4.1. Eliciting data

From the variable-length component of PB and PM in the Publication sub-corpus, tokens of the 25 loanwords and their native alternatives in Table 5.6 have been elicited based on the following criteria (which is the same as in Table 5.4). Investigation is limited to samples written by a single author. The number of tokens of each variable is shown in Table 5.7 (75,163 tokens in total). In the following sections we will observe the distribution of these loanwords in social and stylistic contexts.

N: a word that ends with *hiragana* + target lemma + particle/auxiliary

V: a word that ends with *hiragana* + target lemma + *suru* ('do') / *dekiru* ('be able to')⁴³

A: a word that ends with *hiragana* + target lemma + auxiliary

⁴³ If it is a NS verb, e.g. *sasaeru* ('to support'), the following context, i.e., + *suru* ('do')/*dekiru* ('be able to'), is disregarded.

Table 5.7 Distribution of LWs and non-LWs (BCCWJ)

	lemmaJP	lemma	meaning (etymology)	usage	LW	non-LW	total	LW%
1	アドバイス	<i>adobaisu</i>	advice	V	73	125	198	36.9
2	バランス	<i>baransu</i>	balance	N	920	186	1106	83.2
3	ケース	<i>keesu</i>	case	N	1627	21796	23423	6.9
4	チャレンジ	<i>charenji</i>	challenge	V	149	390	539	27.6
5	チャンス	<i>chansu</i>	chance	N	572	1508	2080	27.5
6	イベント	<i>ibento</i>	event	N	324	1035	1359	23.8
7	イメージ	<i>imeeji</i>	image	N	402	357	759	53.0
8	キープ	<i>kiipu</i>	keep	V	124	2274	2398	5.2
9	レベル	<i>reberu</i>	level	N	856	572	1428	59.9
10	マスター	<i>masutaa</i>	master	V	134	18	152	88.2
11	メリット	<i>meritto</i>	merit	N	450	594	1044	43.1
12	ニーズ	<i>niizu</i>	needs	N	489	1016	1505	32.5
13	オープン	<i>oopun</i>	open	V	162	7093	7255	2.2
14	プロセス	<i>purosesu</i>	process	N	523	7072	7595	6.9
15	リラックス	<i>rirakkusu</i>	relax	V	216	164	380	56.8
16	リスク	<i>risuku</i>	risk	N	606	1242	1848	32.8
17	ルール	<i>ruuru</i>	rule	N	501	230	731	68.5
18	シフト	<i>shifuto</i>	shift	V	102	432	534	19.1
19	シンプル	<i>shinpuru</i>	simple	A	450	594	1044	43.1
20	スムーズ	<i>sumuuzu</i>	smooth	A	359	190	549	65.4
21	スタート	<i>sutaato</i>	start	V	431	12706	13137	3.3
22	サポート	<i>sapooto</i>	support	V	251	2475	2726	9.2
23	テーマ	<i>teema</i>	theme	N	946	192	1138	83.1
24	タイプ	<i>taipu</i>	type	N	1260	492	1752	71.9
25	ユニーク	<i>yuniiku</i>	unique	A	187	296	483	38.7
total number of token					12114	63049	75163	

Alphabetically ordered by “meaning (etymology)”.

5.4.2. Independent variables

This section examines four language-external factors: the author’s a) Year of Birth, b) Sex, c) Education and d) Register (i.e. Medium Type: books (PB) vs. magazines (PM)).

As mentioned in 3.1.2, I identified educational history of the author using data from the questionnaire survey of authors conducted separately by NINJAL between 2006 and 2010 with special permission from NINJAL to use this additional information.

5.4.3. Year of Birth

Appendix B2a shows the distribution of the LW and non-LW variants by the author's year of birth. To reduce the effect of irregular distributions due to small-sized samples, especially from the youngest and oldest generations, I have chosen to focus on those born between the 1920s and 1960s, each of which, on average, contribute more than 5% of the whole sample (45,554 tokens). Note that the sample size for generations other than these groups, those born in the 1910s and before and those born in the 1970s and after, is very small (less than 5% of the whole sample, see the bottom row labeled "total (%)" in Appendix B2a).⁴⁴

By limiting the scope of investigation to these generations, a general tendency becomes more visible: the occurrence of LW increases as authors become younger. A regression analysis is conducted in order to verify the upward trend in the occurrence of LW in apparent time. The statistical significance of age-effect (at the 5 % level) is confirmed for 10 loanwords as shown in Figure 5.4. Each loan variant is presented by its etymology, e.g., *case*, instead of its Japanese writing *keesu*. The figure demonstrates that the occurrence of these loanwords is not independent of the author's year of birth, i.e., there are changes in apparent time toward more frequent use of loanwords as opposed to their native lexical variants. The weak complementarities in their distributions provide supporting evidence that the loanwords are in competition with the selected native forms. The rate of change, however, is different across variables. For example, *keesu* (< *case*), *oopun* (< *open*) and *purosesu* (< *process*) have relatively

⁴⁴ We have seen in 5.2.2 that we should interpret irregular distributions of these groups with care. For example, *keesu* (< *case*) shows irregular distributions among those born in the 1910s and before as well as those born in the 1970s and after (Appendix B2a).

gentle slopes, while the rest of the loanwords show relatively steeper slopes. Figure 5.5 and Figure 5.6 show examples of a regression model of age-distribution of LW and its goodness of fit to the data (R^2) (p refers to the result of an t -test). See Appendix B2b for the distribution and result of the regression analysis for each linguistic variable.

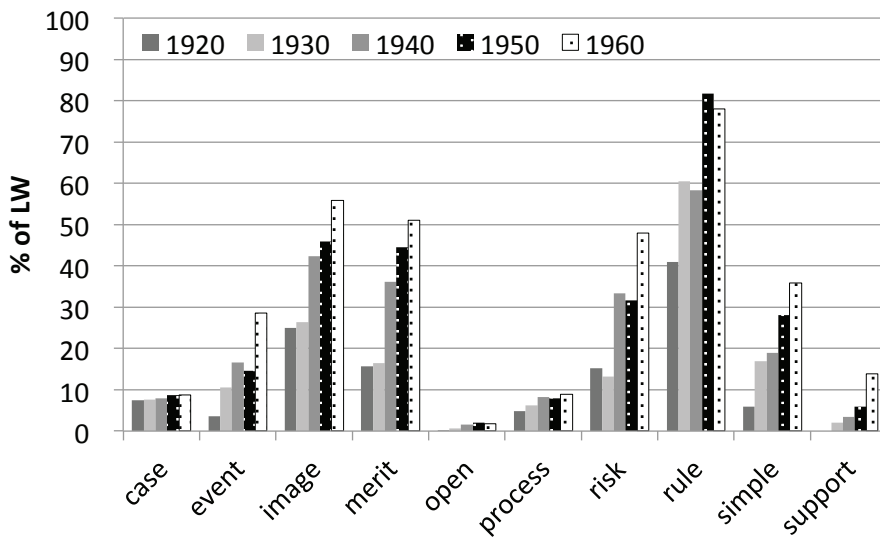


Figure 5.4 Effect of Year of Birth on the occurrence of LWs (BCCWJ)

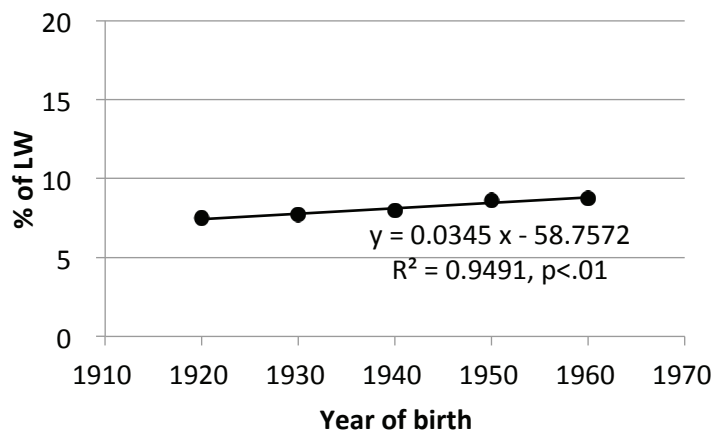


Figure 5.5 Regression analysis of *keesu* (< case)

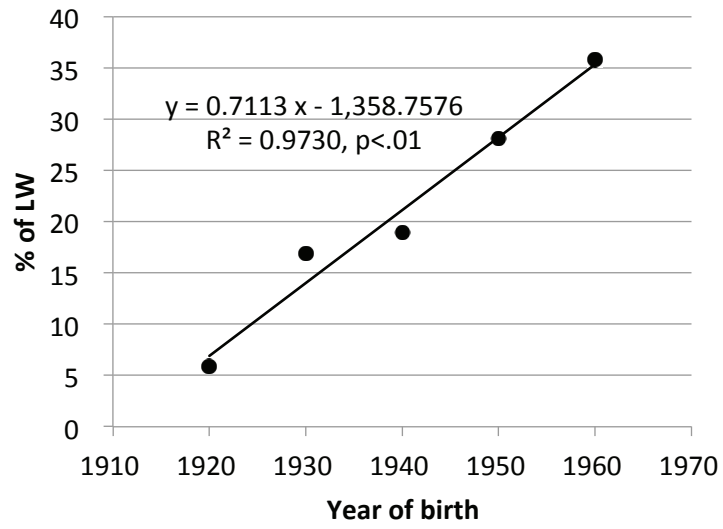


Figure 5.6 Regression analysis of *shinpuru* (< simple)

5.4.4. Sex

The statistical significance of Sex is evaluated by the chi-square test. The difference between men and women proved to be significant at the 5% level for 6 loanwords as shown in Figure 5.7. For a full list of the frequencies and results of the chi-square test for individual linguistic variables, see Appendix B3. Among these, 4 variables show a tendency for female authors to use a higher proportion of the loanword than male authors. The remaining two variables, marked with an asterisk (*), i.e., *rirakkusu* (< *relax*) and *risuku* (< *risk*), shows an opposite tendency. We should note that we cannot simply conclude that women are ahead of men in their use of LW as a general tendency, considering that such a tendency is observed only in 4 out of the 25 variables.

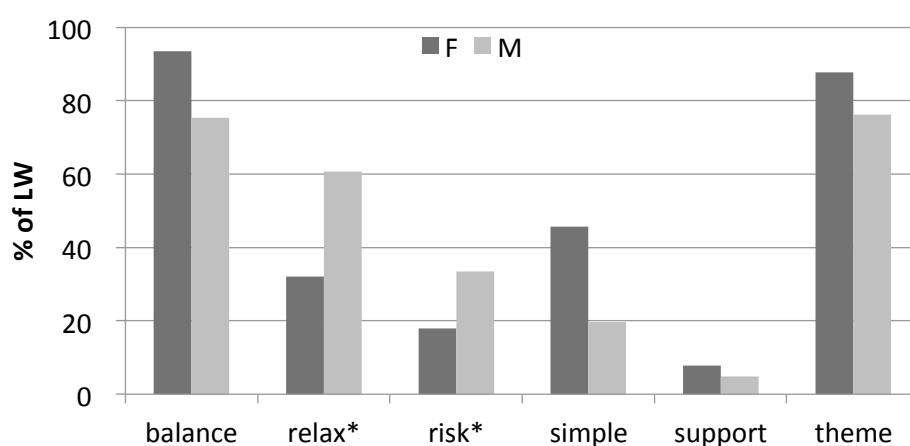


Figure 5.7 Effect of Sex on the occurrence of LWs (BCCWJ)

5.4.5. Education

Figure 5.8 shows the distribution of individual loanwords by educational background of the authors.⁴⁵ There are 14 loanwords for which the difference between those with postgraduate education and others proved to be significant at the 5% level by the chi-square test. For a full list of the frequencies and results of the chi-square test for individual linguistic variables, see Appendix B4. It should be noted that the majority of these (11 out of 14), are favored by those whose highest obtained education is university (up to UNIV) than graduate school educated authors (GRAD). Those marked with an asterisk (*), i.e., *imeeji* (< *image*), *risuku* (< *risk*) and *shifuto* (< *shift*), however, show an opposite tendency.

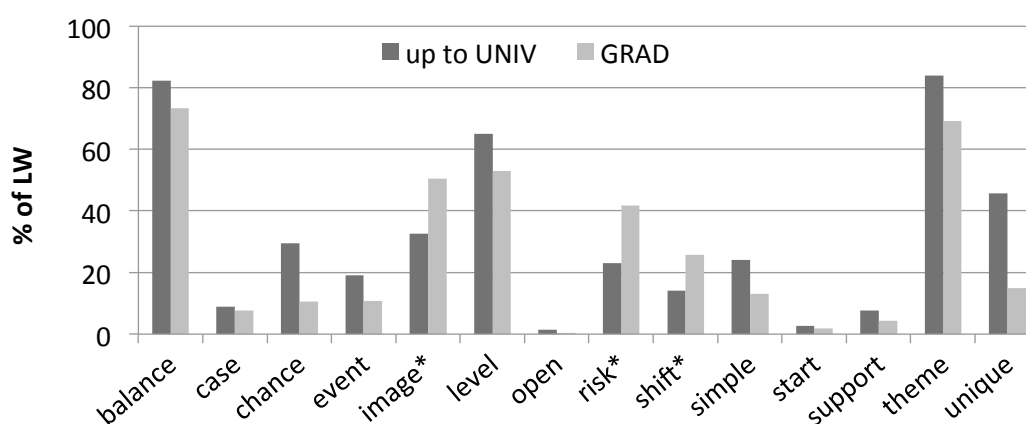


Figure 5.8 Effect of Education on the occurrence of LWs (BCCWJ)

The figure exhibits a general tendency for those with a university education to use loanwords more frequently than those with a postgraduate education. The reason for the exceptional patterns, i.e., the reason why some loanwords are favored by those with a

⁴⁵ Note that the educational variable is binary here as the majority (more than 90%) of samples come from either university or postgraduate school educated authors.

postgraduate education should be investigated further by consulting the use of individual words in context. For example, there may be some cases in which these loanwords are favored in academic/professional contexts.

5.4.6. Register (Medium Type)

Figure 5.9 shows the distribution of individual loanwords by register. The difference between books (PB) and magazines (PM) proved to be significant at the 5% level by the chi-square test for 9 variables. For a full list of the frequencies and results of the chi-square test for individual linguistic variables, see Appendix B5. The figure shows that the occurrence of these loanwords is not independent of the effect of register, with the occurrence of LWs consistently favored in PB than in PM.

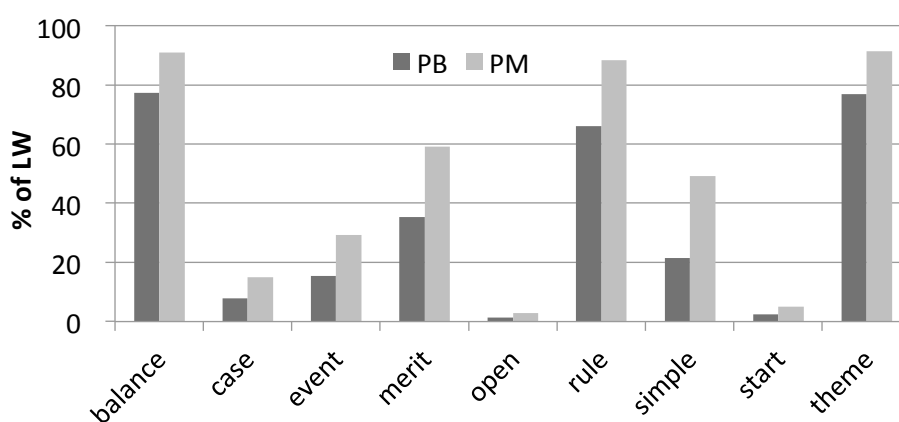


Figure 5.9 Effect of Register on the occurrence of LWs (BCCWJ)

Now the question arises as to what this difference implies. One way of exploring this is to examine the correlations between register and the other factors. Table 5.8 shows (Pearson's) correlations between every pair of independent variables considered in this section. It does not reveal a very high correlation between any pairs.

Table 5.8 Correlation matrix (BCCWJ)

	Year of Birth	Sex	Education	Register
Year of Birth	1	.149	.098	.117
Sex (M=0, F=1)	.149	1	-.059	.061
Education (UNIV=0, GRAD=1)	.098	-.059	1	-.119
Register (PB=0, PM=1)	.117	.061	-.119	1

Provided that Register is not largely dependent on the other independent variables, we should think of other possible factors that might induce the register difference.

First, it is possible that the formality of a text differentiates the two registers. Tanaka (2012) points out that, among registers contained in BCCWJ, magazines and online text are characterized as having attributes such as “private,” “daily,” “entertaining” and “intragroup,” while books and newspapers as having properties such as “public” and “social.” It is then predictable that the fact of more frequent LW use in magazines than in books is an indication of a stronger association of the loan variants, as opposed to their native variants, with less public or formal domains of writing.

It is doubtful, however, whether the difference in the type of media, i.e., books or magazines, is directly interpretable as a difference in the degree of “formality.” It may be necessary to consider empirical evidence of a stylistic difference between these media. To test the claim that the degree of formality of magazines is relatively lower than that of books, I took a close look at the occurrence of some phonologically contracted copulas and particles across these media. Since contracted forms are considered to be non-standard or colloquial, a heavier use of them can be identified as a marker of a relatively lower degree of formality of text.

Examined forms include the following: (i) *ja* (copula), (ii) *tte* (particle), (iii) *tte* (auxiliary)⁴⁶ and (iv) *n* (particle). They are contracted (therefore, colloquial) forms of (i) *de-wa* (COP-TOP ‘is/are not’), (ii) the complementizer *to* (‘that’), (iii) *toiu* (‘what they say’) and (iv) the particle *no*, respectively. Examples are presented in (5.4) – (5.7).

(5.4) *ja_copula*

a. *Kore-wa inu ja nai.* (contracted)

b. *Kore-wa inu de-wa nai.*

This-TOP dog COP-TOP NEG

‘This is not a dog.’

(5.5) *tte_particle*

a. *Inu-wa kawaii tte omoi-masu.* (contracted/colloquial)

b. *Inu-wa kawaii to omoi-masu.*

god-TOP cute COMP think-COP

‘I think that dogs are cute.’

(5.6) *tte_auxiliary*

a. *Inu tte no-wa hoeru kara kirai-da.* (contracted)

b. *Inu toiu no-wa hoeru kara kirai-da*

dog what.they.say.about COMP-wa bark because hate-COP

‘As for dogs, I do not like them since they bark.’

(5.7) *n_particle*

a. *Inu-wa kawaii to omou n desu.* (contracted)

⁴⁶ UniDic distinguishes the two *tte* forms: (i) the one called “particle” (Example (5.5)), which is the colloquial form of the complementizer *to* followed by a verb, such as *iu* (‘say’), *omou* (‘think’), and so on, and (ii) the one called “auxiliary” (Example (5.6)), which is the merger of *toiu* (‘what they say’) followed by a noun or the particle *no*.

- b. *Inu-wa kawaii to omou no desu.*
 dog-TOP cute COMP think COMP COP
 ‘I do think that dogs are cute.’

The distribution of the four contracted forms (per million word) is shown in Figure 5.10.⁴⁷ The figure demonstrates that there is a consistent tendency for the contracted forms to occur more frequently in PM than in PB. Now it is more persuasive to say that magazines can be characterized as a medium that adopts a more colloquial or less formal style than books. So, in turn, the lower proportion of loanwords in PB than in PM implies a tendency for the loan variants to be disfavored in the more formal (or less colloquial) medium.

In summary, the difference between the two registers is interpreted, in one possibility, as the difference in the stylistic factor of formality, as represented by the presence/absence of contracted forms. In the next section, the effect of style will be further investigated with different parameters.

⁴⁷ I used the variable component of the Publication sub-corpus. The criteria for eliciting these forms are as follows:

ja_copula: *ja* + *nai* (NEG),
tte_particle: *tte* + {*iu* (‘say’), *omou* (‘think’), *kiku* (‘hear’), *kangaeru* (‘think’)},
tte_auxiliary: *tte* + {*koto* (‘thing’), *no* (particle), *kanji* (‘feeling’), *wake* (‘reason’), *yatsu* (‘thing, fellow’), *mono* (‘thing’), *tokoro* (‘place’)}
n_particle: *n* + {*da* (COP), *desu* (COP.POL)}

As quotations tend to promote the occurrence of these colloquial forms, tokens occurring in quotations are excluded by using the two tags: “quotation” and “quotes” (see 6.2 for more detail about these tags). This allows us to eliminate the effect of direct and indirect speech by people other than the author.

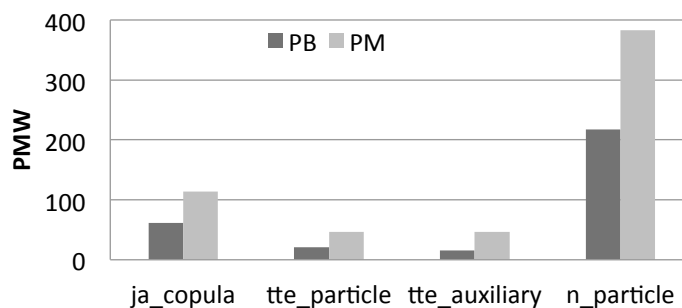


Figure 5.10 Phonologically contracted forms by Register

5.5. Analysis of the spoken corpus (CSJ)

5.5.1. Eliciting data

Following the discussion above, this section puts a special focus on the effect of stylistic factors on the occurrence of loanwords. CSJ provides useful information on styles based on detailed evaluation of several factors such as formality and spontaneity.

Following the procedures presented in Table 5.4, frequencies of loanwords in APS and SPS were counted by their usage to make frequency lists. Appendices B6a-c show those that occur more than 100 times in nominal usage (113 words), more than 50 times in verbal usage (11 words) and more than 50 times in adjectival usage (4 words). Among those included in Appendices B6a-c, only those that are also listed in Table 5.6, i.e., those investigated with BCCWJ in the previous section, are left for our attention. Consequently, the 14 loanwords shown in Table 5.9, including 9 in nominal usage, 3 in verbal usage and 2 in adjectival usage, have been selected for further investigation (they correspond to those marked with “1” in Appendices B6a-c). 16,894 tokens in total have been extracted.

Table 5.9 Selected lexical variables (CSJ)

	lemma_JP	lemma	meaning (etymology)	usage	LW	non-LW	total	LW%
1	バランス	<i>baransu</i>	balance	N	110	1	111	99.1
2	ケース	<i>keesu</i>	case	N	185	7717	7902	2.3
3	イベント	<i>ibento</i>	event	N	159	290	449	35.4
4	イメージ	<i>imeeji</i>	image	V	68	69	137	49.6
5	レベル	<i>reberu</i>	level	N	446	68	514	86.8
6	メリット	<i>meritto</i>	merit	N	152	186	338	45.0
7	プロセス	<i>purosesu</i>	process	N	139	2433	2572	5.4
8	リラックス	<i>rirakkusu</i>	relax	V	61	5	66	92.4
9	ルール	<i>ruuru</i>	rule	N	458	259	717	63.9
10	シンプル	<i>shinpru</i>	simple	A	52	389	441	11.8
11	スムーズ	<i>sumuuzu</i>	smooth	A	54	40	94	57.4
12	スタート	<i>sutaato</i>	start	V	82	2288	2370	3.5
13	テーマ	<i>teema</i>	theme	N	522	145	667	78.3
14	タイプ	<i>taipu</i>	type	N	390	126	516	75.6
total number of token					2878	14016	16894	

Alphabetically ordered by “meaning (etymology)”

5.5.2. Independent variables

In the following, besides the four variables considered in the previous section, i.e., a) Year of Birth, b) Sex, c) Education, of the authors and d) Register (Speech Type: APS vs. SPS), three stylistic variables are incorporated into analysis: e) Spontaneity, f) Formality and g) Audience Size of a talk.

5.5.3. Year of Birth

For the same reason mentioned in 5.4.3, age groups that do not account for more than 5% of the whole sample are not included in the analysis (see Appendix B7a for more detail). As a result, only those born between the 1940s and 1970s are included in the scope of investigation here (15,823 tokens). Figure 5.11 shows the distribution of the 14 loanwords by the speaker's year of birth.

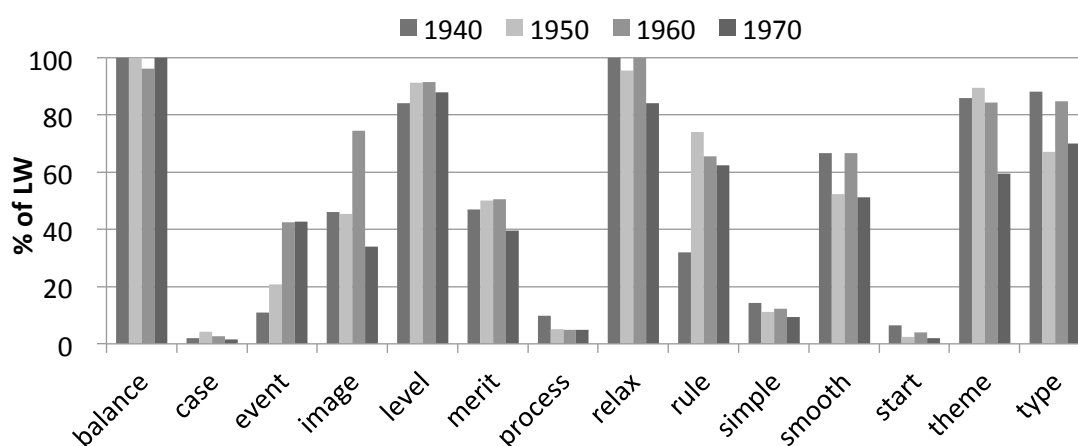


Figure 5.11 Effect of Year of Birth on the occurrence of LWs (CSJ)

The statistical significance of birth-year effect is evaluated by the regression analysis. The distributions of the selected loan variants in CSJ do not exhibit consistent upward tendencies in apparent time except that *ibento* (< *event*) barely shows a tendency toward significance (at the 10% level in an t-test) of birth-year effect. For the distribution and result of the regression analysis for each linguistic variable, see Appendix B7b.

One of the reasons why age gradient is not visible may be attributed to the relatively small sample size of the spoken corpora. A small difference in raw frequency might

have resulted in a great difference in proportion, and in turn, irregular distributions across generations. Furthermore, it is also possible that the effect of Year of Birth is distorted by other independent variables (this point will be discussed later in 5.5.10).

5.5.4. Sex

The effect of Sex was evaluated by the chi-square test and proved to be significant at the 5% level for 6 out of the 14 loanwords as in Figure 5.12, with 4 of them being favored by female speakers while 2 of them (marked with an asterisk *) being favored by male speakers. (See Appendix B8 for a full list of the frequencies and results of the test for individual linguistic variables.) It seems that women tend to use loanwords more frequently than men, but we should again take care not to draw a sweeping conclusion from this small number of samples.

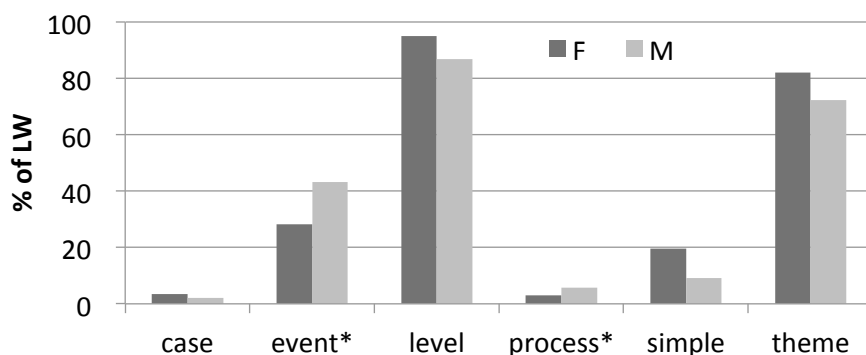


Figure 5.12 Effect of Sex on the occurrence of LWs (CSJ)

5.5.5. Education

The educational effect proved to be significant at the 5% level in the chi-square test for 5 out of the 14 variables as in Figure 5.13 (See Appendix B9 for a full list of the frequencies and results of the test for individual linguistic variables). There is a tendency for those in GRAD to disfavor the use of loan variants than those in UNIV except for the use of *purosesu* (< *process*) (marked with *). Although this tendency is not very striking, it is noteworthy that this trend, i.e., the tendency for more educated people to disfavor the use of LW, is consistent with what we have seen in analysis of BCCWJ in 5.4.5.

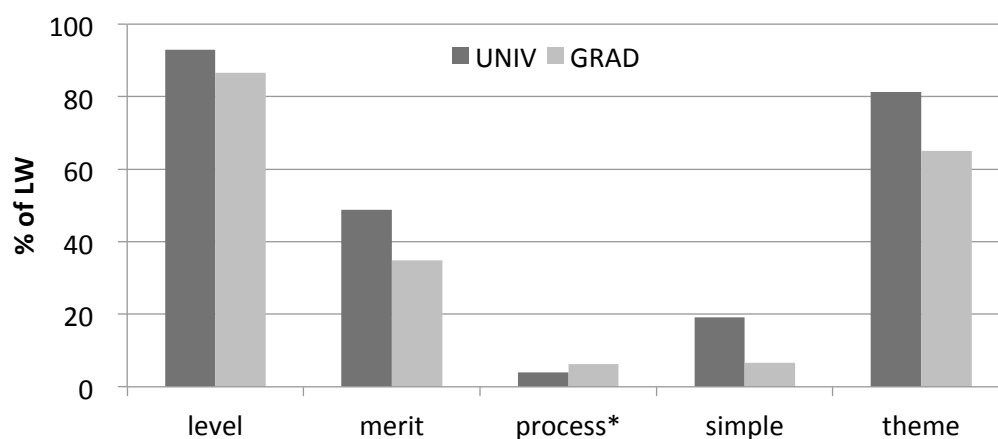


Figure 5.13 Effect of Education on the occurrence of LWs (CSJ)

5.5.6. Register (Speech Type)

Figure 5.14 shows the distribution of loanwords by register. The effect of Register was proven to be significant at the 5% level in the chi-square test for 7 out of the 14 variables, which demonstrates that the occurrence of loanwords is generally more frequent in SPS than in APS except for *purosesu* (< *process*). (See Appendix B10 for a full list of the frequencies and results of the test for individual linguistic variables.)

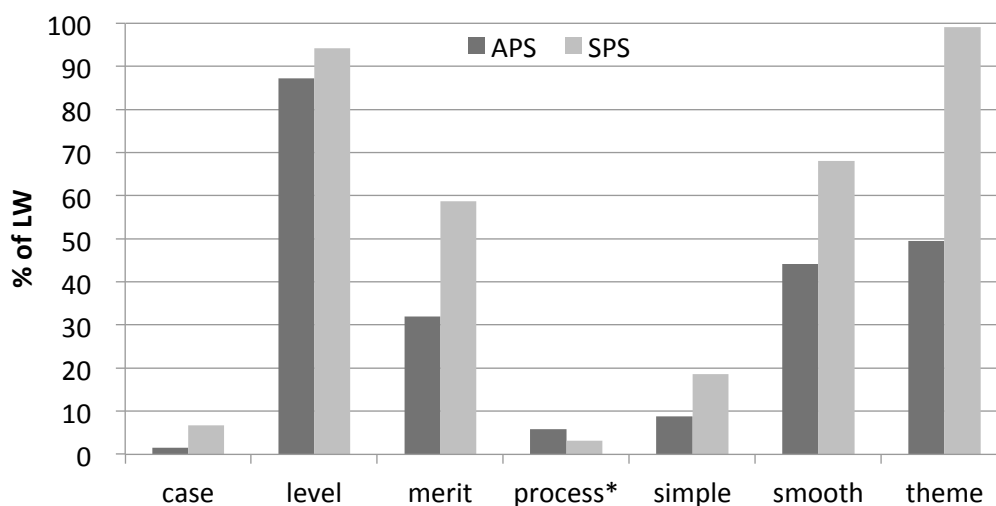


Figure 5.14 Effect of Register on the occurrence of LWs (CSJ)

5.5.7. Spontaneity

Even though there is an *a priori* generalization that APS is characterized as a more formal speech setting than SPS, it would be useful to look at factors that are more closely associated with style, e.g., the degree of “spontaneity” or “formality” of the speech, in order to confirm the effect of style on the occurrence of loanwords.

Figure 5.15 shows the occurrence of LWs by the degree of spontaneity of the talk (spontaneous vs. others).⁴⁸ The effect of Spontaneity proved to be significant at the 5% level in the chi-square test for 7 out of the 14 variables (see Appendix B11 for a full list of the frequencies and results of the test for individual linguistic variables). The figure shows that the use of LWs is more frequent in spontaneous speech except for *keesu* (< *case*) (marked with *). It implies that the loan variants are more strongly associated with spontaneous speech.

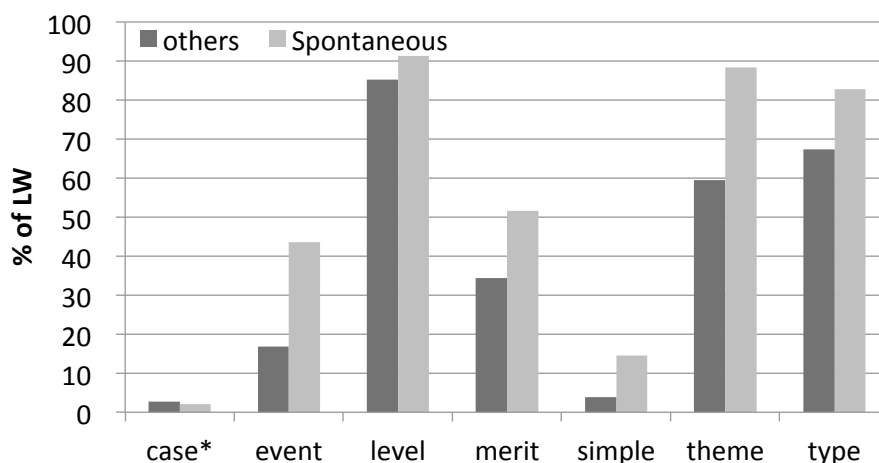


Figure 5.15 Effect of Spontaneity on the occurrence of LWs (CSJ)

⁴⁸ Spontaneity is originally evaluated in a five-point scale: (1) not spontaneous—(2) not really spontaneous—(3) neither—(4) somewhat spontaneous—(5) highly spontaneous. Here (4) and (5) are grouped together into “spontaneous” as against “others” including (1), (2) and (3).

5.5.8. Formality

Figure 5.16 shows the distribution of LWs by the degree of formality of the speech (formal vs. others).⁴⁹ The effect of Formality on the occurrence of a loan variant proved to be significant at the 5% level in the chi-square test for 7 out of the 14 variables (see Appendix B12 for a full list of the frequencies and results of the test for individual linguistic variables). It demonstrates that loanwords are disfavored in formal speech except for *purosesu* (< *process*) (marked with *). This implies that LW use is more strongly associated with informal speech.

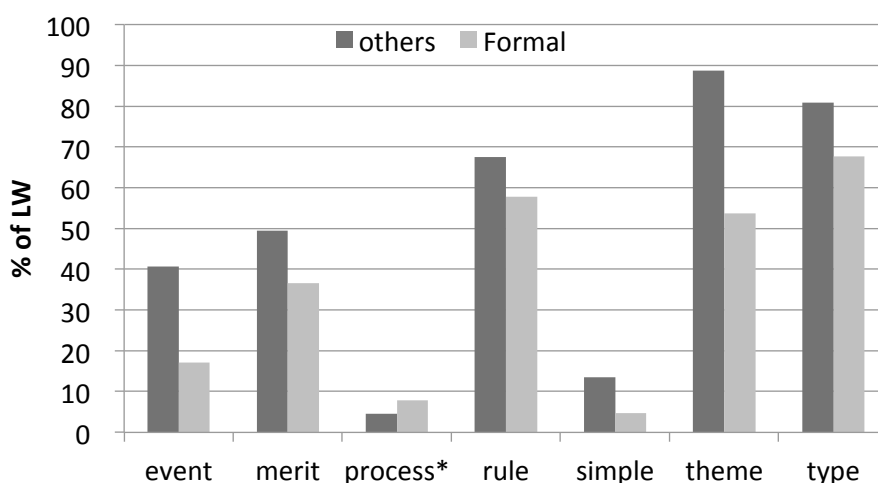


Figure 5.16 Effect of Formality on the occurrence of LWs (CSJ)

⁴⁹ Formality is originally evaluated in a five-point scale: (1) casual—(2) somewhat casual—(3) neither—(4) somewhat formal—(5) formal. Here (4) and (5) are grouped together into “formal” as against “others” including (1), (2) and (3).

5.5.9. Audience Size

Figure 5.17 shows the distribution of LWs by audience size. The effect of Audience Size proved to be significant at the 5% level in the chi-square test for 8 out of the 14 variables (see Appendix B13 for a full list of the frequencies and results of the test for individual linguistic variables). Although the distribution is not always hierarchically ordered from small audience to large audience, it is often the case that the loan variants are least frequent in front of the largest audience, except for *purosesu* (< *process*) (marked with *). It implies that LWs are less frequently used in front of a larger audience.

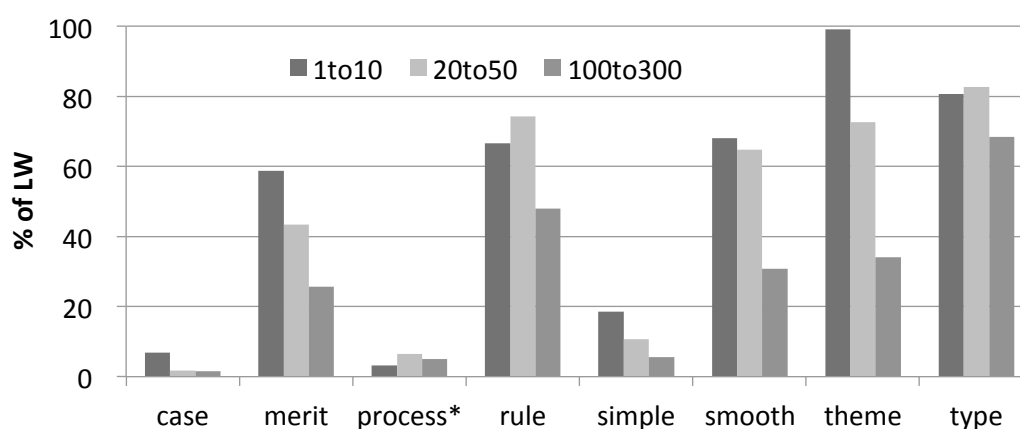


Figure 5.17 Effect of Audience Size on the occurrence of LWs (CSJ)

It is not difficult to imagine that audience size tends to affect the speaker's attention to the formality level of the setting, or in other words, the speaker's attention to speech. A larger audience would mean that the speaker is required to show more consideration to the various linguistic backgrounds of the addressees, such as age, for instance. For a similar discussion on the relationship between audience size and the occurrence of loanwords, see Chapter 7. It examines the stylistic distinction between (i) the situation

of having a chat with friends and (ii) that of public speaking with a large audience, and shows that the latter setting discourages some speakers from using LW.

5.5.10. Correlation among independent variables

Although data from CSJ has showcased the roles of different stylistic variables, it should be noted that they are correlated with each other. Table 5.10 shows a correlation matrix of the seven independent variables considered in this section. It is notable that many pairs of independent variables are moderately/strongly correlated.

Table 5.10 Correlation matrix (CSJ)

	Register	Year of Birth	Sex	Education	Spontaneity	Formality	Audience Size
Register (SPS=0, APS=1)	1	.135	-.355	.612	-.180	.221	.862
Year of Birth	.135	1	-.033	-.028	-.174	.040	.154
Sex (M=0, F=1)	-.355	-.033	1	-.210	-.079	.019*	-.306
Education (UNIV=0, GRAD=1)	.612	-.028	-.210	1	-.039	.143	.519
Spontaneity (others=0, high=1)	-.180	-.174	-.079	-.039	1	-.261	-.161
Formality (others=0, high=1)	.221	.040	.019	.143	-.261	1	.264
Audience Size (1-10=1, 20-50 =2, 100-300=3)	.862	.154	-.306	.519	-.161	.264	1

Most noticeably, Register or Audience Size is moderately/strongly correlated with many factors. First, Register has a very strong correlation with Audience Size ($r=0.862$), where the academic presentation speech (APS) predicts a larger audience size. Second, both Register and Audience Size show moderate/strong correlations with Education, where more educated speakers are associated with APS ($r=0.612$) and speech with a larger audience ($r=0.519$). Third, these two factors (Register or Audience Size) also

show weak/moderate correlations with Sex, where female speakers are associated with SPS ($r=-0.355$) and speech with a smaller audience ($r=-0.306$). Lastly, Register and Audience Size have weak/moderate correlations with Formality and Spontaneity, where APS is associated with more formal speech ($r=0.221$) and with less spontaneous speech ($r=-0.180$), and speech with a larger audience is associated with more formal speech ($r=0.264$) and with less spontaneous speech ($r=-0.161$). It can also be noted that Formality and Spontaneity are weakly correlated with each other ($r=0.261$).

These observations imply that factors such as Audience Size, Education, Sex, Formality and Spontaneity affect the phenomenon as subcategories of Register (for similar discussions on correlations among external factors in CSJ, see Sano 2009: 136). That is, register differentiation is attributed to these factors. The fact that SPS shows higher proportions of LWs than APS is attributed to the predominance of a smaller audience, less educated speakers, female speakers, and less formal/more spontaneous speech in SPS.

As for Year of Birth, it only shows a very weak correlation with Register ($r=0.135$), Spontaneity ($r=-0.174$) and Audience Size ($r=0.154$), where younger speakers are associated with APS, less spontaneous speech and a larger audience, all of which generally disfavor the use of LWs. This explains to some degree that why age gradient is not observed in CSJ.

5.6. Discussion and conclusion

This chapter has confirmed that the vocabulary in the loanword stratum as a whole is on the increase in apparent time in the Japanese lexicon in both written and spoken modes.

After identifying potential lexical variables, individual variables are further examined with regard to synchronic distribution by generation as well as sex, education register and several stylistic parameters. Table 5.11 shows the summary of the factor-by-factor analyses, where the effect of an external factor proved to be significant is marked either with ** ($p<0.01$), * ($p<0.05$), + ($p<0.1$), or *ns* if not significant. A result is bracketed if it shows a pattern that is different from the general tendency.

Table 5.11 Summary of the factor-by-factor analyses

BCCWJ					CSJ						
factor	Year of Birth	Sex	Education	Register	Year of Birth	Sex	Education	Register	Spontaneity	Formality	Audience Size
general tendency	younger < older	M<F	UNIV> GRAD	PB<PM	younger < older	M<F	UNIV> GRAD	APS<SPS	others < spontaneous	others > formal	small > large
advice	+	ns	ns	-							
balance	ns	**	*	**	ns	-	-	-	-	-	-
case	**	+	*	**	ns	**	+	**	(*)	ns	**
challenge	ns	ns	+	ns							
chance	ns	ns	**	ns							
event	*	ns	**	**	+	(**)	ns	ns	**	**	+
image	**	ns	(**)	ns	ns	ns	ns	ns	ns	ns	ns
keep	ns	-	-	-							
level	ns	ns	**	ns	ns	*	*	*	*	+	ns
master	+	-	-	-							
merit	**	ns	ns	**	ns	ns	*	**	**	*	**
needs	+	ns	ns	ns							
open	*	ns	*	*							
process	*	ns	ns	+	ns	(*)	(*)	(*)	ns	(**)	(*)
relax	ns	(**)	ns	ns	ns	-	-	-	-	-	-
risk	*	(**)	(**)	ns							
rule	*	ns	ns	*	ns	ns	ns	ns	+	**	**
shift	ns	-	(*)	ns							
simple	**	**	**	**	ns	**	**	**	**	*	**
smooth	ns	ns	ns	ns	ns	ns	+	*	ns	ns	**
start	ns	ns	*	**	ns	+	+	ns	ns	ns	ns
support	*	*	*	ns							
theme	ns	**	**	**	ns	**	**	**	**	**	**
type	+	+	ns	ns	ns	ns	+	+	**	**	**
unique	+	ns	**	+							

** , * and + refers to being significant at the 1%, 5% and 10% level, respectively. () indicates that it shows tendency different from a general one.

In BCCWJ, many of variables showed a tendency for younger generations to use loanwords more often than older generations. In CSJ, this generational tendency was not visible, partly because the correlations between Year of Birth and other factors. That is, age distribution might be distorted because younger speakers are associated with factors that prevent the use of LW, such as APS, low spontaneity and larger audience (5.5.10). Another possibility is that the diffusion of loanwords is not in progress in the spoken mode, or, that it is in progress but too slow to be captured by observing just a 40-year period.

Nevertheless, CSJ exhibited the significance of the effect of other social and stylistic variables. This is consistent with what was observed in the analysis of BCCWJ.

The effect of Education proved to be significant for many lexical variables under consideration. In both BCCWJ and CSJ, those from GRAD use loanwords less frequently than those from UNIV.

The effect of Sex was not as consistent as that of the above variables. It seems that the sex differentiation is specific to the context of the words in question rather than a universal tendency.

The effect of Register was attested in both corpora. In the written corpus, I pointed out that the difference between books (PB) and magazines (PM) may possibly be due to stylistic difference; for example, the formality of a text, evaluated by the frequency of non-standard contracted forms. To examine the effect of stylistic factors by means of other parameters, CSJ has also been investigated. The spoken corpus also exhibited the effect of Register, Spontaneity, Formality and Audience Size on the occurrence of loan variants. Yet these results should be interpreted with care because various external

factors are moderately/strongly correlated with each other. In particular, the correlations between Register and (i) Audience Size, (ii) Education, (iii) Spontaneity and (iv) Formality could imply that the register difference is attributable to these factors.

In summary, many loanwords exhibited age-gradient, indicating the existence of an ongoing change toward heavier use of loanwords (BCCWJ). This means that the claim that there is an ongoing change toward more frequent use of loanwords, which had previously only been attested to by questionnaire-based studies of linguistic attitudes, has been further supported by empirical evidence from people's linguistic behavior. Besides, it has also been shown that loanwords are disfavored by more educated people and in more formal settings.

It is noteworthy that the comprehensive approach succeeded in capturing *weak complementarity* (Sankoff, D. 1982) in synchronic distribution of loan variants and their native variants not only by age but also by other social and stylistic factors. This approach, however, still has room for improvement.

One may wonder if the identification of the native variants is correct or appropriate. While the dictionary-based approach that I used to identify lexical variants is objective, it is also problematic because dictionaries put an emphasis on the denotational meaning of the synonymous words, not their usage (Xiao & McEnery 2006: 111). It is possible that weak complementarity in distribution could become more striking if the native variants of a given loanword were selected on the basis of usage. In addition, one should go a step further to answer the question of to what extent one should limit the envelope of variation based on usage.

In the next chapter I will conduct a case study of *keesu* (< *case*) based on the previous in-depth study of this loanword (Kim 2011). I will take a more contextual approach to lexical variation, in which the envelope of variation is identified in a more careful manner. This provides clues to tackle the above problems.

6. A case study of *keesu* (< case): A contextual approach

6.1. Introduction

6.1.1. Aim: Towards a more fine-grained analysis of lexical variation

This chapter attempts to make a more contextual approach to lexical variation.⁵⁰ The comprehensive approach shown in the previous chapter has several methodological limitations to be tackled from a traditional variationist point of view. Although we have succeeded in finding the general sociolinguistic patterns in loanword distribution based on a less fine-grained but large-scale analysis, we still need to give a more detailed analysis of the phenomenon by tackling these problems.

This chapter further investigates the distribution of the loanword *keesu* (< case) and its native alternatives in BCCWJ and CSJ as a case study. I will first provide a more detailed description of the selected linguistic variable by observing all the contexts where the variants that I identified appear. I will then identify the shared function of the variants (the envelope of variation) in a more careful manner. I will also introduce a multivariate model to analyze the phenomenon multidimensionally.

The reason for choosing *keesu* as a linguistic variable is four-fold. First, previous studies such as Kim (2011) and Hashimoto (2010: 52) confirm an upward tendency in the occurrence of the loanword in question in real time. This means that the diffusion of this loanword is a bona fide example of an ongoing change. Second, the loanword is high in frequency, guaranteeing the necessary volume for a quantitative analysis. Note that *keesu* is the most frequently occurring loanword among those examined in Chapter

⁵⁰ Chapter 6 is a revised version of the paper published as: Kuya, A. (2013) Gendai kakikotoba ni okeru gairaigo no kyoojiteki bunpu: *Keesu* o jirei to shite. *Kokuritsu Kokugo Kenkyuujo Ronshuu* [NINJAL Research Papers], 6.

5 (see Table 5.7). Third, the distribution pattern of this loanword is not inconsistent with the overall tendencies of the distributions of the 25 variables considered in Chapter 5. Last but most importantly, the loanword and its native equivalents have been fully investigated in terms of their usages and functions based on a corpus (Kim 2011). The last point is particularly crucial in identifying the envelope of variation. In short, *keesu* is a good example of lexical variable to be studied within the framework of variationist approach.

6.1.2. Backgrounds: Kim's study (2011) of the loanword *keesu*

In this chapter the loanword *keesu* (< *case*) is considered a new sociolinguistic lexical variant of its native equivalents: *jirei*, *baai* and *rei*.⁵¹ Kim (2011) suggests that some loan nouns, including *keesu*, have become “basic words” in spite of the existence of existing native equivalents in competition with them.⁵² Figure 6.1 shows that *keesu* first appears in newspapers in the 1960s and shows a consistent increase over the latter half of the 20th century.⁵³

⁵¹ Note that *rei* is not included in the dictionary-based approach in the previous chapter.

⁵² According to her explanation, “basic words” are defined as those being broadly and frequently used in certain registers by shifting from the periphery to the core of the lexicon of a language. She showed in her case studies that (i) *toraburu* (< *trouble*) and (ii) *keesu* (< *case*) have established their status as “basic words” in newspapers by expanding their meaning and usage.

⁵³ According to Irwin (2011: 60), the first written attestation of *keesu* is in 1872.

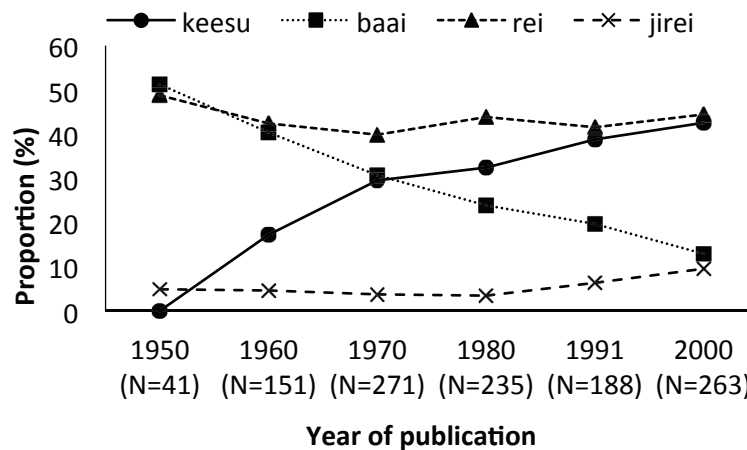


Figure 6.1 Diachronic distribution of *keesu*, *baai*, *rei* and *jirei* (newspapers) (source, Kim 2011: 107, Table 1a)⁵⁴

6.1.2.1. Shared function of the four variants

According to Kim, *keesu*, *jirei*, *rei*, and *baai* are synonymous in that they all serve as a formal noun, in particular as a head of a gap-less relative clause (see 6.3.1 for more detail) as in Examples (6.1a-d) (2011: 93). In the examples below, each variant forms a nominal phrase (NP), modified by the preceding square-bracketed gap-less relative clause.

(6.1) (Kim 2011: 93, Examples (1) – (4))

a. [NP [債務超過状態でも認められた] N ケース]がある。

[_{NP} [saimuchookajootai-de-mo mitome-rare-ta] N keesu]-ga aru

[[in.the.state.of.insolvency-COP-too admit-PASS-PAST] case]-NOM exist

‘There is [a case [in which one obtained a permission even in the state of insolvency]].’

⁵⁴ The figures are given in PMW (per million word) in Kim but I modified them by converting PMW into proportions to show the probabilities of the occurrence of a given word as opposed to the other variants.

- b. [NP[「情報教育」を掲げながら、実際はパソコンの利用法の習得どまりとなる] _N事例]もある。

[NP [Joohoo-kyooiku-o kakage-nagara jissai-wa pasokon-no riyoo-hoo-no shuutoku-domari tonaru] _Njirei]-mo aru.
 [[information-education-ACC emphasize-although actual-TOP computer-GEN use-method-GEN acquisition-mere remain] case]-too exist
 ‘There are [cases [in which students learn merely how to use a computer in spite of the emphasis in education being on promoting their comprehensive skills of making the best use of information]].’

- c. 植物状態のような慢性の意識障害では、[NP[目は開いているが話せないなど、JCSが適合しない] _N例]がある。

Shokubutsu-jootai-no yoona mansei-no ishiki-shoogai-de-wa,
[NP [me-wa ai-teiru ga hanase-nai nado,
JCS-ga tekigoo-shi-nai] _Nrei]-ga aru.
 vegetative-state-GEN such.as chronic-GEN conscious-disturbance-in-TOP
 [[eyes-NOM open-PROG but can.speak-NEG such.as JCS(Japan Coma Scale)-NOM be.applicable.to-do-NEG] case]-NOM exist
 ‘In the example of chronic disturbance of consciousness such as a vegetative state, there are [cases [in which the patient cannot speak a word despite the fact that his/her eyes are open]].’

- d. 水汲みは子供たちの仕事。[NP[数キロ離れた所からやってくる] _N場合]もある

Mizukumi-wa kodomotachi-no shigoto.
[NP [Suu-kiro hanareta tokoro kara yattekuru] _Nbaai]-mo aru.
 Fetching.water-TOP children-GEN routine.
 [[Several-kilometers distant place from come] case]-too exist.
 ‘Fetching water is a daily routine for children. There are [cases [in which they come from several kilometers away to do this]].’

Based on Kim’s observation, we will regard this function (as the formal noun ‘case’) as one that the four variants typically share. In contrast, the context in which these variants are used in other functions does not fall within the scope of further investigation. For example, as Kim (2011: 94-95) points out, environments where the variants have idiomatic or word-specific meanings are to be excluded from the analysis. First, unlike the other three variants, *baai*, if not followed by a particle, can be used to refer to (i) ‘as for,’ simply raising the topic of the sentence as in (6.2), or (ii) ‘if/when,’ adding a hypothetical subordinate clause to the sentence as in (6.3). Second, the four variants occur in their own idiomatic usages. For example, *baai* refers to ‘time’ (6.4); *rei* refers to ‘in question’ or ‘that’ as a demonstrative determiner (6.5) and ‘as usual’ (6.6); and *keesu* refers to ‘container’ as a concrete noun (6.7). These usages are distinguished from the usage as the general formal noun ‘case.’ In addition, cases where these words serve as a component of a compound or a nonce word (*rinji ichigo*) as in (6.8) are also excluded from the analysis. Examples (6.2) – (6.7) were taken from BCCWJ.

(6.2) Topicalizing (‘as for’)
 船橋警察署の場合、(PB13_00095)
Funabashi keisatsu-sho-no baai,
 Funabashi police-office-GEN case
 ‘As for Funabashi police office,’

(6.3) Hypothetical subordinate clause ('if/when')

対馬でタクシーを利用する場合は、島が大きくそのまま利用すると非常に交通費が高くなってしまうため、(PM51_00890)

Tsushima-de takushii-o riyoo-suru baai-wa,

shima-ga ookiku sonomama riyoo-suru-to

hijoo-ni kootsuuhi-ga takaku-natteshimau tame,

Tsushima-in taxi-ACC use-do case-TOP,

island-NOM big without.consideration use-do-if

very-COP travel.expenses-NOM expensive-result.in because,

'If/when you use a taxi on Tsushima Island, traveling expenses can get very high if you keep riding without thinking since the island is so big.'

(6.4) 'time'

だが、今はそんなことを言っている場合ではない。(PB43_00366)

Daga, ima-wa sonna koto-o itte-iru baai-de-wa-nai

but, now-TOP such thing-ACC say-PROG case-COP-TOP-NEG

'It is not the time to be talking about a thing like that.'

(6.5) Demonstrative determiner ('in question/that')

そして、例の十字路に着いた。(PB22_00242)

Soshite, rei-no juujiro-ni tsui-ta

then in.question-GEN crossroad-at arrive-PAST

'Then I got to that crossroads (the crossroads in question)'

(6.6) 'as usual'

ところが、例によってエレベーター前のドアには鍵がかかっている。

(PB52_00020)

Tokoroga, rei-niyotte erebeetaa mae-no doa-ni-wa kagi-ga kakat-teiru

however, as.usual lift front-GEN door-on-TOP key-NOM locked-PROG

'However, as usual, the door in front of the lift is locked.'

(6.7) Concrete noun (‘container’)

革製のケース

kawa-sei-no keesu

lether-made.of-GEN container

‘A container made of leather’

(6.8) Part of a compound

モデルケース、 具体的事例、 典型例

moderu keesu, gutaiteki jirei, tenkei rei

model case, specific case, typical case

‘model case,’ ‘specific example (case),’ ‘typical example (case)’

6.1.2.2. Co-occurring predicates

Kim (2011: 115) points out that *keesu* co-occurs with the predicates listed in Table 6.1 (particularly with *ooi* (‘many’) and *aru* (‘present’)). Noticing that these predicates assign the nominative particle *ga* to the noun *keesu*, I will focus on the contexts where *keesu* and other variants are used in a subject position followed by these predicates. Examples (6.9) – (6.14) from BCCWJ provide sentences with each predicate.

Table 6.1 Type of predicate that co-occurs with *keesu*
(source, Kim 2011: 115⁵⁵)

Type of predicate	examples
[present/absent]	<i>aru</i> ('present'), <i>mi-rareru/miuke-rareru</i> ('can be observed'), <i>nai</i> ('absent'), <i>mitome-rare-nai</i> ('cannot be observed'), etc.
[many/few]	<i>ooi/medatsu/hotondo-da/attooteki-da/ooku mirareru</i> ('many'), <i>ippanteki-da</i> ('(the most) common'), <i>sukunai</i> ('few'), <i>mezurashii/hotondo nai</i> ('rare'), etc.
[increase/decrease]	<i>hueru/zooka-suru</i> ('increase'), <i>heru</i> ('decrease'), etc.
[occur]	<i>okiru/shoojiru/hassei-suru</i> ('occur'), etc.
[assume]	<i>sootei-sareru</i> ('be assumed'), <i>kangae-rareru</i> ('be predicted'), <i>yosoo-sareru</i> ('be expected'), etc.
[report]	<i>age-rareru</i> ('be given as an example'), <i>hooji-rareru</i> ('be reported'), etc.

(6.9) present/absent

手順 7 の写真の方が自然なケースもありますので、(PB10_00093)

Tejun 7-no Shashi-no hoo-ga shizen-na keesu- mo

[present]ari-masu *node*,

procedure 7-GEN photo-GEN this.way-NOM natural-COP case-too
exist-POL so

'There is a case in which the photo taken in the procedure 7 looks natural, so...'

(6.10) many/few

また回線料も現状の専用線や VAN と比べると安くできるケースが多い。

(PB10_00063)

Mata kaisen-ryoo-mo genjoo-no sen'yoo-sen ya VAN to

kuraberu to yasuku dekiru keesu-ga [many]ooi.

also connection-fee-too present.state-GEN special-line or VAN with
compare if cheap can.do case-NOM many

'There are many cases in which the connection fee is cheaper than that of an existing line or VAN.'

⁵⁵ Note that mainly those examples that assign the nominative particle to the subject are selected from the source.

(6.11) increase/decrease

最近は地方銀行が「株主優遇定期作成優待券」を贈呈したり、地元の名産品を贈るケースも増えています。(PB23_00254)

Saikin-wa chihoo-ginkoo-ga "kabunushi yuuguu teiki sakusei yuutaiken"-o zootei-shi tari, jimoto-no meisanhin-o okuru keesu-mo [increase]hue-tei-masu.

nowadays local-bank-NOM stockholder preferential.treatment periodical preparation discount.coupons-ACC give-do or local-GEN specialty-ACC give case-too increase-PROG-POL

‘Nowadays the cases in which local banks present “stockholder discount coupons” or specialties of the district (to stockholders) are increasing.’

(6.12) occur

最近では、こうした危機管理以前の「安全管理」に問題あるケースが続発している。(PB13_00434)

Saikin-de-wa, koushita kiki-kanri izen-no "anzen-kanri"-ni mondai aru keesu-ga [occur]zokuhatsu-shi-teiru.

nowadays-in-TOP this risk-management rather.than-GEN safety-manegement-in problem exist case-NOM occur.in.succession-do-PROG

‘Nowadays cases occur in succession in which there is a problem of “safety management” rather than risk management.’

(6.13) assume

銀行で借り直すことになるケースも考えられる (PB13_00381)

Ginkou-de kari-naosu koto ninaru keeus-mo [assume]kangae-rareru.

bank-in ask.for.loan-again thing result.in case-too expect-PASS

‘It is expected that (they) have to ask a bank for a loan again.’

(6.14) report

ネグレクトが疑われるというケースも報告されている。(PB13_00249)

Negurekuto-ga utagaw-areru toiu keesu-mo [report]*hookokus-are-teiru.*

neglect-NOM suspect-PASS COMP case-too report-PASS-PROG

‘The cases in which there is some suspicion of neglect (of children) are also reported.’

6.1.3. Envelope of variation

Based on Kim’s in-depth investigation of the loanword *keesu* and its native equivalents as shown above, the envelope of variation is limited to the contexts as listed in Table 6.2, in which the four variants serve as formal nouns in subject positions, followed by either the nominative case marker *ga* (NOM), its topicalized variation *wa* (TOP) or its emphasized version *mo* (‘too’), and then followed by the predicates that have the semantic features of [many/few, present/absent, occur, increase/decrease, assume, reported].

Table 6.2 Envelope of variation

Function	serves as a formal noun
Grammer	serves as a subject with the NOM marker <i>ga</i> , <i>wa</i> (TOP) or <i>mo</i> (‘too’)
Predicates	present/absent, many/few, increase/decrease, occur, assume, report
Subject	{ <i>keesu, jirei, baai, rei</i> } + Particle { <i>ga, ha, mo</i> } + Predicacte {present/absent, many/few, increase/decrease, occur, assume, report}

6.2. Eliciting data (BCCWJ)

The language-internal criteria for eliciting data have now been identified. Now, tokens that satisfy the criteria in Table 6.2 will be extracted from the variable-length component of the Publication sub-corpus in BCCWJ.

But before doing so, I would also like to draw attention to two issues in using BCCWJ as a source of synchronic data, in order to make our analysis more fine-grained.

First, there is sometimes a discrepancy between the publication date of a book and its first printed date. Generally, unlike magazines or newspapers, books published at a certain point include those reprinted from their previous, original publications. That is, there is a risk that language originally written and published, say, in the 1980s is treated as language written in the 2000s. Moreover, it is not necessarily the case that a given book was firstly and exclusively written for publication as a book, i.e., it is possible that it was formerly written and first made public elsewhere, e.g., in magazines, newspapers, etc., but later published as a book. In other words, in the Publication sub-corpus there is always a certain proportion of text from reprinted books or from articles that were originally written for other types of media. NINJAL made the maximum effort to trace and provide the first publication date of texts and to mark samples when they know that the text in question was originally made public before or originally written for other types of media (Maruyama 2012). To reduce the effect of these types of text, I will consider only those that were first published during the period 2001-2005. If clearly indicated, those that were originally written for other sources of media are also excluded.

Second, quotations should be excluded from the analysis because language that appears in such environments cannot be treated as the language of the author. BCCWJ marks such components by the tags “quotation” and “quote” (Yamaguchi et al. 2011: 117-124). These tags allow us to exclude the direct and indirect speech of people other than the author. By the same token, translated articles are also excluded.

The resultant number of tokens is shown in Table 6.3: there are 1,405 in total, including 312 of *keesu* (22.2%), 76 of *jirei* (5.4%), 776 of *baai* (55.2%) and 241 of *rei* (17.2%). We first notice that *keesu* accounts for approximately 20% of the total tokens; in contrast, it accounts for only 8% in the comprehensive approach taken in the previous chapter. This is mainly due to the exclusion of a great number of examples of *baai* when it is used in the topicalizing or hypothetical usages shown in Examples (6.2) and (6.3).

Table 6.3 Tokens of the lexical variants (BCCWJ) ⁵⁶

lemma	N	%
<i>keesu</i>	312	22.2%
<i>jirei</i>	76	5.4%
<i>baai</i>	776	55.2%
<i>rei</i>	241	17.2%
total	1405	100.0%

6.3. Internal variables

6.3.1. Usage

Kim (2011: 101) demonstrated that *keesu* is often modified by a gap-less relative clause as shown earlier in Example (6.1). She adds that *keesu*, however, is also used independently (without a modifier) as in (6.15). It can otherwise be preceded by a non-clause modifying component; for example, a nominal phrase marked by the genitive case *no* (6.16a), an adjectival phrase (6.16b), and an adnominal (6.16c). In Figures 6.2 – 6.4, one would notice that the occurrence of *keesu* in the usage with a

⁵⁶ Articles written by a single author for whom birth year, sex and educational history are all identified are counted.

clause modifier (Figure 6.4) exhibits a clearer and sharper upward tendency in real time, compared with that in the remaining two usages (Figures 6.2 and 6.3).⁵⁷

(6.15) (Kim 2011: 99, Example (15))

男女の賠償額にケースによっては 1000 万円近い差が生じている。

danjo-no baishoo-gaku-ni keesu-niyotte-wa

1000-man-yen chikai sa-ga shooji-teiru.

men.and.women-GEN amount.of.compensation-in case-depending.on-TOP

1000-ten.thousand-JPY nearly difference-NOM occur-PROG

‘However, as the average wage for women is lower than that for men, the difference in the amount of compensation could be as much as 10 million yen in some cases.’

(6.16) (Kim 2011: 99)

a. 京都市のケース

Kyoto-shi-no keesu

Kyoto-city-GEN kesu

‘the case of Kyoto’

b. 悪質なケース

akushitsu-na keesu

vicious-COP case

‘a vicious case’

c. そんなケース

sonna keesu

that case

‘such a case’

⁵⁷ The figures are given in PMW (per million word) in Kim but I modified them by converting PMW into proportions to show the probabilities of the occurrence of a given word as opposed to the other variants.

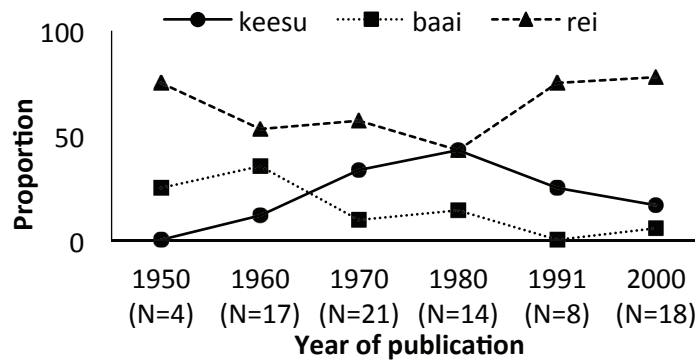


Figure 6.2 Diachronic distribution of *keesu* (without a modifier) (source, Kim 2011: 112, Table 5)

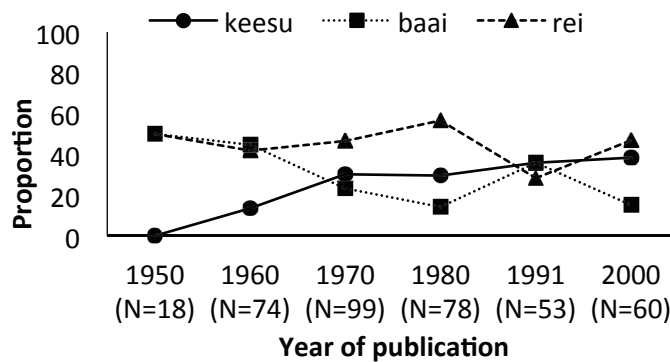


Figure 6.3 Diachronic distribution of *keesu* (with a non-clause modifier) (source, Kim 2011: 112, Table 7)

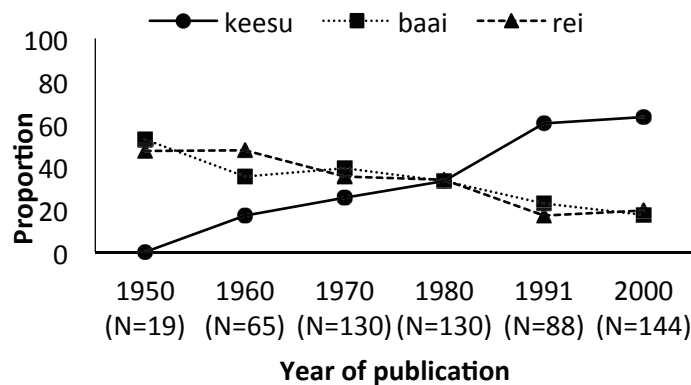


Figure 6.4 Diachronic distribution of *keesu* (with a clause modifier) (source, Kim 2011: 112, Table 8)

Furthermore, Kim emphasizes that the majority, i.e., 225 out of 233, of *keesu* used with a clause modifier serve as a head of a gap-less relative clause (ibid: 101). This is clearly distinguished from an ordinary relative clause construction in that the noun *keesu* does not fill the grammatical gap in the preceding relative clause. In (6.17a), *keesu* fills the object position of the preceding bracketed clause. While in (6.1a), repeated here as (6.17b), there is no syntactic gap for *keesu* to fill in the preceding bracketed relative clause. We call the latter use of *keesu* the “gap-less relative clause head” (henceforth, gap-less RCH).

(6.17)

a. 参照すべきケース

[_{NP} [*sanshoo-su beki*] _N *keesu*]

[[refer.to-do should] case]

‘[A case_i [that we should refer to _{t_i]]’}

b. 債務超過状態でも認められたケースがある。

[_{NP} [*saimuchookajootai-de-mo mitome-rare-ta*] _N *keesu*]-*ga aru*

[[in.the.state.of.insolvency-COP-too admit-PASS-PAST] case]-NOM exist

‘There is [a case [in which one obtained permission even in the state of insolvency]].’

Based on Kim’s observation, the present data shown in Table 6.3 will be distinguished by usage: Gap-less RCH or others. Figure 6.5 shows the proportion of *keesu* and its native equivalents by usage. It is most noticeable that *keesu* is more frequently used as a gap-less RCH (23.1%) than others (15.5%). *Baai* also show a similar distribution whereas *jirei* and *rei* shows the opposite tendency. Table 6.4

confirms the statistical significance of the effect of usage on the occurrence of the LW ($X^2 = 4.695$, $df = 1$, $p < 0.05$).

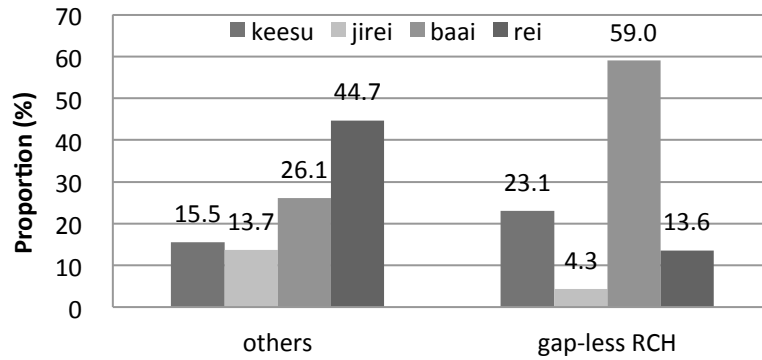


Figure 6.5 Distribution of the variants by usage (BCCWJ)

Table 6.4 Effect of usage on the occurrence of *keesu* (BCCWJ)

		others	gap-less RCH	total
LW	N	25	287	312
	%	15.5%	23.1%	22.2%
non-LW	N	136	957	1093
	%	84.5%	76.9%	77.8%
Total	N	161	1244	1405
	%	100.0%	100.0%	100.0%

$X^2 = 4.695$, $d.f. = 1$, $p < 0.05$

6.3.2. Co-occurring Predicate

Figure 6.6 demonstrates that *keesu* shows a relatively high proportion of use with the predicates [increase/decrease] and [occur]. *Baai* shows a relatively high proportion of use with [present/absent], [many/few] and [assume], whereas *jirei* and *rei* often co-occur with [report]. Table 6.5 indicates that the occurrence of the LW is not independent of the type of co-occurring predicate ($X^2 = 81.832$, $df = 5$, $p < 0.001$).

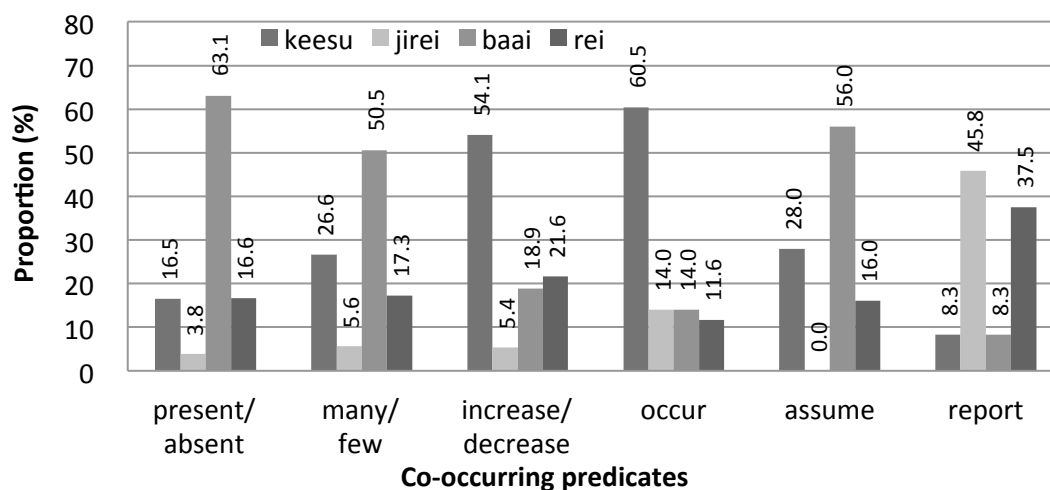


Figure 6.6 Distribution of the variants by co-occurring predicate (BCCWJ)

Table 6.5 Effect of co-occurring predicate on the occurrence of *keesu* (BCCWJ)

		present/ absent	many/ few	increase/ decrease	occur	assume	report	total
LW	N	134	123	20	26	7	2	312
	%	16.5%	26.6%	54.1%	60.5%	28.0%	8.3%	22.2%
non-LW	N	679	340	17	17	18	22	1093
	%	83.5%	73.4%	45.9%	39.5%	72.0%	91.7%	77.8%
total	N	813	463	37	43	25	24	1405
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

$X^2 = 81.832$, d.f. = 5, $p < 0.001$

6.4. External variables

6.4.1. Year of Birth

In light of the existence of an ongoing change in favor of the newer lexical variant *keesu* in the latter half of the 20th century as reported by Kim (2011), it is expected that the proportion of the loan variant systematically increases as speaker's generation become younger.

The hypothesis proved to be true as shown in Figure 6.7: the proportion of *keesu* increases from the oldest birth-year group (-1930, 19.7%) to the youngest birth-year group (1960-70, 29.2%). We notice, in contrast, that *rei* shows a systematic decrease, being surpassed largely by *keesu* among the youngest authors. *Baai* retains the highest proportion, while *jirei* shows the lowest proportion throughout generations without striking tendencies. Table 6.6 confirms the statistical significance of the effect of year of birth on the preference for the LW ($X^2=11.320$, d.f.=2, $p<0.01$).

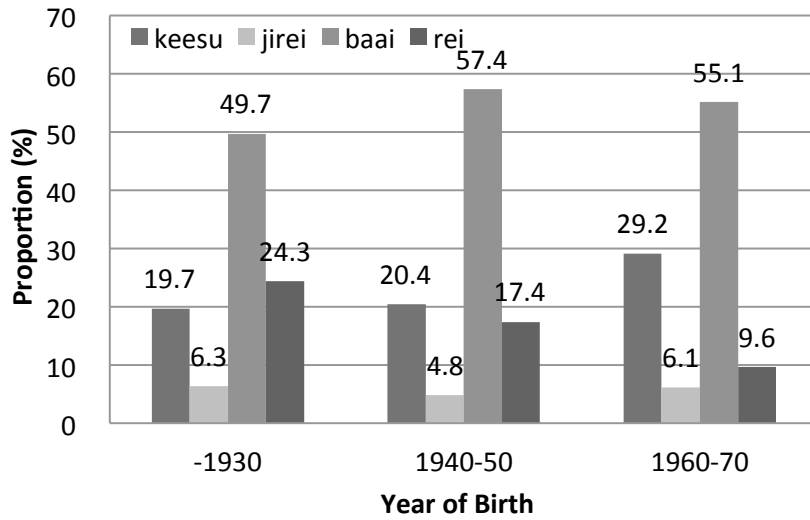


Figure 6.7 Distribution of the variants by year of birth (BCCWJ)

Table 6.6 Effect of year of birth on the occurrence of *keesu* (BCCWJ)

		-1930	1940-50	1960-70	total
LW	N	59	162	91	312
	%	19.7%	20.4%	29.2%	22.2%
non-LW	N	241	631	221	1093
	%	80.3%	79.6%	70.8%	77.8%
total	N	300	793	312	1405
	%	100.0%	100.0%	100.0%	100.0%

$X^2=11.320$, d.f.=2, $p<0.01$

6.4.2. Sex

Figure 6.8 shows that women (22.9%) are ahead of men (22.1%) in the use of *keesu*, but it does not seem to be a striking difference. Table 6.7 shows, as expected, the difference is not statistically significant ($X^2=0.045$, d.f.=1, $p=0.833$).

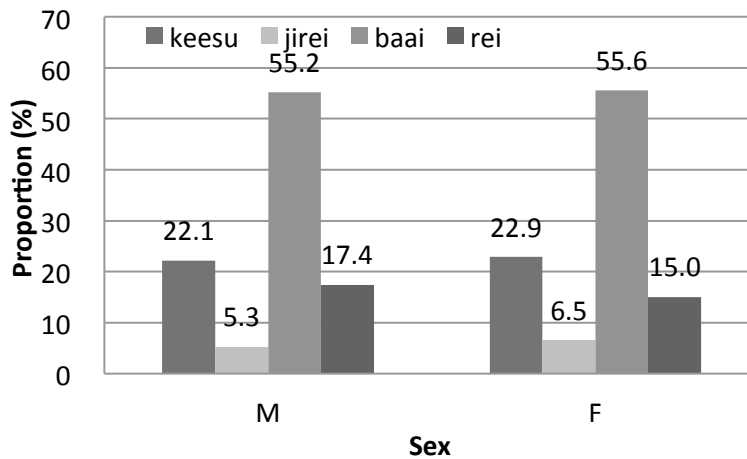


Figure 6.8 Distribution of the variants by sex (BCCWJ)

Table 6.7 Effect of sex on the occurrence of *keesu* (BCCWJ)

		M	F	total
LW	N	277	35	312
	%	22.1%	22.9%	22.2%
non-LW	N	975	118	1093
	%	77.9%	77.1%	77.8%
Total	N	1252	153	1405
	%	100.0%	100.0%	100.0%

$X^2 = 0.045$, d.f. = 1, $p = 0.833$

6.4.3. Education

Figure 6.9 shows that the proportion of *keesu* systematically decreases from left to right: from those whose highest obtained education is high school (HS, 33.8%), then university educated authors (UNIV, 25.2%) to postgraduate school educated authors (GRAD, 18.1%). *Baai*, in contrast, decreases leftwards consistently. It implies that there is a tendency for the more educated people to refrain from using the LW. Table 6.8 confirms the significance of the effect of education on the use of the loanword ($X^2=15.412$, d.f.=2, $p<0.001$).

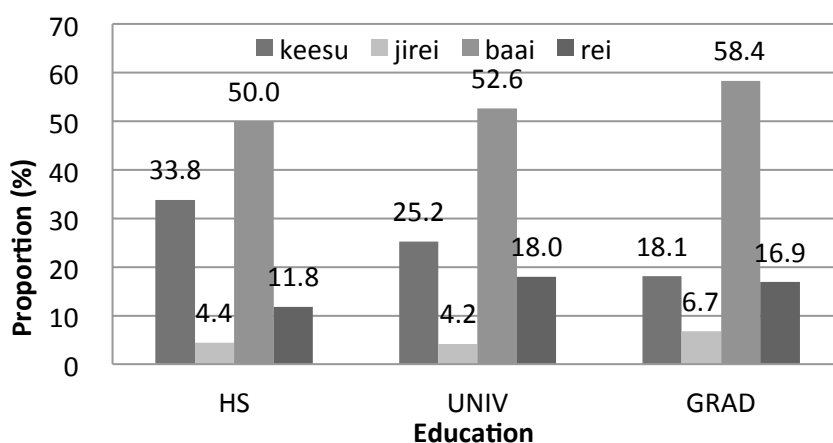


Figure 6.9 Distribution of the variants by education (BCCWJ)

Table 6.8 Effect of education on the occurrence of *keesu* (BCCWJ)

		HS	UNIV	GRAD	total
LW	N	23	168	121	312
	%	33.8%	25.2%	18.1%	22.2%
non-LW	N	45	499	549	1093
	%	66.2%	74.8%	81.9%	77.8%
Total	N	68	667	670	1405
	%	100.0%	100.0%	100.0%	100.0%

$X^2 = 15.412$, d.f. = 2, $p < 0.001$

6.4.4. Register (Medium Type)

Figure 6.10 shows that the proportion of *keesu* is rather higher in magazines (PM) (35.6%) than in books (PB) (21.5%). Table 6.9 shows that the effect of register (type of medium) on the occurrence of the loanword is significant ($X^2 = 8.016$, d.f. = 1, $p < 0.01$).

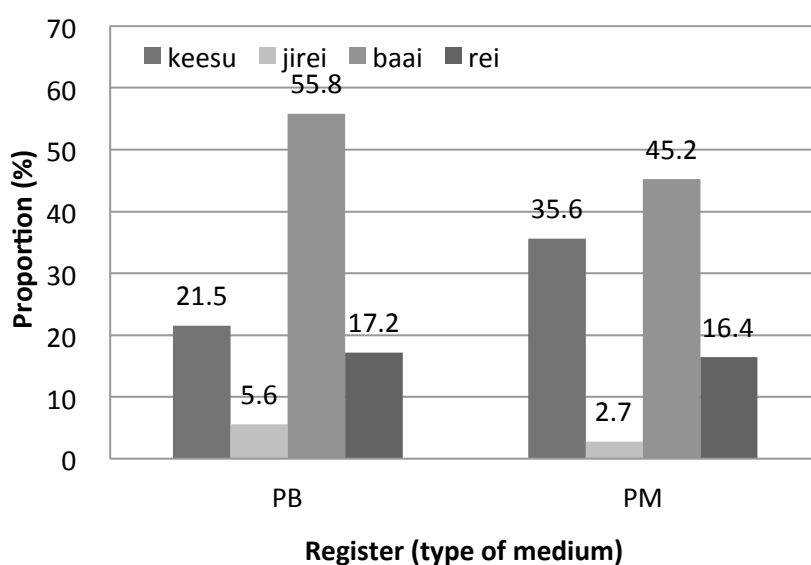


Figure 6.10 Distribution of the variants by register (BCCWJ)

Table 6.9 Effect of register on the occurrence of *keesu* (BCCWJ)

		PB	PM	total
LW	N	286	26	312
	%	21.5%	35.6%	22.2%
non-LW	N	1046	47	1093
	%	78.5%	64.4%	77.8%
Total	N	1332	73	1405
	%	100.0%	100.0%	100.0%

$X^2 = 8.016$, d.f. = 1, $p < 0.01$

6.4.5. Genre (PB only)

Table 6.10 shows the proportion of the LW and the non-LWs by genre. The effect of genre proved to be significant on the occurrence of the LW ($X^2=27.711$, d.f.=1, $p<0.01$). The residual analysis shows that Social Science ($r=2.3$, $p<0.05$) has a positive effect on the occurrence of the LW whereas Technology & Engineering ($r=-2.6$, $p<0.01$) and Language ($r=-3.0$, $p<0.01$) have a negative effect on it. The positive effect of Social Science on the occurrence of the LW is apparently because *keesu* is associated more closely with the contexts in which social/legal problems are described as in (6.18a-b).⁵⁸

Table 6.10 Effect of genre on the occurrence of *keesu* (BCCWJ, PB only)

	General	Philosophy	History	Social Sciences	Natural Sciences	Tech.& Engineering	Industry	Arts	Language	Literature	Total
LW	N 10	20	14	131	39	13	30	13	8	7	285
	% 19.2%	26.7%	15.7%	24.7%	18.8%	11.8%	28.6%	31.0%	9.0%	29.2%	21.5%
	<i>r</i> -0.4	1.1	-1.4	*2.3	-1.0	** -2.6	1.8	1.5	** -3.0	0.9	
non-LW	N 42	55	75	399	168	97	75	29	81	17	1038
	% 80.8%	73.3%	84.3%	75.3%	81.2%	88.2%	71.4%	69.0%	91.0%	70.8%	78.5%
	<i>r</i> 0.4	-1.1	1.4	-2.3	1.0	2.6	-1.8	-1.5	3.0	-0.9	
total	N 52	75	89	530	207	110	105	42	89	24	1323
	% 100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

$X^2=27.711$, d.f. = 1, $p < 0.01$

r refers to adjusted residual. **/*: significant at the 1%/5% level in the residual analysis.

(6.18)

a. このころは家事事件に弁護士が関わるケースが増えてきたが、(PB43_00044)

Konogor-wa kaji-jiken-ni bengoshi-ga kakawaru keesu-ga

hue-tekita ga,

recently-TOP family-affairs-DAT lawyer-NOM be.involved case-NOM

increase-has.been but

‘Recently, cases in which a lawyer gets involved in family affairs have increased,’

⁵⁸ This does not necessarily mean that *keesu* has been developed particularly as a technical term used by professionals in legal or other related fields.

- b. また、日本と外国との商慣習のズレにより国際問題になるケースもあり、
(PB43_00557)

*Mata, nihon to gaikoku to-no shoo-shuukan-no
zure-niyori kokusai-mondai ninaru keesu-mo ari,*

Also, Japan and foreign.country and-GEN business-practice-GEN
difference-due.to international.problems result.in case-too exist

‘There also are cases in which differences in business practices can develop into
international problems,’

6.5. Multivariate analysis

6.5.1. Variables

A multivariate logistic regression analysis was conducted to evaluate the statistical significance of the independent variables examined above. A multivariate analysis examines the effect of an independent variable on a dependent variable while the other independent variables are controlled; this is not immediately observable in a univariate cross-tabulation analysis. We can also include an interaction term in a model to see whether the given independent variables interact with each other. The model takes the form of [Formula 1] or [Formula 1’], where p corresponds to the probability of the occurrence of the LW *keesu* (< *case*) as opposed to that of the non-LWs: *jirei*, *baai* and *rei*.

$$\log[p/(1-p)] = a_1x_1 + a_2x_2 + \dots + a_nx_n + b \quad \text{[Formula 1]}$$

$$p = 1/[1+\exp(-1)(a_1x_1 + a_2x_2 + \dots + a_nx_n + b)] \quad \text{[Formula 1’]}$$

The dependent variable is binary: the occurrence (=1) or non-occurrence (=0) of the loanword *keesu*. There are six independent variables, including two language-internal ones: (x_1) Usage and (x_2) Predicate, as well as four language-external ones: (x_3) Year of Birth, (x_4) Sex, (x_5) Education and (x_6) Register (Medium Type). All independent variables are categorical and the reference categories are those coded with the smallest numbers.

x_1 = Usage (gap-less RCH = 1, others = 0)

x_2 = Predicate (present/absent = 1, many/few = 2, increase/decrease = 3, occur = 4, assume = 5, report = 6)

x_3 = Year of Birth (-1930 = 1, 1940-50 = 2, 1960-70 = 3)

x_4 = Sex (M = 0, F = 1)

x_5 = Education (HS = 1, UNIV = 2, GRAD = 3)

x_6 = Register (Medium Type) (PB = 0, PM = 1)

Table 6.11 shows correlations between each pair of independent variables under consideration. It demonstrates that each variable is not highly correlated with any of the others.

Table 6.11 Correlation matrix: *keesu* (BCCWJ)

	Usage	Predicate	Year of Birth	Sex	Education	Register
Usage	1.000	-.187	.076	-.003	-.034	-.006
Predicate	-.187	1.000	.004	.012	.022	-.044
Year of Birth	.076	.004	1.000	.151	.068	.099
Sex	-.003	.012	.151	1.000	.088	.103
Education	-.034	.022	.068	.088	1.000	-.018
Register	-.006	-.044	.099	.103	-.018	1.000

6.5.2. Results

Table 6.12 shows the result of the logistic regression analysis (stepwise) using SPSS.

“B” (regression coefficient) corresponds to a_n in [Formula 1] and reveals how much influence the independent variable x_n has on the output p , compared with its reference category. A positive coefficient indicates that a given category has a higher probability of the occurrence of the LW than its reference category whereas a negative coefficient indicates that it has a lower probability than its reference category. 0 indicates that the effect of the independent variable in question on the output is not significant at the 5% level.

“Wald” is a test statistic that is used for testing the significance of B. The significance of a Wald statistic is evaluated based on a degree of freedom (df) and its p-value is shown in the column “Sig.”

“Exp(B)” represents the odds ratio, which is the value converted from B. The interpretation of Exp(B) is easier than that of B; for example, Exp(B)=2.070 in the independent variable “Register” means that the probability of occurrence of LW is 2.070 times stronger in PM than that in its reference category PM. The effect size is smallest (i.e. none) when Exp(B)=1 and grows as Exp(B) becomes distant from 1 (Exp(B)=10 would mean that the effect size is 10 times more, and, Exp(B)=0.1 would mean that it is 10 times less, in contrast to a reference category).

In the logistic regression model, unlike the linear regression model, the effect size of an independent variable on a dependent variable is not understood from the value of B by intuition, since B represents the amount of change in the left-hand side of [Formula 1], i.e. $\log[p/(1-p)]$, not the probability p itself, when x_n is in increment of 1 (see 6.5.1).

Table 6.12 Result of the logistic regression analysis: *keesu* (BCCWJ)

		B	S.E.	Wald	df	Sig.	Exp(B)
Usage	others						
	gap-less RCH	.595	.249	5.703	1	*	1.814
Predicate				76.784	5	**	
	present/absent						
	many/few	.609	.144	17.867	1	**	1.839
	increase/decrease	2.224	.338	43.408	1	**	9.246
	occur	2.078	.362	32.925	1	**	7.991
	assume	.815	.477	2.916	1	+	2.259
	report	-.421	.755	.310	1		.656
Year of Birth				9.988	2	**	
	-1930						
	1940-50	.095	.179	.285	1		1.100
	1960-70	.552	.203	7.404	1	**	1.737
Sex		0					
Education				17.284	2	**	
	HS						
	UNIV	-.267	.286	.875	1		.765
	GRAD	-.802	.290	7.629	1	**	.448
Register	PB						
	PM	.727	.267	7.401	1	**	2.070
	Constant	-1.928	.376	26.343	1	**	.145

$R^2 = 0.33$ (Nigelkerke $R^2 = 0.120$)

** $p < 0.01$, * $p < 0.05$, + $p < 0.10$

It demonstrates that the all independent variables except for Sex: Usage, Predicate, Year of Birth, Education and Register, proved to be significant.

First, we notice that, taking “others” as a reference, the gap-less RCH usage promotes the occurrence of the LW. Second, the occurrence of the LW is higher with predicates referring to [many/few], [increase/decrease] and [occur], compared with the reference category [present/absent]. Third, taking the oldest birth-year group as a reference, the LW occurs more frequently in those born in the 1960-70s but the difference between the reference category and those born in the 1940-50s does not prove to be significant. Fourth, it is shown that GRAD disfavors the use of the LW in

comparison with the reference category HS but the difference between HS and UNIV are not significant. Lastly, the LW occurs more frequently in PM than in PB. The coefficient of determination (R^2) is 0.33.

6.6. Additional analysis: *keesu* in CSJ

6.6.1. Eliciting data (CSJ)

In order to examine further the effect of stylistic factors, the spoken corpus, CSJ, is also investigated. Following the criteria shown in Table 6.2, tokens of the four variants are collected (Table 6.13). The table shows that after eliminating irrelevant contexts, the loanword *keesu* accounts for 12.0% of the all tokens, which is less frequent than that in the written corpus (BCCWJ), i.e., 22.2%.

Table 6.13 Tokens of the lexical variants (CSJ)

lemma	N	%
<i>keesu</i>	87	12.0
<i>jirei</i>	24	3.3
<i>baai</i>	431	59.4
<i>rei</i>	183	25.2
total	725	100.0

6.6.2. Internal variables

Table 6.14 shows that the LW occurs as a gap-less RCH (13.2%) more frequently than as others (8.1%) but this does not prove to be significant at the 5% level despite a tendency toward significance. Table 6.15 shows that the LW occurs more often with the predicate [many/few] (23.1%) than with the predicate [present/absent] (9.4%). There was no occurrence of the LW with the predicate [increase/decrease].

Table 6.14 Effect of usage on the occurrence of *keesu* (CSJ)

		others	gap-less RCH	total
LW	N	14	73	87
	%	8.1%	13.2%	12.0%
non-LW	N	158	480	638
	%	91.9%	86.8%	88.0%
Total	N	172	553	725
	%	100.0%	100.0%	100.0%

$X^2 = 3.182$, d.f. = 1, $p = 0.074$

Table 6.15 Effect of co-occurring predicate on the occurrence of *keesu* (CSJ)

		present/ absent	many/ few	increase/ decrease	occur	assume	report	total
LW	N	52	28	0	4	2	1	87
	%	9.4%	23.1%	0.0%	20.0%	20.0%	5.9%	12.0%
non-LW	N	503	93	2	16	8	16	638
	%	90.6%	76.9%	100.0%	80.0%	80.0%	94.1%	88.0%
total	N	555	121	2	20	10	17	725
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-square test is not conducted as there are several cells with expect count less than 5.

6.6.3. External variables

Table 6.16 shows that the occurrence of the LW is not independent from the speaker's year of birth. The use of the LW decreases from older generations to younger ones, indicating a "decrease" of the use of the LW, opposite to what we saw in Table 6.6.

Table 6.17 shows that the difference between sexes proved to be significant, with the LW favored by female speakers (17.4%) more than male speakers (10.2%). Table 6.18 shows that the educational effect proved to be significant, with the LW occurring less frequently among GRAD (9.1%) in particular. This is consistent with the result in Table 6.8.

Table 6.16 Effect of year of birth on the occurrence of *keesu* (CSJ)

		-1940	1950-60	1970-80	total
LW	N	20	46	21	87
	%	17.2%	14.2%	7.5%	12.1%
non-LW	N	96	277	258	631
	%	82.8%	85.8%	92.5%	87.9%
total	N	116	323	279	718
	%	100.0%	100.0%	100.0%	100.0%

$X^2 = 9.750$, d.f. = 2, $p < 0.01$

Table 6.17 Effect of sex on the occurrence of *keesu* (CSJ)

		M	F	total
LW	N	56	31	87
	%	10.2%	17.4%	12.0%
non-LW	N	491	147	638
	%	89.8%	82.6%	88.0%
Total	N	547	178	725
	%	100.0%	100.0%	100.0%

$X^2 = 6.553$, d.f. = 1, $p < 0.05$

Table 6.18 Effect of education on the occurrence of *keesu* (CSJ)

		HS	UNIV	GRAD	total
LW	N	9	42	36	87
	%	16.7%	15.8%	9.1%	12.2%
non-LW	N	45	223	361	629
	%	83.3%	84.2%	90.9%	87.8%
Total	N	54	265	397	716
	%	100.0%	100.0%	100.0%	100.0%

$X^2 = 7.962$, d.f. = 2, $p < 0.05$

Tables 6.19 – 6.22 show the effect of Register, Spontaneity, Formality and Audience Size on the occurrence of the LW. The effect of these variables proved to be significant except for Formality. First, the LW is more frequent in SPS (25.3%) than in APS (7.0%). Second, the LW is disfavored in “less spontaneous” speech in particular (5.9%), more than in “neither” or “more spontaneous” speech (14.0%, 14.2%). Third,

the LW is used more frequently in front of small audience (24.3%) than in front of larger audiences (5.6% and 8.9%).

Table 6.19 Effect of register (speech type) on the occurrence of *keesu* (CSJ)

		SPS	APS	total
LW	N	50	37	87
	%	25.3%	7.0%	12.0%
non-LW	N	148	490	638
	%	74.7%	93.0%	88.0%
Total	N	198	527	725
	%	100.0%	100.0%	100.0%

$X^2 = 45.303$, d.f. = 1, $p < 0.001$

Table 6.20 Effect of spontaneity on the occurrence of *keesu* (CSJ)

		low spontaneity	neither	high spontaneity	total
LW	N	10	25	52	87
	%	5.9%	14.0%	14.2%	12.2%
non-LW	N	159	154	313	626
	%	94.1%	86.0%	85.8%	87.8%
Total	N	169	179	365	713
	%	100.0%	100.0%	100.0%	100.0%

$X^2 = 8.175$, d.f. = 2, $p < 0.05$

Table 6.21 Effect of formality on the occurrence of *keesu* (CSJ)

		casual	neither	formal	total
LW	N	12	44	28	84
	%	13.5%	13.9%	10.6%	12.6%
non-LW	N	77	272	236	585
	%	86.5%	86.1%	89.4%	87.4%
Total	N	89	316	264	669
	%	100.0%	100.0%	100.0%	100.0%

$X^2 = 1.523$, d.f. = 2, $p = 0.467$

Table 6.22 Effect of audience size on the occurrence of *keesu* (CSJ)

		small	medium	large	total
LW	N	53	18	14	85
	%	24.3%	5.6%	8.9%	12.2%
non-LW	N	165	301	144	610
	%	75.7%	94.4%	91.1%	87.8%
Total	N	218	319	158	695
	%	100.0%	100.0%	100.0%	100.0%

$$X^2 = 44.211, \text{ d.f.} = 2, p < 0.001$$

6.6.4. Multivariate analysis

Table 6.23 shows correlations between the above-mentioned independent variables. Register, Audience Size and Education have relatively strong correlation with each other ($r > 0.5$). Spontaneity and Formality are moderately correlated ($r = -0.483$). To avoid the multicollinearity problem, Audience size and Education are excluded from the analysis, considering that Register serves as a good indicator of these two variables. By the same token, Formality is excluded from the analysis, considering that Spontaneity serves as a good indicator of this variable.⁵⁹

Table 6.23 Correlation matrix: *keesu* (CSJ)

	Usage	Predicate	Year of Birth	Sex	Education	Register	Spontaneity	Formality	Audience
Usage	1.000	-.151	.101	-.006	-.141	-.073	.099	-.118	-.088
Predicate	-.151	1.000	.015	.015	.105	.102	-.006	.017	.132
Year of Birth	.101	.015	1.000	.012	.054	.224	-.077	.039	.188
Sex	-.006	.015	.012	1.000	-.163	-.172	-.067	.044	-.151
Education	-.141	.105	.054	-.163	1.000	.648	-.169	.258	.556
Register	-.073	.102	.224	-.172	.648	1.000	-.289	.365	.804
Spontaneity	.099	-.006	-.077	-.067	-.169	-.289	1.000	-.483	-.288
Formality	-.118	.017	.039	.044	.258	.365	-.483	1.000	.399
Audience	-.088	.132	.188	-.151	.556	.804	-.288	.399	1.000

⁵⁹ Note that the value of VIF is 3.553 for Register, 2.990 for Audience size, 1.793 for Education, 1.346 for Spontaneity, 1.468 for Formality, and less than 1.1 for the rest if all the variables in Table 6.23 are included. After eliminating the variables in the shaded cells in Table 6.23, however, the remaining variables no longer exhibit a relatively high VIF value (1.181 at the highest).

Consequently, the following internal and external factors are incorporated into a multivariate analysis. An interaction term (Register×Spontaneity) is added in order to examine the existence of an interaction effect between the two variables. It is noted that the tokens with the value [increase/decrease] for the variable Predicate are excluded from the analysis, since the occurrence of the LW was categorical with this predicate (Table 6.15).

x_1 = Usage (gap-less RCH = 1, others = 0)

x_2 = Predicate (present/absent = 1, many/few = 2, occur = 4, assume = 5, report = 6)

x_3 = Year of Birth (-1940 = 1, 1950-60 = 2, 1970-80 = 3)

x_4 = Sex (M = 0, F = 1)

x_5 = Register (SPS = 0, APS = 1)

x_6 = Spontaneity (low spontaneity = 1, neither = 2, high spontaneity = 3)

$x_5 .x_6$ = Register×Spontaneity (interaction term)

Table 6.24 shows the result of the logistic regression analysis (stepwise) by SPSS. It demonstrates that the effect of Predicate, Register and the interaction term Register×Spontaneity proved to be significant. Usage, Year of Birth, Sex and Spontaneity have been shown not to be significant. The coefficient of determination (R^2) is 0.24.

Table 6.24 Result of the logistic regression analysis: *keesu* (CSJ)

	B	S.E.	Wald	df	Sig.	Exp(B)
Usage	0					
Predicate			21.873	4	**	
present/absent						
many/few	1.082	.277	15.288	1	**	2.951
occur	1.645	.627	6.875	1	**	5.179
assume	1.632	.838	3.791	1	+	5.116
report	-.047	1.065	.002	1		.954
Year of Birth	0					
Sex	0					
Register						
SPS						
APS	-2.692	.543	24.611	1	**	.068
Spontaneity	0					
Register×Spontaneity			8.788	2	*	
SPS×low spontaneity						
SPS×neither						
SPS×high spontaneity						
APS×low spontaneity						
APS×neither	1.690	.579	8.512	1	**	5.419
APS×high spontaneity	1.122	.572	3.847	1	*	3.071
Constant	-1.339	.185	52.204	1	**	.262
R ² = 0.24 (Nigelkerke R ² = 0.179)		** p<0.01, * p<0.05, + p<0.10				

First, the LW occurs more frequently with the predicates [many/few] and [occur], compared with [present/absence] as a reference category. The result is consistent with what we have seen in the analysis of BCCWJ (Table 6.12).

Second, there is an interaction between Register and Spontaneity. Figure 6.11 illustrates the relation between Register and Spontaneity by showing the predicted probabilities of the occurrence of *keesu* followed by Predicate [many/few] as an example. The predicted values were calculated with the regression equation shown in [Formula 1']. We can see that Spontaneity comes into play in APS, with [APS×neither] and [APS×high spontaneity], as compared with [APS×low spontaneity], showing higher probabilities of the occurrence of the LW. In contrast, Spontaneity does not play a role

in SPS. This indicates that a higher degree of spontaneity promotes the use of the LW only in APS. SPS, however, shows higher probabilities of the occurrence of the LW than APS does, regardless of the degree of spontaneity.

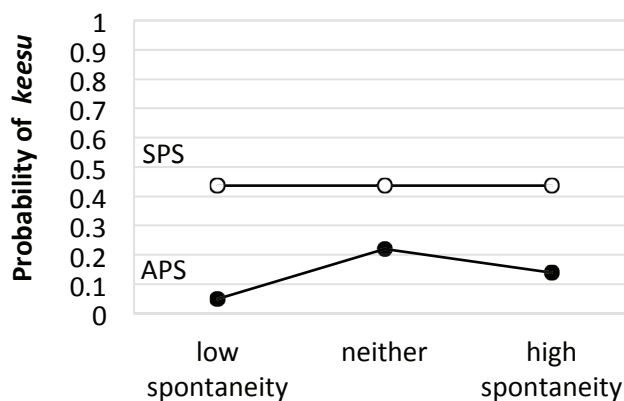


Figure 6.11 Interaction between Register and Spontaneity (those co-occurring with Predicate [many/few])

Last but not least, it is noted that Year of Birth, which showed a slightly odd distribution in the univariate analysis (Table 6.16), did not prove to be significant in the multivariate one.

6.7. Discussion and conclusion

This chapter identified both language-internal and external factors that have effects on the occurrence of the loan variant *keesu* as opposed to its native variants *jirei*, *baai* and *rei*. The contextual approach was attempted in order to take a step closer to a more fine-grained analysis of lexical variation. I have carefully specified environments in which the four variants alternate actively based on the findings from the previous study of these particular lexical items (Kim 2011). We have witnessed that the four variants

overlap in function as a formal noun (in particular as a gap-less RCH). Of course, the four variants also have their own unique functions (a division of labor), which explains the co-existence of these near-synonyms. These environments were excluded from the analysis.

As a result of the multivariate analysis, two internal variables: (i) Usage and (ii) co-occurring Predicate, proved to have an influence on the occurrence of the loan variant. First of all, the loanword proved to be more strongly associated with the gap-less RCH usage (BCCWJ). This provided statistical evidence in favor of Kim's (2011) observation.

Second, compared with the predicate [present/absent], predicates referring to (a) [many/few], (b) [increase/decrease] and (c) [occur] co-occur more often with the loan variant (BCCWJ, CSJ). It may be worth mentioning that, as opposed to the predicate [present/absent], (a) [many/few] and (b) [increase/decrease] are characterized having the semantic feature of "quantitative evaluation (+)," while (b) [increase/decrease] and (c) [occur] having the feature of "change of state (+)" as shown in Table 6.25. This may imply that *keesu* tends to be used in the context in which the *occurrence* or a *quantitative aspect* of a certain event is emphasized, co-occurring with the predicates (a) [many/few], (b) [increase/decrease] and (c) [occur]. In contrast, *baai* occurs most frequently with the predicate [present/absent], indicating that it is used more often in the context in which only the *existence/non-existence* of a certain fact matters.

Table 6.25 Classification of co-occurring predicates by their semantic features

	Change of state (-)	Change of state (+)
Quantitative evaluation (-)	present/absent	(c) occur
Quantitative evaluation (+)	(a) many/few	(b) increase/decrease

With regard to external factors, the author's (i) Year of Birth, (ii) Education, (iii) Register and (iv) Spontaneity proved to have impacts on the use of the loan variant. First, there is a tendency for the youngest generation to use the loanword more often (BCCWJ). This can be considered an indication of a generational change in favor of *keesu* as opposed to its Japanese counterparts. It is important to notice that the generational effect on the adoption of loanwords, which has been attested to by linguistic attitudes of the public (Chapter 4), has gained additional support from the empirical evidence of their linguistic behavior.

Second, the effect of the level of education also proved to be significant, with the loanword *keesu* being disfavored by the most educated authors (BCCWJ).

Third, register is another significant factor, with the proportion of *keesu* being higher in magazines than in books (BCCWJ). I pointed out in Chapter 5 that the register difference in BCCWJ is attributable to a stylistic difference associated with the degree of "formality" of the text. In other words, the loanword is disfavored in more formal text like books, compared with magazines.

Additional analysis with the spoken corpus (CSJ) provided a more comprehensive understanding of the relationship between register and stylistic parameters such as Spontaneity. In the present data, the use of the LW is inhibited in APS, which is characterized having a more formal speech style than SPS. Recall that APS is associated with larger audiences and speakers with higher educational backgrounds, implying that

the loan variant is disfavored in such environments. Furthermore, Spontaneity is in interaction with Register, with more spontaneous speech promoting the use of the LW in APS. The results imply that, roughly speaking, the occurrence of the LW is more closely associated with more informal and more spontaneous speech, and, additionally, smaller audiences and those with lower educational backgrounds.

To conclude, a more detailed analysis of lexical variation based on the “contextual approach” has revealed that the occurrence of the LW *keesu* is dependent on both language-internal and external factors. In addition to distribution based on internal factors, sociolinguistically asymmetric distributions of *keesu* as a loan variant give an interesting insight into its status as opposed to its native counterparts in contemporary written and spoken Japanese. Age gradient indicates the presence of a generational change in favor of the loan variant. The tendency for more educated people as well as more formal or less spontaneous settings to disfavor the loan variant implies that the loanword is not fully established in certain social categories or settings. This asymmetry implies, although *keesu* may have become an integral part of basic Japanese vocabulary in some respect (Kim 2011), it still remains non-standard and subject to various sociolinguistic restrictions to a certain degree.

In the next chapter I will examine stylistic constraints on loanword use more deeply using interview-based data on the loanword *sapooto* (< *support*) (NINJAL 2005a).

7. Stylistic constraints on the use of loanwords: The alternation between *sapooto* (< *support*) and its native equivalents as intra-speaker variation

7.1. Introduction

7.1.1. Aim: Elucidating the effect of style on lexical variation

This chapter aims to elucidate the mechanism of the occurrence of loanwords from a stylistic point of view, through a case study of the loanword *sapooto* (< *support*) and its native alternatives.⁶⁰

The earlier chapters have verified the statistical significance of the influence of age on the adoption of loanwords: there is a general tendency for younger people to be more receptive to loanwords. Such age differentiation has also been repeatedly reported in previous surveys (BBK 1999, 2000, 2003, 2008, 2013; NINJAL 2004, 2005a; Tanaka 2007; Jinnouchi 2007). However, stylistic effect on the phenomenon has not yet been investigated sufficiently in spite of its importance to an extensive elucidation of the theory of language change.

The study of style, i.e., intra-speaker variation, is an attempt to understand variation that occurs within one speaker's language depending on various contexts. Inoue (2000: 557) points out that linguistic change progresses in different manners across speech style. In order to express this, he proposed a three-dimensional model: the Water Tank Model, in which the diffusion of a given variant is captured by two variables: time and speech style as seen in Figure 7.1. It shows that the higher the level of formality of a setting is, the faster an innovative form spreads. Note that his model is based on dialect

⁶⁰ Chapter 7 is a revised version of the paper published as: Kuya, A. (in press) *Gairaigo shiyoo ni kakaru sutairu no seiyaku: Sapooto to kisongo to no tsukaiwake ni miru washanai barieeshon. Shakai Gengo Kagaku [The Japanese Journal of Language in Society]*, 19(1).

standardization, i.e., change from above. In the case of change from below, it is expected that the higher the level of formality of a setting is, the slower an innovation spreads.

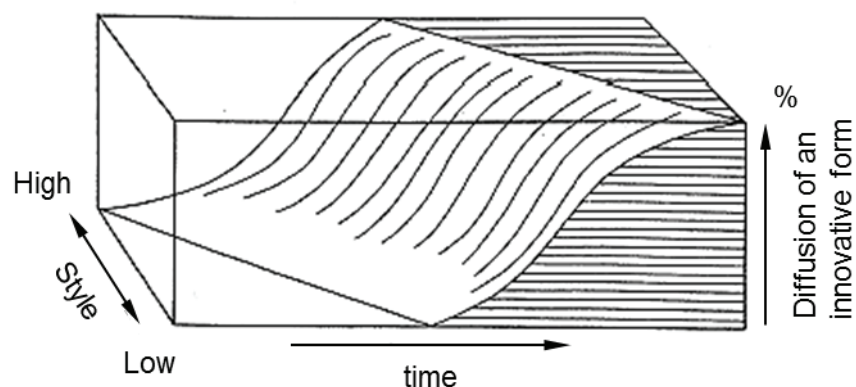


Figure 7.1 "Water Tank Model"
(Inoue 2000: 556, Figure 27-7)

According to NINJAL (2004: 32), almost 40% of people turned out to decide the use or non-use of loanwords depending on where the utterance takes place and to whom the utterance is directed. This implies that the diffusion of a given loanword progresses in a complicated way, i.e., it spreads over a speech community at different speeds across speech style. This chapter, therefore, will take into consideration a stylistic effect on the use of loanwords, in order to elucidate a more elaborate process of loanword diffusion.

First it will be shown that the occurrence of the loanword *sapooto* is predicted from (i) the speaker's year of birth (age), (ii) their educational background and (iii) speech style: That is, the diffusion of the loanword is in concert with stylistic variation. Second, we will see the mechanism in which stylistic variation occurs. The present research will show that the alternation between the loanword and native alternatives occurs (i) as a response to the formality of a speech setting, (ii) as a response to the addressees and (iii)

as a speaker's proactive act of proving their identity, which can be explained with the (i) Attention to Speech Model, (ii) Audience Design Model and (iii) Speaker Design Model, respectively.

7.1.2. Previous studies on stylistic variation

In spite of a number of studies on the relation between language and style, a unified theory of stylistic variation has not been reached (Takano 2012). The earliest approaches, such as Labov (1972) and Trudgill (1974), claim that one's linguistic behavior is determined by the level of formality of speech style or, in other words, the amount of attention paid to speech (Attention to Speech Model). This model predicts that a prestigious variant is favored in a formal setting as a result of an increased amount of attention paid to speech (careful speech).

In contrast, the Audience Design Model (Bell 1984) points out that stylistic variation occurs in response to audience. For example, Rickford & McNair-Knox (1994) reported that the probability of the occurrence of vernacular speech varies depending on to whom the utterance is addressed.

The Speaker Design Model (Schilling-Estes 2002) puts more emphasis on the speaker's intention. A speaker uses language selectively in order to prove his/her social identity. Ekert (1988: 206) points out that the high school student group called *Burnouts* in Detroit tends to adopt ongoing urban sound changes proactively as a result of their motivation to orient themselves to the urban community that they are associated with.

The previous studies have shown that stylistic variation is interpretable from various perspectives. Analysts have to consider which model is most convincing to explain the process in which certain stylistic variation occurs. However, it should also be noted that

the mechanism of a given style-shifting is sometimes not fully explainable by a single model. This chapter thus attempts to employ more than one model to obtain a more comprehensive understanding of a phenomenon.

Furthermore, we should aware that stylistic factors often interact with other external factors. Figure 7.2 shows the distribution of the variable /ng/ in Norwich across style and socioeconomic class (Trudgill 1974: 92). Index score (ng) represents the proportion of [n] to the standard [ŋ] (ibid: 84). Different speech styles are arranged along a scale of formality: the most formal “Word List Style (WLS),” “Reading Passage Style (RPS),” “Formal Speech style (FS)” to the least formal “Casual Speech style (CS).” The consistent tendency for the index to decrease from CS to WLS indicates that the occurrence of the variable (ng) is not independent from the social context in which a speech is delivered, i.e., speech style. It should also be noted that the manner in which the stylistic factor affects this variation is different across social classes (ibid: 92-93). For example, the two middle-class (MC) groups show the greatest amount of stylistic differentiation between CS and FS, whereas the three working-class (WC) groups do not make a clear stylistic difference between these two speech styles; they instead show the sharpest stylistic differentiation between FS and RPS. Trudgill points out that this appears to show the ability of the MC groups to control the (ng) forms in FS to a level nearer to that of the more formal styles, i.e., RPS and WLS (ibid: 93).

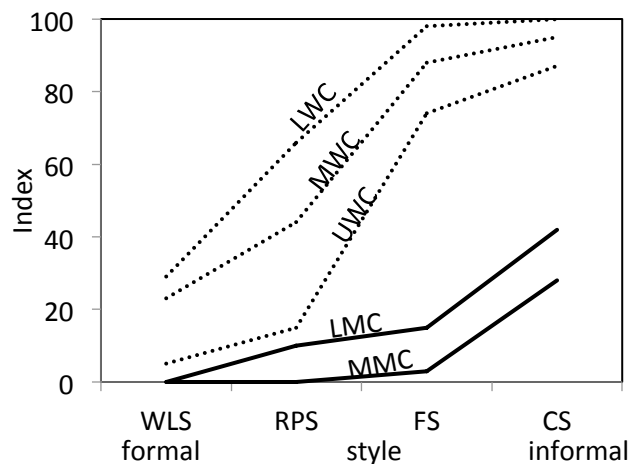


Figure 7.2 Variable /ng/ in Norwich by class and style (Trudgill 1974: 92, Figure 14)

Here there is an interaction between social class and style, with the stylistic effect on the variable (ng) modified by social class, i.e., the gradient of the data is different across social classes. An interaction effect can sometimes demonstrate the “crossover pattern.” We can see the most famous example of this in a study of New York English (Labov 1972: 114, see also Labov 1994: 87), wherein people from the second highest socioeconomic class use the prestigious variant of the variable (r) more frequently than those from the highest socioeconomic class in formal settings, although the use of this (r) by these two groups in more informal settings is the other way around (hypercorrection). Since previous studies on variation in Japanese (e.g., Sano 2008, 2011) have also reported the presence of an interaction between style and other social factors, this chapter will also take this point into consideration (7.3.4).

7.1.3. Research questions

The following sections aim:

- (i) To provide a statistical model that predicts stylistic variation with regard to the use of the loanword (quantitative analysis)

- (ii) To elucidate the mechanism of stylistic variation (qualitative analysis)

In the first part of the chapter we will start with univariate analyses of the effect of several external factors (including style) on the loanword distribution (7.3). A logistic regression analysis will then be conducted to give a multivariate model that predicts the use/non-use of the loanword *sapooto* (< *support*) by style (7.4). This model will help give a multidimensional view of how the loanword spreads into a community across styles.

The latter part of the chapter aims to describe a qualitative aspect of stylistic variation (7.5). It concerns the question of how and why individual speakers switch from one lexical item to another. An in-depth analysis of individual speakers will let us discuss the nature of the observed style-shifting as well as speakers' evaluation of the loanword in question.

7.2. Data

7.2.1. *The National Survey on Attitudes to Loanwords II* (NINJAL 2005a)

NINJAL conducted *The National Survey on Attitudes to Loanwords II* in 2004 (NINJAL 2005a).⁶¹ The targeted population is the entire Japanese-speaking population in Japan who are aged 15 and above. The statistical sample is 4,500 people selected randomly with a stratified sampling method (stratified by administrative divisions and area). There were valid responses from 3,090 people (68.7%), 1,421 male speakers and 1,669 female speakers. The survey was conducted in the form of a face-to-face

⁶¹ I owe this chapter to the generosity of NINJAL, who allowed me to access the raw data of their survey.

interview in which an informant was interviewed individually based on the prepared questionnaire.

This was the second attempt by NINJAL to investigate the public attitudes to loanwords, following their first survey conducted in 2003: *The National Survey on Attitudes to Loanwords* (NINJAL 2004). The first survey concerns people's overall attitudes to loanwords in general. For example, it includes questions on whether they have a positive attitude towards a further influx of loanwords in the future, and whether or not they usually use loanwords in general. The second survey, by contrast, puts more focus on the question of how speech style or social setting affects the use of loanwords. Lexical choice between the loanword *sapooto* (< *support*) and its Japanese equivalents *shien* (SJ) and *tedasuke* (NJ), all referring to “support/help”, was used by NINJAL as a case study.

As shown below, informants were asked to select one lexical item from the three in the distinct three speech settings: (i) when talking to friends (Friends), (ii) when talking in public (Public) and (iii) when talking to elderly people (Elders). An example context was presented to the informants to help them make choices more easily, such as “[] at the community level should be given when one starts a new farming business.” They were asked to select one word among the above three to fill the blank “[]” in this example sentence.

The informants were subsequently asked the reason why they chose that word in each case from two different perspectives. Those categorized as A concern (purely) the linguistic meanings that a word delivers, whereas those categorized as B concern things

outside of a purely linguistic meaning, i.e., the nuances socially embedded in each word.

Interview conducted by NINJAL (2005a: 125)

Example context:
Atarashiku noogyoo-o hajimeru tame-ni-wa,
chiiki-no []-ga hitsuyoo-desu.

newly farming.business -ACC to.start in.order-to-TOP
 community-GEN []-NOM necessary-COP.POL

‘[] should be given at the community level when one starts a new farming business’

Question: In the above context,

Q2: Which of the following three words do you usually use to fill the blank when talking to **friends**? (Friends)

Q3: Which of the following three words do you usually use to fill the blank when talking in **public**? (Public)

Q4: Which of the following three words do you usually use to fill the blank when talking to **elderly people** who you are unfamiliar with? (Elders)

Choices: (all the following items refer to ‘support’ or ‘help’)

Loanword (LW)	Sino-J (NJ)	Native-J (NJ)
<i>sapooto</i> (< <i>support</i>)	<i>shien</i>	<i>tedasuke</i>



Q2-4: Why do you select that word? (select one each from A and B)

A	B
It is concise.	It sounds novel.
It is easy to understand.	It sounds intelligent.
It has a precise meaning.	It sounds stylish/fashionable.
It is most familiar to me.	It sounds colloquial.
No applicable choice.	It sounds dignified.
Not sure.	No applicable choice.
	Not sure.

Table 7.1 shows that the distribution of each word varies across settings. The LW *sapooto* occurs most frequently in Friends. In contrast, the use of the LW is least frequent in Elders (3.3%). The LW is selected in Public (27.3%) at a level little lower than that for Friends (33.4%).

Table 7.1 Lexical choice across different settings
(NINJAL 2005a: 18)

	LW	SJ	NJ	Not sure	Total
Friends	1033	790	1204	63	3090
	33.4%	25.6%	39.0%	2.0%	100%
Public	843	1455	708	84	3090
	27.3%	47.1%	22.9%	2.7%	100%
Elders	102	402	2520	66	3090
	3.3%	13.0%	81.6%	2.1%	100%

7.2.2. Contributions of the present data to the study of stylistic variation

One of the advantages of using the present data is that the sample is a good representative of the population: We are reasonably able to extend the result obtained here to the whole population, i.e., the entire Japanese-speaking population in Japan.

In addition, the loanword *sapooto* is regarded as a typical example of a well-attested loanword in Japanese. First, according to NINJAL (2007b: 275), people understand the meaning of this word 81.8% of the time on average and even the oldest age group (aged 60 and above) shows a relatively high level of comprehension (61.8%). Moreover, the loanword has been shown to be well attested in written and spoken language (Chapter 5).⁶² Second, the distribution pattern of *sapooto* is not inconsistent with the general trend that the loanwords under investigation in Chapter 5 have exhibited. These studies

⁶² Kim (2011: 26) also points out that the loanword in question has become a basic word in written Japanese.

provide evidence in support of the assumption that *sapooto* can be seen as representative of loanwords that are well adopted in Japanese. Hence, it is not unreasonable to generalize a conclusion from an analysis of *sapooto*, within the bounds of these well-adopted kinds of loanwords.

7.2.3. Scope of analysis

Henceforth, the focus is put on two of the three settings originally studied by NINJAL: Friends and Public. The reason for focusing only on these two settings is that it is difficult to align the three settings, i.e. Friends, Public and Elders, on the same continuum of style. Based on the traditional definition of style, in which various speech settings are aligned along the continuum of formality (Casual vs. Careful, Labov 2001b), we are likely to place a setting like Friends close to the Casual pole (Labov 1972: 85). Speech addressed to a general audience, what Labov calls “Soapbox,” is in contrast placed closer to the Careful pole (2001b: 91).

However, a setting like Elders, which is defined as a situation where one talks to elderly people one is unfamiliar with, cannot be placed on the continuum of Informal-Formal. In addition, we see that the occurrence of the loanword in Elders is too rare across all generations to be applicable to the research being undertaken (Figure 7.3).

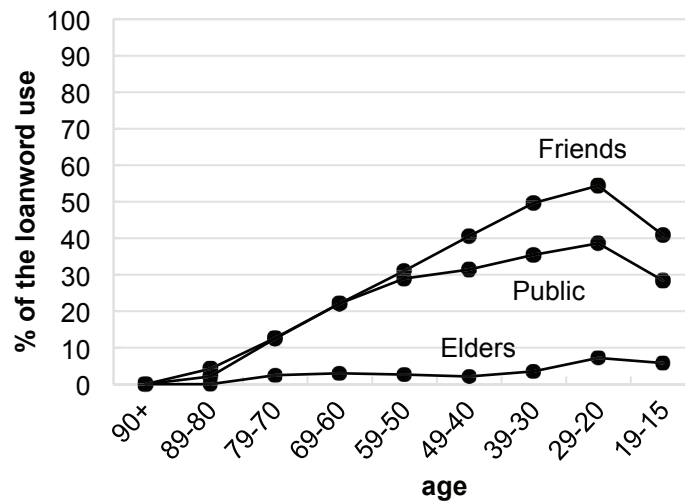


Figure 7.3 Distribution of *sappoto* by age and style

7.3. Effect of external factors

This section examines the effect of individual social factors on the choice to use the loanword using cross tabulation.

7.3.1. Age

Figure 7.4 shows the distribution of the LW, SJ and NJ across age groups and styles. In Friends, the proportion of the LW increases from older to younger age groups, with those in their 40s or younger selecting the LW as their first choice. Likewise, there is a similar age gradient in Public, but the LW is no longer the first choice for any age group as a result of the jump in the relative frequency of the SJ in all the age groups.

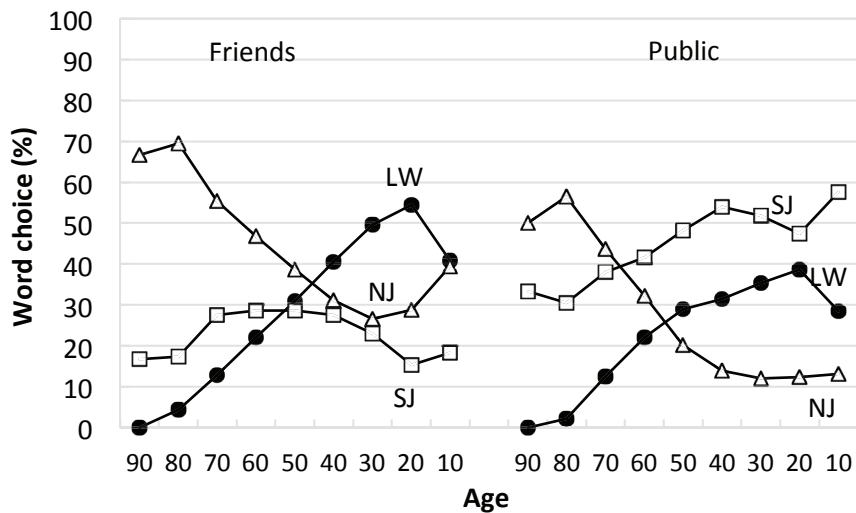


Figure 7.4 Lexical variation by age

It is notable that, in both Friends and Public, those in their teens use the LW less frequently than those in their twenties in spite of the general tendency for younger people to favor loanwords. This phenomenon is, as discussed in Chapter 4, explainable in terms of a delay in lexical acquisition among teens: they have not acquired the loanword *sapooto* to the same extent as older generations have. The example sentence presented to the informants in the interview describes a rather social/political issue, which is unlikely to be a common topic of conversations for teens. Hence, it is possible that they are not familiar enough with the loanword in question to judge whether it can be used in this specific context.

Henceforth, those in their teens (N=137) are excluded from the analysis as they are shown to behave differently from those in other age groups. Moreover, those in their 80s (N=46) and 90s (N=6) are also excluded from further analysis because of their very small sample size. The number of the remaining informants (those aged 20s-70s) is 2,901.

Table 7.2 shows the use and non-use of the LW across age groups (the occurrence of the SJ and NJ are grouped together as “Non-LW”). The chi-square test shows that age is not independent from the occurrence of the LW in both Friends and Public.

Table 7.2 Effect of age on the use/non-use of *sapooto*

		20s	30s	40s	50s	60s	70s	Total	
Friends	LW	149	252	193	192	138	51	975	$X^2(5)=238.163$ $p<0.001$
	Non-LW	125	256	283	427	487	348	1926	
	Total	274	508	476	619	625	399	2901	
Public	LW	106	180	150	179	138	50	803	$X^2(5)= 91.332$ $p<0.001$
	Non-LW	168	328	326	440	487	349	2098	
	Total	274	508	476	619	625	399	2901	

7.3.2. Education

Figure 7.5 shows the distribution of lexical items by speakers’ educational background with JHS, HS and UNIV referring to elementary/junior high school graduates, high school graduates and university graduates, respectively. It demonstrates the tendency for those with higher educational backgrounds to use the LW more frequently. It is remarkable that those in UNIV select the LW as their first choice in Friends, but its proportion in Public drops largely to a level close to that of HS. Instead, their first choice switches to the SJ.

We should, however, pay attention to the possibility that the educational factor is interrelated with the age factor. Table 7.3 shows that the order of the educational categories in LW use, i.e., JHS < HS < UNIV in Friends and JHS < HS ≤ UNIV in Public, holds true for most age groups, even when each educational category is stratified by age. It is then not unreasonable to claim that the educational factor is independent of the age factor to a certain degree (see also 7.4.1).

Statistical testing shows that, as shown in Table 7.4, the effect of one's educational background on the occurrence of the LW is significant in both Friends and Public.

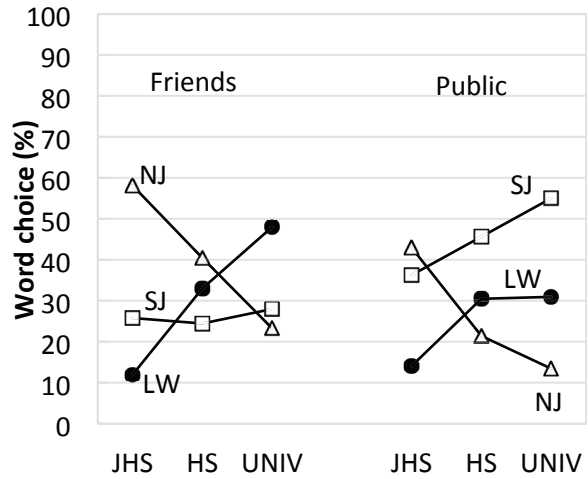


Figure 7.5 Lexical variation by education

Table 7.3 Proportion (%) of LW to non-LW by education and age

		20s	30s	40s	50s	60s	70s	Total
Friends	JHS	38.9	16.7	30.8	9.0	11.9	9.5	12.0
	HS	49.3	49.0	33.6	29.7	25.2	14.8	33.1
	UNIV	63.5	52.4	49.8	44.9	30.2	26.3	48.0
Public	JHS	16.7	25.0	30.8	21.8	13.0	10.5	14.1
	HS	40.7	37.9	33.2	29.7	27.3	12.6	30.6
	UNIV	39.1	34.0	29.2	30.5	21.7	23.7	31.0

Table 7.4 Effect of education on the use/non-use of *sapoto*

		JHS	HS	UNIV	Total	
Friends	LW	62	482	425	969	$X^2(2)=190.499$ $p<0.001$
	Non-LW	456	976	460	1892	
	Total	518	1458	885	2861	
Public	LW	73	446	274	793	$X^2(2)= 58.646$ $p<0.001$
	Non-LW	445	1012	611	2068	
	Total	518	1458	885	2861	

Samples whose educational background is unknown (N=40) are excluded.

7.3.3. Sex

Sex differentiation does not appear to be very striking for the LW, compared with that for the SJ or NJ (Figure 7.6). The chi-square test (Table 7.5) indicates that sex is independent of the occurrence of the LW.

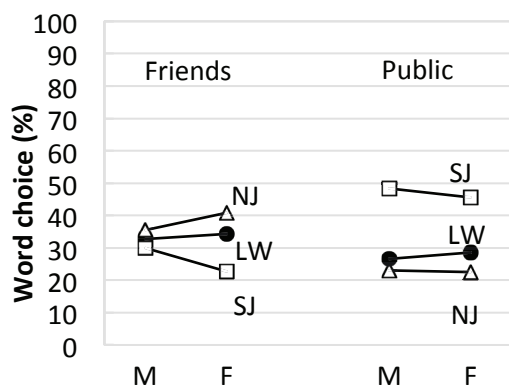


Figure 7.6 Lexical variation by sex

Table 7.5 Effect of sex on the use/non-use of *sapooto*

		Male	Female	Total	
Friends	LW	434	541	975	$X^2(1)=0.798$
	Non-LW	891	1035	1926	$p=0.372$
	Total	1325	1576	2901	
Public	LW	352	451	803	$X^2(1)=1.512$
	Non-LW	973	1125	2098	$p=0.219$
	Total	1325	1576	2901	

7.3.4. Style

As we saw earlier in Table 7.1, the occurrence of the LW varies across the two speech styles. Table 7.6 confirms that the difference reaches statistical significance.

Table 7.6 Effect of style on the use/non-use of *sapoto*

		Friends	Public	Total	
	LW	975	803	1778	$X^2(1)=23.991$ $p<0.001$
Friends	Non-LW	1926	2098	4024	
	Total	2901	2901	5802	

Figures 7.7 and 7.8 are provided to examine the interaction between style and other social factors: (i) age and (ii) education. In Figure 7.7, style-shifting occurs only among those aged 40s and below. Likewise, Figure 7.8 shows that style-shifting occurs only among those from UNIV. These figures imply the possibility that the effect of style on the use of the LW comes into play in interaction with the effect of age and/or education.

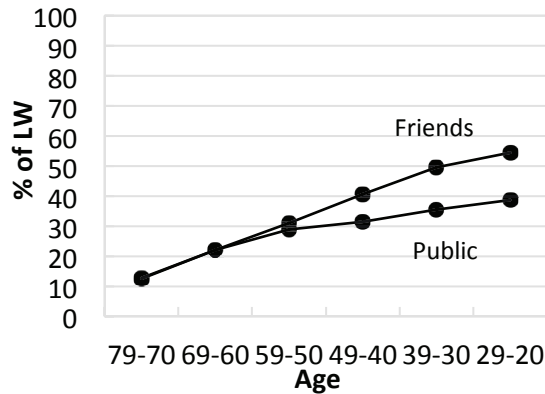


Figure 7.7 Distribution of *sapoto* by age and style

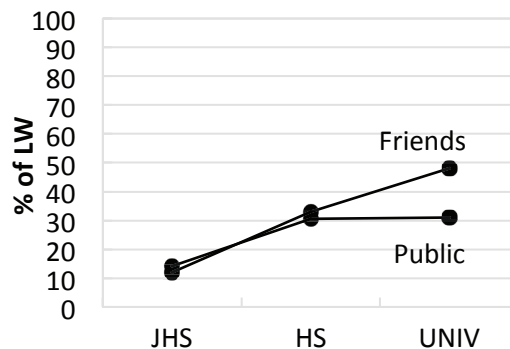


Figure 7.8 Distribution of *sapoto* by education and style

7.4. Multivariate analysis

7.4.1. Variables

A multiple logistic regression model is employed to evaluate the effect of each independent variable on the phenomenon. The general formulae of the model is repeated here as [Formula 1] and [Formula 1']. p in the equations represents the probability of the occurrence of the LW *sapooto* as opposed to its native equivalents *shien* and *tedasuke* (a and b are constants).

$$\log[p/(1-p)] = a_1x_1 + a_2x_2 + \dots + a_nx_n + b \quad \text{[Formula 1]}$$

$$p = 1/[1+\exp(-1)(a_1x_1 + a_2x_2 + \dots + a_nx_n + b)] \quad \text{[Formula 1']}$$

The dependent variable is considered to be binary, i.e., the occurrence (=1) or non-occurrence (=0) of the LW.⁶³ There are six independent variables including two interaction terms as follows:

x_1 = Year of Birth (converted from Age)

x_2 = Education

x_3 = Style

x_4 = Sex

$x_1 . x_3$ = Year of Birth×Style (interaction term)

$x_2 . x_3$ = Education×Style (interaction term)

⁶³ Although the dependent variable actually has 3 values (LW, SJ, NJ), binary logistic regression is employed since the focus of the present study is on the occurrence/non-occurrence of the LW.

The year of birth of the speakers (converted from their age) is inputted as a numeral variable. In practice, a median value is adopted to represent each age range: 1929.5 (70-79), 1939.5 (60-69), 1949.5 (50-59), 1959.5 (40-49), 1969.5 (30-39), 1979.5 (20-29). The other variables are all categorical ones. Education, Style and Sex are coded as (1 = JHS, 2 = HS, 3 = UNIV), (1 = Friends, 2 = Public) and (1 = Male, 2 = Female), respectively. Reference categories are those having the lowest value (=1).⁶⁴ As discussed in 7.3.4, two interaction terms, i.e., Year of Birth×Style and Education×Style, are added into the model. Table 7.7 is provided to show how samples are coded, with those in their 20s as an example.

⁶⁴ It seems that independent variables do not have high multicollinearity. Correlation matrix showed that the highest correlation exists between age and education: $r = -0.421$ (a low-moderate correlation). However, the highest VIF remains below 2 (≤ 1.215). Hence, both age and education are treated as distinct variables here.

Table 7.7 Example of data coding (samples of those in 20s)

Age Range	Year of Birth	Sex	Education	Style	Occurrence of LW	No. of Sample
20-29	1979.5	1	1	1	1	4
20-29	1979.5	1	2	1	1	31
20-29	1979.5	1	3	1	1	28
20-29	1979.5	2	1	1	1	3
20-29	1979.5	2	2	1	1	38
20-29	1979.5	2	3	1	1	45
20-29	1979.5	1	1	1	0	3
20-29	1979.5	1	2	1	0	40
20-29	1979.5	1	3	1	0	22
20-29	1979.5	2	1	1	0	8
20-29	1979.5	2	2	1	0	31
20-29	1979.5	2	3	1	0	20
20-29	1979.5	1	1	2	1	1
20-29	1979.5	1	2	2	1	26
20-29	1979.5	1	3	2	1	18
20-29	1979.5	2	1	2	1	2
20-29	1979.5	2	2	2	1	31
20-29	1979.5	2	3	2	1	27
20-29	1979.5	1	1	2	0	6
20-29	1979.5	1	2	2	0	45
20-29	1979.5	1	3	2	0	32
20-29	1979.5	2	1	2	0	9
20-29	1979.5	2	2	2	0	38
20-29	1979.5	2	3	2	0	38

7.4.2. Results

Table 7.8 shows the result of the logistic regression analysis (stepwise) using SPSS. “B” (regression coefficient) in the table corresponds to a_n in [Formula 1] and represents the factor weight of independent variable x_n . A positive value indicates the preference for the use of the LW whereas a negative value indicates the preference for the non-use of the LW. 0 indicates that the effect of x_n on the occurrence of the LW is not significant at the 5% level.

The analysis shows that (i) Year of Birth, (ii) Education and (iii) Education×Style is significant. The coefficient of determination (R^2) is 0.67.⁶⁵

First, we can see that Year of Birth has a positive effect on the occurrence of the LW, which indicates that the use of the LW increases as speakers' year of birth does. The interaction between Year of Birth and Style is not statistically significant.

Second, there is an interaction between Education and Style. This means that the effect of Education, in spite of its statistical significance, should not be discussed alone, because it is modified by the stylistic factor. The interpretation of the effect of this interaction term on the occurrence of the LW is illustrated in the following section.

⁶⁵ Goodness of fit for *sapooto* in Table 7.8 ($R^2=0.67$) is higher than that for *keesu* in Table 6.12 ($R^2=0.33$) and in Table 6.24 ($R^2=0.24$). It seems that this results from the difference in the nature of the data used here. First, the sample size in the first case ($N=2,901$) is much larger than that in the latter two cases ($N=1,405$ and $N=725$). Second, the data examined for *sapooto* is well controlled and well balanced in terms of the factors that are useful to a sociolinguistic study, e.g., linguistic context, style, age, and education, taking advantage of the interview method. In contrast, the data examined for *keesu*, collected from the two corpora, is not as well controlled for linguistic context nor balanced in terms of the author's/speaker's social attributes since the examined corpora are not specifically designed for sociolinguistic studies. Thus, the samples from the corpora show more discrepancy between the observed and predicted values than the samples from the interviews.

Table 7.8 Result of the logistic regression analysis: *sapooto* (interview)

	B	S.E.	Wald	df	Sig.	Exp(B)
Year of birth	.0258	.002	147.806	1	**	1.026
Education				2	**	
JHS			121.865			
HS	.8462	.112	56.873	1	**	2.331
UNIV	1.3366	.121	121.690	1	**	3.806
Style	0					
Sex	0					
Year of birth×Style	0					
Education×Style			56.819	2	**	
JHS×Friends						
HS×Friends						
UNIV×Friends						
JHS×Public						
HS×Public	-.1174	.081	2.112	1		.889
UNIV×Public	-.7431	.100	54.712	1	**	.476
Constant	-51.9468	4.121	158.910	1	**	<.001
$R^2 = 0.67$					**	$p < 0.01$

7.4.3. Predicting change

This section illustrates how each independent factor affects the probability of the occurrence of the LW and, at the same time, predicts in what manner a change will occur in future. This prediction is not only for those aged 20s to 70s, but also for those excluded from the analysis in 7.3.1, i.e., those in the age groups 15-19, 80-89 and 90-94. The median values of each age range assigned to the equation are: 1912 (90-94), 1919.5 (80-89), 1929.5 (70-79), 1939.5 (60-69), 1949.5 (50-59), 1959.5 (40-49), 1969.5 (30-39), 1979.5 (20-29) and 1987 (15-19). In addition to that, future generations, i.e., those born in 1997 and 2007, 10 and 20 years after the youngest surveyed group are also taken into consideration. This enables us to predict variation across 100 years of age.

Figure 7.9 shows the predicted probabilities of the occurrence of the LW by Education, Style and Year of Birth. The predicted values were calculated with the

regression equation shown in [Formula 1']. The lines with filled marks represent the predicted probabilities of the occurrence of the LW, and the empty marks represent the observed values. The round marks (○/●) represent the occurrence of the LW in Friends and the square marks (□/■) represent the same in Public. It should be noted that there are two observed values for each birth-year group as samples are divided by Sex. The predicted values are the same for both sexes as the effect of Sex has been shown not to be significant.

Note that JHS shows more discrepancy between the predicted and observed values than HS or UNIV, apparently because JHS is more sensitive to the effect of outliers due to the relatively smaller sample size among the three (the sample sizes of JHS, HS and UNIV are N=518, N=1458 and N=885, respectively. See Table 7.4). Among the younger age groups, that is, those in their 20s, 30s and 40s, i.e., those born in the 1960s-1980s, is especially rare for their highest obtained education to be JHS. We observe, accordingly, the largest discrepancy in these age groups in JHS.

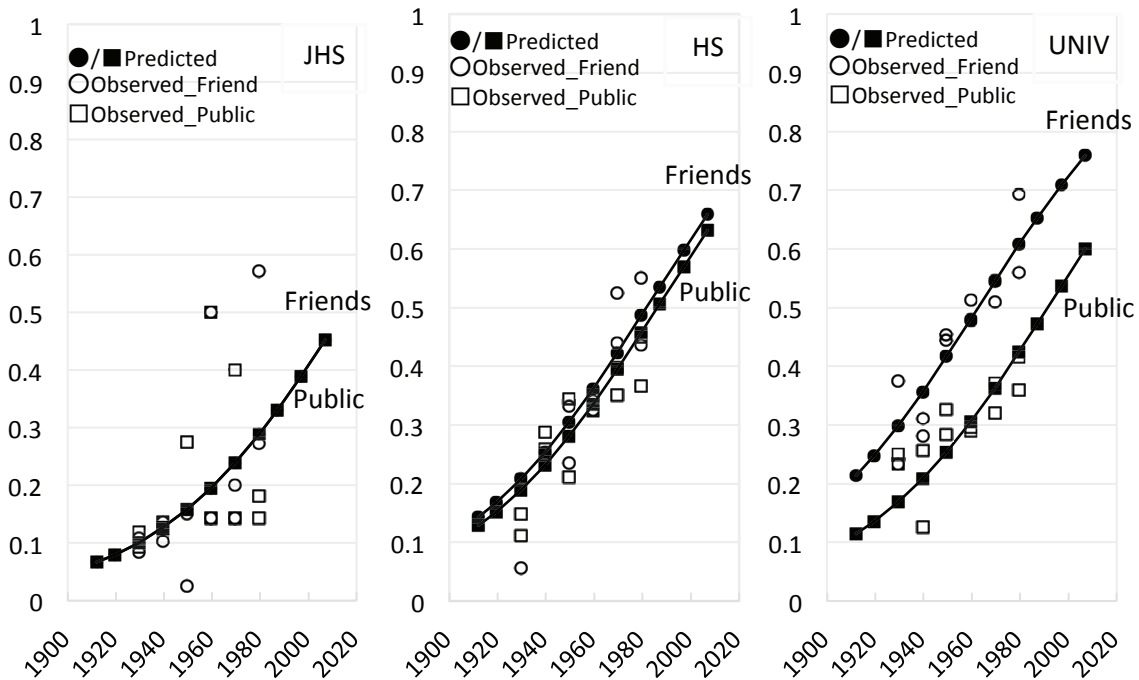


Figure 7.9 Probability of *sapoto* by Education, Style and Year of Birth

The figure shows that the probability of the occurrence of the LW increases from older generations to younger ones in every educational and stylistic category, forming part of an s-curve. However, such linguistic change progresses at different rates depending on the combination of [Education×Style], as a result of the interaction between these variables.

JHS exhibits no stylistic difference in the use of the LW: The two curves completely overlap each other. In HS, there is a very small difference in the use of the LW between Friends and Public, though the difference is shown not to be significant. Notice that [HS×Public] in Table 7.8 does not show significance at the 5% level. On the other hand, the stylistic effect comes into play in UNIV, where the probability is approximately 0.1–0.2 lower in Public than in Friends. Figure 7.9 shows that loanword adoption is greatest in the category [UNIV×Friends]. It appears that this group is reaching the

advanced stage of the s-curve, where the rate of change starts to slow down at the youngest generations, such as those born in 1997 and 2007. Such deceleration is not observed in the other groups. In [UNIV×Public], in contrast, the spread of the LW is delayed for 20-30 years: the diffusion reaches 0.6 among those born in 1979 in [UNIV×Friends] whereas it reaches the same level among those born in 2007 in [UNIV×Public].⁶⁶

Figure 7.10 illustrates the relation between Education and Style in the case of those in the 30s age group by showing predicted values. As seen above, those in [UNIV×Friends] shows the greatest adoption of the LW, though it drops largely in [UNIV×Public] to a level slightly lower than that of [HS×Public]. Consequently, the ranking of the educational categories seen in Friends, i.e., JHS < HS < UNIV, is no longer maintained in Public. This implies that the effect of Education varies as a function of Style.

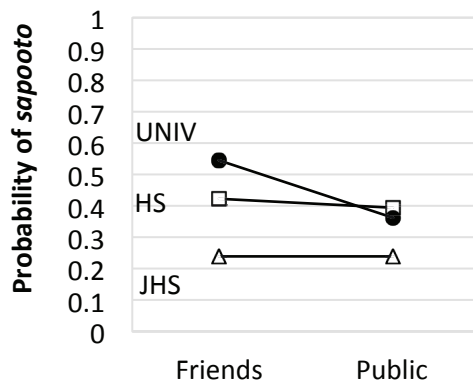


Figure 7.10 Interaction between Education and Style (those in their 30s)

⁶⁶ We can compare this prediction with real-time data in 2028 or later as those born in 2007 are expected to graduate from university when they are age 21 or older.

7.4.4. Summary of quantitative analysis

The multivariate analysis has confirmed statistical significance of the effect of speaker's year of birth (or age) on the use of the LW. The age differentiation indicates that there is an ongoing change toward more frequent use of the LW. It is shown, however, that the use of the LW is context-bound: people use the LW quite selectively depending on where the utterance is taking place or to whom it is directed. The change, as a result, spreads at different rates across speech style: informants' readiness to use the LW becomes weaker when talking in public than when chatting with friends. The stylistic variable, however, is shown to work in interaction with the educational one and that such a tendency is prominent only among the most educated group. As a result, the present study has succeeded in providing another piece of evidence in support of the claim made in the previous variationist studies that it is often the case that style interacts with other social variables (e.g. class, education, sex) (Labov 1972, Trudgill 1974, Sano 2008, Sano 2011, etc.).

Following Inoue (2000) (see also Figure 7.1), Figure 7.11 is provided to illustrate a three-dimensional model of the loanword diffusion (in the case of those born in 1987). It succeeds in capturing the spread of the loanword *sapooto* in a three-dimensional space by Style and Education in interaction.

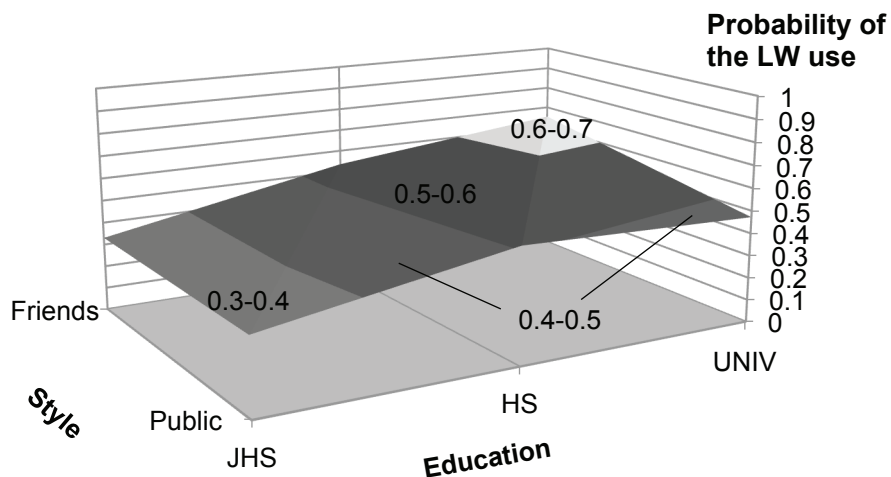


Figure 7.11 A three dimensional model of the diffusion of *sapooto* (those born in 1987)

7.5. Qualitative analysis of stylistic variation

This section takes a qualitative approach to the phenomenon. It aims to describe the mechanism of style-shifting, i.e., the reason why such shifting occurs, by contrasting those who alternate between LW and non-LW across styles with those who do not.

7.5.1. Distribution of style-shifting

We first observe what types of word-switching occur when speech style shifts, say, from Friends to Public.⁶⁷ Table 7.9 shows that there are 9 (3×3) types of switching/non-switching (excluding the cells “Not sure”). The shaded cells, i.e., those numbered as [2], [3], [4], [6], [7], [8], represent the groups that had style shifting in lexical choice. In total, half of the informants made some sort of switch across styles (48.9%). The major patterns are the switches from non-SJ to SJ, including those from

⁶⁷ Shifting does occur in the opposite direction too. Here, the more informal setting “Friends” is interpreted as the reference (default) category since an interpretation of the results seems to become more straightforward when we look at the shift from the less formal setting (Friends) to the more formal setting (Public) than the other way around.

NJ to SJ ([6], 15.6%) and the switch from LW to SJ ([4], 13.8%). The switch from LW to non-LW is 17.5% in total ([4], 13.8% + [7], 3.7%) whereas the switch from non-LW to LW is 11.7% in total ([2], 4.6% + [3], 7.1%). Because of the overall decrease in the proportion of LW use from 33.6% in Friends to 27.7% in Public, it is easy for us to assume that the switch from LW to non-LW occurred. However, the proportion of switching in the opposite direction, i.e., from non-LW to LW, is not negligible.

Table 7.9 Patterns of style-shifting

		Public				
	LW	SJ	NJ	Not sure	Total	
LW	[1]	[4]	[7]			
	461 15.9%	401 13.8%	107 3.7%	6 0.2%	975 33.6%	
SJ	[2]	[5]	[8]			
	132 4.6%	500 17.2%	121 4.2%	3 0.1%	756 26.1%	
Friends	[3]	[6]	[9]			
	206 7.1%	453 15.6%	429 14.8%	26 0.9%	1114 38.4%	
Not sure						
	4 0.1%	6 0.2%	4 0.1%	42 1.4%	56 1.9%	
Total						
	803 27.7%	1360 46.9%	661 22.8%	77 2.7%	2901 100.0%	

Table 7.10 shows the proportion (%) of types of switching (to non-switching) by education. It is also visualized in Figure 7.12 with a little simplification, where SJ and NJ are grouped together as “non-LW.”

Figure 7.12 shows that the total proportion of switching from LW to non-LW, of which the majority comes from the switching to SJ, increasing from JHS to UNIV. On the other hand, the occurrence of switching from non-LW to LW remains more or less at 10% for all educational categories. Focusing on the difference between switching to

non-LW (LW→non-LW) and switching to LW (non-LW→LW), it is less than 3% in JHS and HS, whereas the difference is outstanding (17%) in UNIV. A clear stylistic difference in LW use among UNIV is then attributed to the predominance of switching to non-LW over switching to LW. It also shows that the smaller amount of style-shifting in HS and JHS does not mean that no style-shifting occurred in these groups. Both types of switching in fact occurred, but canceled each other out.

Table 7.10 Patterns of shifting by Education (%)

	Friends→Public	JHS	HS	UNIV	Others	Total
[1]	LW→LW	6.2	17.1	19.9	7.5	15.9
	LW→non-LW	5.6	15.7	28.0	5.0	17.5
[4]	LW→SJ	3.5	12.1	23.1	5.0	13.8
[7]	LW→NJ	2.1	3.6	5.0	0.0	3.7
	non-LW→LW	7.9	13.2	11.0	17.5	11.7
[2]	SJ→LW	1.5	4.9	5.4	10.0	4.6
[3]	NJ→LW	6.4	8.3	5.5	7.5	7.1
	non-LW→non-LW	73.4	50.9	40.3	60.0	51.8
[5]	SJ→SJ	18.0	15.5	19.1	30.0	17.2
[8]	SJ→NJ	6.2	3.9	3.5	2.5	4.2
[6]	NJ→SJ	14.5	17.8	12.8	15.0	15.6
[9]	NJ→NJ	34.7	13.7	5.0	12.5	14.8
	Others	6.9	3.0	0.8	10.0	3.1
	Total	100.0	100.0	100.0	100.0	100.0
	(N)	(518)	(1458)	(885)	(40)	(2901)

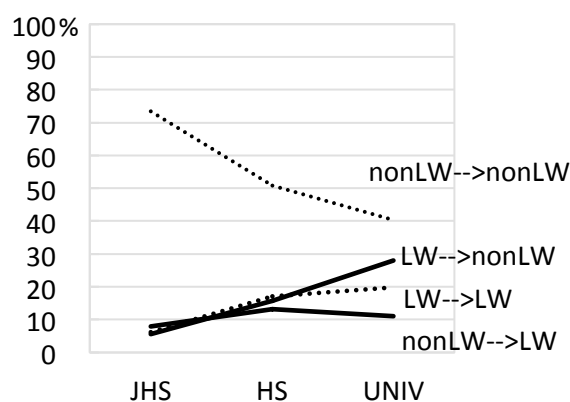


Figure 7.12 Patterns of shifting by Education (simplified)

7.5.2. Mechanism of switching from LW to non-LW

In this section we will explore the mechanism of switching from LW to non-LW. It is assumed that, among those who choose LW in Friends, those who do not show style-shifting in Public, and those who do show switching in Public each have different reasons for their word choice in Public.

First, let us focus on those who chose LW in Friends ([1][4][7] in Table 7.10). Tables 7.11 and 7.12 demonstrate that the reasons for word choice in Public are different between those who show style-shifting to SJ ([4] LW→SJ) and those who do not ([1] LW→LW), with the difference significant at the 0.1% level in the chi-square test.⁶⁸ The residual analysis shows that the former puts more emphasis on “easy to understand” and “precise in meaning” in A (Table 7.11) and “intelligent” and “dignified” in B (Table 7.12) than the latter. “Easy to understand” and “precise in meaning” seem to represent the speakers’ consideration for the addressee. The other two, “intelligent” and “dignified,” in contrast, seem to reflect speakers’ attention to the formality of the setting, in which they are required to sound proper for the occasion. In summary, those who switch from LW to SJ seem to consider SJ more suitable to use in Public than LW in terms of addressee’s ease of understanding and the formality of the setting.

⁶⁸ It is noted that those who chose “No applicable choice” and “Not sure” are excluded from the analysis. Consequently, the total number of each switching group is not equal to that in Table 7.9.

Table 7.11 Reason for word choice in Public (A) (LW→LW vs. LW→SJ)

A		concise	easy to understand	precise	familiar	Total
[1]LW→LW	N	122	147	18	170	457
	%	26.7%	32.2%	3.9%	37.2%	100.0%
	Residual	5.5	-5.8	-10.9	11.6	
[4]LW→SJ	N	47	206	128	17	398
	%	11.8%	51.8%	32.2%	4.3%	100.0%
	Residual	-5.5	**5.8	**10.9	-11.6	
Total	N	169	353	146	187	855
	%	19.8%	41.3%	17.1%	21.9%	100.0%

$$X^2(3)=248.315, p<0.001$$

** significant at 1% in residual analysis

Table 7.12 Reason for word choice in Public (B) (LW→LW vs. LW→SJ)

B		novel	intelligent	stylish, fashionable	colloquial	dignified	Total
[1]LW→LW	N	121	52	33	152	9	367
	%	33.0%	14.2%	9.0%	41.4%	2.5%	100.0%
	Residual	9.7	-6.0	4.2	-.4	-6.9	
[4]LW→SJ	N	11	106	5	135	57	314
	%	3.5%	33.8%	1.6%	43.0%	18.2%	100.0%
	Residual	-9.7	**6.0	-4.2	.4	**6.9	
Total	N	132	158	38	287	66	681
	%	19.4%	23.2%	5.6%	42.1%	9.7%	100.0%

$$X^2(4)=163.536, p<0.001$$

** significant at 1% in residual analysis

Second, Tables 7.13 and 7.14 demonstrate that the reasons for word choice in Public among those who show style-shifting to NJ ([7] LW→NJ) are different from the reasons given by those who do not ([1] LW→LW), with the difference significant at the 0.1% level in the chi-square test. The data shows that the former group consider “easy to understand” in A (Table 7.13) and “colloquial” in B (Table 7.14) more important than the latter.⁶⁹ Both “easy to understand” and “colloquial” seem to reflect the speakers’

⁶⁹ “Precise” in A (Table 7.13) and “dignified” in B (Table 7.14) are excluded from the analyses, because these categories include cells in which the expected number of occurrence is 5 and under. By the same token, “dignified” in B is excluded from the analysis in Table 7.18.

consideration for the addressee. In summary, those who switch from LW to NJ seem to consider NJ more suitable to use in Public than LW in terms of addressee's ease of understanding.

Table 7.13 Reason for word choice in Public (A) (LW→LW vs. LW→NJ)

A		concise	easy to understand	familiar	Total
[1]LW→LW	N	122	147	170	439
	%	27.8%	33.5%	38.7%	100.0%
	Residual	4.4	-9.1	5.7	
[7]LW→NJ	N	7	83	9	99
	%	7.1%	83.8%	9.1%	100.0%
	Residual	-4.4	**9.1	-5.7	
Total	N	129	230	179	538
	%	24.0%	42.8%	33.3%	100.0%

$$X^2(2)=83.695, p<0.001$$

** significant at 1% in residual analysis

Table 7.14 Reason for word choice in Public (B) (LW→LW vs. LW→NJ)

B		novel intelligent	stylish, fashionable	colloquial	Total
[1]LW→LW	N	121	52	33	152
	%	33.8%	14.5%	9.2%	42.5%
	Residual	5.5	2.2	3.0	-8.0
[7]LW→NJ	N	4	5	0	79
	%	4.5%	5.7%	0.0%	89.8%
	Residual	-5.5	-2.2	-3.0	**8.0
Total	N	125	57	33	231
	%	28.0%	12.8%	7.4%	51.8%

$$X^2(3)=64.533, p<0.001$$

** significant at 1% in residual analysis

7.5.3. Mechanism of switching from non-LW to LW

Now let us turn to those who chose non-LW in Friends but switch to LW in Public ([2] and [3] in Table 7.10). Tables 7.15 and 7.16 demonstrate that reasons for this word choice in Public are different between those who show style-shifting ([2] SJ→LW) and

those who do not ([5] SJ→SJ), with the difference significant at the 0.1% level in the chi-square test. The residual analysis shows that the former group places more stress upon “concise” in A (Table 7.15) and “novel” and “stylistic/fashionable” in B (Table 7.16) than the latter. Evaluations like “concise” seem to indicate consideration of the speakers’ ease of use as well as their consideration for the addressee. “Novel” and “stylistic/fashionable” seem to be associated with the question of how the speakers want to express themselves in front of unspecified general audience. In summary, those who switch from SJ to LW seem to consider LW more suitable to use in Public than SJ in terms of the ease of use for the speakers themselves, the ease of understanding of the addressee, and the speakers’ identity construction.

Table 7.15 Reason for word choice in Public (A) (SJ→SJ vs. SJ→LW)

A		concise	easy to understand	precise	familiar	Total
[5]SJ→SJ	N	47	281	70	98	496
	%	9.5%	56.7%	14.1%	19.8%	100.0%
	Residual	-8.1	4.0	1.7	0.9	
[2]SJ→LW	N	50	48	11	21	130
	%	38.5%	36.9%	8.5%	16.2%	100.0%
	Residual	**8.1	-4.0	-1.7	-0.9	
Total	N	97	329	81	119	626
	%	15.5%	52.6%	12.9%	19.0%	100.0%

$X^2(3)=66.726, p<0.001$

** significant at 1% in residual analysis

Table 7.16 Reason for word choice in Public (B) (SJ→SJ vs. SJ→LW)

B		novel	intelligent	stylish, fashionable	colloquial	dignified	Total
[5]SJ→SJ	N	13	77	11	184	32	317
	%	4.1%	24.3%	3.5%	58.0%	10.1%	100.0%
	Residual	-8.9	-1.0	-3.5	7.5	1.5	
[2]SJ→LW	N	41	33	14	19	6	113
	%	36.3%	29.2%	12.4%	16.8%	5.3%	100.0%
	Residual	**8.9	1.0	**3.5	-7.5	-1.5	
Total	N	54	110	25	203	38	430
	%	12.6%	25.6%	5.8%	47.2%	8.8%	100.0%

$X^2(4)=113.043, p<0.001$

** significant at 1% in residual analysis

Tables 7.17 and 7.18 demonstrate that the reasons for word choice in Public are different between those who make style-shifting ([3] NJ→LW) and those who do not ([9] NJ→NJ), with the difference significant at the 0.1 % level in the chi-square test. The residual analysis shows that the former group considers “concise” and “precise in meaning” in A (Table 7.17) and “novel,” “intelligent” and “stylistic/fashionable” in B (Table 7.18) more important than the latter. In summary, those who switch from NJ to LW seem to consider LW more suitable to use in Public than NJ in terms of the ease of use for the speakers themselves or the addressees, the speakers’ identity construction, and the formality of the setting.

Table 7.17 Reason for word choice in Public (A) (NJ→NJ vs. NJ→LW)

A		concise	easy to understand	precise	familiar	Total
[9]NJ→NJ	N	18	292	21	96	427
	%	4.2%	68.4%	4.9%	22.5%	100.0%
	Residual	-10.9	8.8	-4.9	2.3	
[3]NJ→LW	N	76	63	34	30	203
	%	37.4%	31.0%	16.7%	14.8%	100.0%
	Residual	**10.9	-8.8	**4.9	-2.3	
Total	N	94	355	55	126	630
	%	14.9%	56.3%	8.7%	20.0%	100.0%

$$X^2(3)=161.986, p<0.001$$

** significant at 1% in residual analysis

Table 7.18 Reason for word choice in Public (B) (NJ→NJ vs. NJ→LW)

B		novel intelligent	stylish, fashionable	colloquial	Total
[9]NJ→NJ	N	3	8	5	293
	%	1.0%	2.6%	1.6%	94.8%
	Residual	-10.7	-9.5	-5.4	17.5
[3]NJ→LW	N	65	61	25	32
	%	35.5%	33.3%	13.7%	17.5%
	Residual	**10.7	**9.5	**5.4	-17.5
Total	N	68	69	30	325
	%	13.8%	14.0%	6.1%	66.1%

$$X^2(3)=308.116, p<0.001$$

** significant at 1% in residual analysis

7.5.4. People's evaluation of the loanword

In-depth observation of individual speakers reveals that there are actually two types of style-shifting moving in opposite directions. Shifting from LW to non-LW tends to occur when speakers want to add intelligence and dignity to their speech by using SJ in deference to the formality of the setting. It also occurs when speakers want to make understanding easier for the addressee(s) by using non-LW, which are characterized as “easy to understand” and “colloquial.”

On the other hand, switching from non-LW to LW tends to occur when speakers want to place emphasis on the “novelty” and “stylishness” of their speech by selecting LW. In addition, speakers’ intention of adding “intelligence” to their speech in deference to the formality of the setting sometimes induces a switch from NJ to LW.

7.5.5. Interpretation of style

Based on the discussion above, we should revisit the question of what actually induces style-shifting in the present data.

First, there is evidence that style-shifting occurs as a response to the speakers’ attention to their speech (Attention to Speech Model). This includes the cases where a speaker chooses LW in Friends but switches to SJ in Public in order to add “intelligence” or “dignity” to their speech in response to the formality of the setting. Based on the fact that SJ is in general associated with social prestige (Ishiwata 2001: 9), we can conclude that an increase in the level of formality of the setting leads to switching to the more prestigious form, SJ. Evaluations like “intelligent” or “dignified,” however, are also interpretable as speakers’ spontaneous acts of developing their identity in public speaking (Speaker Design Model). That is to say, it is possible that style-shifting occurs when the speaker wants to sound intelligent or dignified in front of an anonymous general audience.

Second, some data show that style-shifting occurs as a response to addressees (Audience Design Model). For example, there are cases where a speaker chooses LW in Friends but switches to non-LW in Public in order to help their addressee understand the speaker more easily. In general, a speaker feels a need to consider their addressees more strongly when talking to an anonymous general audience in public than when

chatting with well-acquainted friends, as they cannot expect as much linguistic homogeneity when talking to the former group as when talking to the latter group. It appears that speakers tend to become more careful about the use of LW when talking to a variety of people in a setting like Public, considering, for example, that not all generations are equally familiar with LW.

The above descriptions point out that the stylistic difference between the two settings, which is often intuitively contrasted in terms of formality (7.2.3), must in fact be interpreted from multiple perspectives. It is therefore important to apply several models complementarily or simultaneously to the data in order to gain a more comprehensive understanding of the nature of style and style-shifting.

7.6. Conclusion

This chapter has examined how a stylistic variable influences the diffusion of the LW *sapooto* as opposed to its native equivalents. The first goal was to provide a multivariate model that includes style as its independent variable. The data showed that the diffusion of the loanword is interpreted as an ongoing change with stylistic variation. It also revealed that a stylistic variable comes into play in interaction with the educational variable: Stylistic difference in word choice is significant only for the most educated group. The multivariate analysis enables us to provide a “Water Tank Model” (Inoue 2000), in which the diffusion of the loanword is captured multi-dimensionally by stylistic and other variables (Figure 7.11). The model is shown to be useful in predicting probabilities of the phenomenon for future generations as well as those excluded from

analysis due to lack of data, verifying its practical utility as a “tool for predicting probability” (Yokoyama & Sanada 2007: 82).

The second goal of this chapter was to elucidate why such stylistic variation occurs. The analysis revealed that when speech style shifts from Friends to Public, shifting from LW to non-LW occurs as the most prominent pattern, but at the same time, a significant minority of people switch from non-LW to LW. Their motives for intra-speaker variation tell us that both types of style-shifting are not fully explainable by a single traditional model and need to be interpreted from various perspectives.

To conclude, examining stylistic variation both quantitatively and qualitatively serves as a clue to elucidate the elaborate process of loanword diffusion. The present result is expected to be applicable to other well-attested loanwords to a certain extent, as mentioned earlier.

8. A real-time approach to the diffusion of loanwords

8.1. Introduction

8.1.1. Aim: Towards more precise predictions of change

This chapter attempts a real-time approach to the diffusion of loanwords. The previous chapters have provided statistical evidence of a change in apparent time toward more adoption of the loan variants. This enables us to predict the existence, direction and rate of change. In this chapter, however, we will see that real-time data exhibits a process of loanword diffusion that is different from what the apparent-time approach predicts (in terms of direction and rate of change).

8.1.2. Background: The need for a real-time approach to lexical variation

In spite of the fact that “apparent time” is “a truly powerful concept in locating the presence of change” (Sankoff, G. 2006: 113, see also Bailey et al. 1991), a number of studies reveal that the apparent-time approach has its limitations.

First, as already mentioned in Chapter 4, age gradient is only an indication, but not evidence, of change in progress. In order to interpret a given age differentiation as an ongoing change, the evidence from real-time observation is essential.

In the study of language change in Japanese, there are two representative series of real-time studies: the study of dialect standardization in Tsuruoka (ISM & NINJAL 2014, 2015) and the study of the use of honorifics in Okazaki (NINJAL 2010a-d). The survey of standardization in Tsuruoka, for example, has been conducted every two decades since 1950 (followed by 1971, 1991 and 2011).⁷⁰ It succeeds in capturing the process of language change in the community in real time on a long-term basis. Figure

⁷⁰ The Okazaki survey was conducted in 1953, 1972 and 2008. (NINJAL 1957, 1983, 2010a-d)

8.1 illustrates the diachronic process of phonological standardization in the community, in which the non-standard (centralized) form of the variable /u/ (as in *karasu* ‘a crow’) is being replaced by its standard form over the 60 years between 1950 and 2011. The degree of standardization of the variable ranges 40 – 65% across generations in 1950 but it seems to have nearly completely standardized by 2011 with most age groups selecting the standard form 100% of the time. We see that the four s-shaped curves form a larger s-curve as a whole, demonstrating that the change in question is in progress in the community overall.

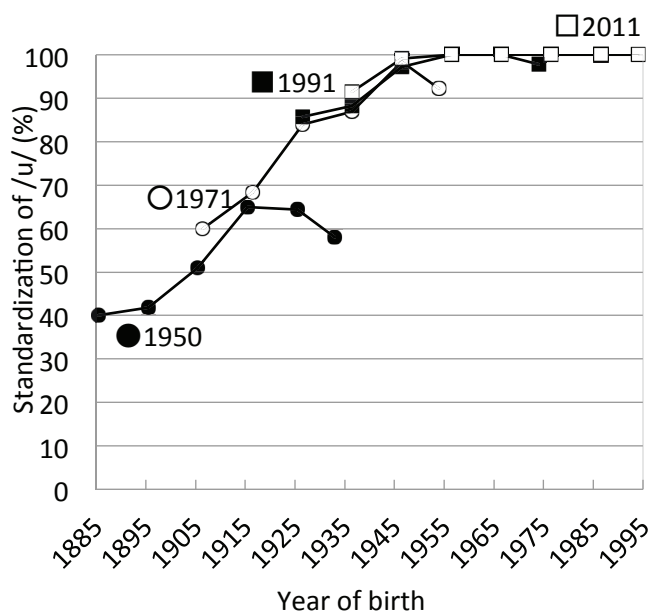


Figure 8.1 Standardization of the variable /u/ in Tsuruoka, Japan (ISM & NINJAL 2014: 51)

Second, the apparent-time approach often fails to capture transition across speakers’ lifetimes (Labov 1982: 67). Figure 8.1 illustrates the case in which a distribution in apparent time is “a faithful reproduction of changes in real time” (ibid: 67). The apparent-time hypothesis predicts the rate of change quite accurately in the case of the

variable /u/ in Tsuruoka (Figure 8.1). Sankoff, G. (2006: 115), however, says that apparent-time observations generally underestimate the rate of change. For example, Pope, Meyerhoff & Ladd (2007: 622-623) points out, in their follow-up study of Labov (1963, revised as 1972: Ch1), that phonological change in Martha’s Vineyard is progressing at a higher rate than the apparent-time distribution indicates. Such acceleration occurs when individuals change their language over their lifetimes in the direction of a change in progress in the community as a whole (Sankoff, G. 2005: 1011). This type of change is called “lifespan change.” In Table 8.1 this is added as a new category of change to refine the classification of language change originated in Labov (1994: 83).

Table 8.1 Types of change defined by (i) age differentiation, (ii) variability of the individuals and (iii) variability of the community (Meyerhoff 2011: 153, modified by the author. See also Labov 1994: 83 and Sankoff, G. 2005)

	Age differentiation	Variability: Individuals	Variability: Community
1 Stability	–	–	–
2 Age-grading	+	+	–
3 Lifespan change (Sankoff, G. 2005)	+	+	+
4 Generational change	+	–	+
5 Communal change	–	+	+

Meanwhile, Yokoyama & Sanada (2010: 40) demonstrates that the rate of change is not only accelerated but also decelerated in the course of time. They show that deceleration of change occurs when individual speakers change in the direction opposite to a change in progress in the community as a whole (e.g., towards “dialectalization” as opposed to “standardization”). Decelerated change is also attested to by the example of

verbal conjugation in Japanese (Matsuda 2013: 87). This type of change, by definition, also falls into a category of lifespan change. Yokoyama et al. (2008: 40) highlights the presence of “life-long memory,” i.e., memory accumulated after one’s critical period, which induces acceleration/deceleration of change. This sort of memory is considered to be dependent on aging and a trend of the times. They suggest that we pay more attention to the effect of these factors in predicting change in order to correct underestimation or overestimation of the rate of change.

In summary, as Table 8.1 shows, there are three possibilities when we observe age differentiation in apparent time. First, it can be interpreted as *age-grading* if the community overall is invariable. Second, it is interpreted as *generational change* if the community as a whole changes but individual speakers remains invariable. The phonological change in Tsuruoka in Figure 8.1 is an example of this type of change. The Third possibility is *lifespan change*, in which both the whole community and individuals are variable. This type of change has been witnessed in phonological change (Pope et al. 2007, Yokoyama & Sanada 2010) but it is said to be well-attested for vocabulary too (Meyerhoff 2011: 152).

The above discussion puts an emphasis on the need for the real-time observation of a phenomenon in order to make a more accurate prediction of language change. In the two well-known real-time studies in Japan (Tsuruoka and Okazaki), the communities were observed at more than three different points of time over 50 to 60 years. These longitudinal studies enable us not only to verify the existence of change but also to

make more accurate predictions of change, paying attention to variability of the community as well as individuals over time.⁷¹

8.1.3. Research questions

With regard to the studies of loanword diffusion, a number of nationwide large-scaled surveys on loanwords have been conducted at different points of time (BBK 2000, 2003, 2008, 2013; NINJAL 2004, 2005a). However, there has never been a sufficient systematic attempt to integrate the results of these surveys into a comprehensive whole within a framework of the real-time approach. This chapter, therefore, aims to answer the following two questions:

The first question concerns whether the attested age gradient in the previous chapters is indeed interpretable as generational change (change in progress). This can be done by observing the variability of the community as a whole, i.e., change across the year of survey when the effect of speakers' generation at the time of survey is fixed.

The second one concerns whether individual speakers are stable over their lifetimes in their receptiveness to loanwords as the Apparent-time Hypothesis predicts, provided that the change is indeed in progress. Considering a general assumption that lexical-level variables are more susceptible to change over one's lifetime than phonological variables (Sankoff, G. 2005: 1010-1011, Meyerhoff 2011: 148), the hypothesis should be applied to lexical variation with care. In the present study, then, more attention will have to be paid to the variability of individual speakers after their critical period. This can be done by examining the effect of year of survey when

⁷¹ For the most recent (4th) survey of Tsuruoka, see ISM & NINJAL (2014, 2015). For the most recent (3rd) survey of Okazaki, see NINJAL (2010a-d). The examination of these data will help evaluate the validity of predictions in Yokoyama & Sanada (2010) and Yokoyama et al. (2008).

speakers' year of birth is fixed, because speakers' critical period is dependent on their year of birth.

8.2. Data

8.2.1. Interview-based data

There is no real-time study of loanwords that is as systematic as that of Tsuruoka or Okazaki. However, several institutes and organizations have conducted one-off surveys on loanwords since the 1970s. All we can do at this point is to pick up mutually related items from these surveys and analyze them as real-time data.

What data sets we can make use of for a real-time study are two-fold: (i) data on the use of individual lexical items and (ii) data on loanwords in general.

(i)

- Which of the two do you usually use: *wain* (< *wine*) or *budoshu*?
- Which of the two do you usually use: *oopun-suru* (< *open*) or *kaiten-suru*?

(ii)

- How do you feel about our present use of loanwords (in general) in our everyday lives?
- How do you feel about a further expansion of loanwords in the future?

As shown in Table 8.2, the sources of these items are all nationwide random sampling surveys that employ a face-to-face interview technique. The results of these surveys can thus reasonably be contrasted with each other and enable us to conduct a trend study of the whole Japanese-speaking population in Japan over decades.

Wain (< *wine*) and *oopun-suru* (< *open*) are only those that were selected as research subjects at more than three different points in time. The surveys were conducted in 1979, 1996 and 2000, and we can observe change in the community over 20 years.

Table 8.2 Sources of data on individual lexical items: *wain* and *oopun-suru*

Survey name (source)	Month/ Year of survey	Targeted populatio n	No. of valid response (%)	Ques tion No.	Diffusio n of <i>wain</i> (< <i>wine</i>)	Diffusion of <i>oopun</i> <i>-suru</i> (< <i>open</i>)
A survey on public attitudes to language (NHK1980: 5)	9/1979	those aged ≥ 16	2639 (73.3%)	---	39.0%	28.0%
The 10 th survey on the linguistic awareness: The Japanese people and spoken language (NHK1996: 55)	3/1996	those aged ≥ 20	1251 (69.5%)	Q1	77.2%	56.9%
<i>Opinion Poll on the National Language in Heisei 11(1999)</i> (BBK 2000: 18)	1/2000	those aged ≥ 16	2196 (73.2%)	Q4	75.6%	34.6%

Between a loanword and its native equivalent: *wain* vs. *budoshu* ('wine'); *oopun-suru* vs. *kaiten-suru* ('to open'), informants were asked to choose one that they use more frequently. "Diffusion of X" in Table 8.2 refers to the proportion of those who chose the loanword X as opposed to those who chose its native equivalent or "Not sure."

Table 8.3 provides information on data concerning the question of how people feel about their present use of loanwords in general in their everyday lives. This question has been repeatedly asked in annual opinion polls by BBK between the 2000s and 2010s (2000, 2002, 2008 and 2013). The same question was also asked in the opinion poll

conducted back in 1977 by NSDKK. The whole data set enables us to examine change in the public attitude in the community over 35 years.

Table 8.3 Sources of data on like or dislike of the present loanword use

Survey name (source)	Month/ Year of survey	Targeted populatio n	No. of valid response (%)	Que stion No.	Positive (%)		
					Like (%)	Neither (%)	
<i>Opinion Poll on the National Language</i> (NSDKK 1977:140)	8/1977	those aged ≥ 20	8170 (81.7%)	Q22	10.0	49.0	59.0
<i>Opinion Poll on the National Language in Heisei 11(1999)</i> (BBK 2000:23*)	1/2000	those aged ≥ 16	2196 (73.2%)	Q6	13.3	48.8	62.1
<i>Opinion Poll on the National Language in Heisei 14(2002)</i> (BBK 2003:140)	11/2002	those aged ≥ 16	2200 (73.3%)	Q17	16.2	45.1	61.3
<i>Opinion Poll on the National Language in Heisei 19(2007)</i> (BBK 2008:144)	3/2008	those aged ≥ 16	1975 (57.3%)	Q16	14.5	43.7	58.2
<i>Opinion Poll on the National Language in Heisei 24(2012)</i> (BBK 2013:92)	3/2013	those aged ≥ 16	2153 (61.1%)	Q8	9.3	54.0	63.3

People were asked to evaluate their feelings on the recent use of loanwords in written and spoken Japanese according to a three-point scale: “(1)Agree—(2)Neither—(3)Disagree” (and “Not sure”). Note that those who chose “agree” and those who chose “neither” are grouped together as “positive” as opposed to the “disagree” group in Table 8.3, as the “agree” and “neither” groups always show similar age distribution patterns (See Appendix C1b for more detail). We notice that, on average, almost half of people (though with variance across age groups) do not display either a like or dislike of the present use of loanwords.

Table 8.4 provides information on data concerning the question of whether or not people have a positive attitude toward a further expansion of the use of loanwords in future. This question was asked in an opinion poll conducted in 1995 by BBK and another conducted in 2003 by NINJAL.

Table 8.4 Sources of data on like or dislike of a further expansion of loanwords

Survey name (source)	Month/Year of survey	Targeted population	No. of valid response (%)	Questi on No.	Positive (%)
<i>Opinion Poll on the National Language in Heisei 7 (1995)</i> (BBK 1995:48*)	4/1995	those aged ≥16	2212 (73.7%)	Q16	57.9
<i>The National Survey on Attitudes to Loanwords</i> (NINJAL 2004:16*)	10-11/2003	those aged ≥15	3087 (68.6%)	Q2	37.0

People were asked to evaluate their feelings on a further expansion of loanwords in the Japanese language in the future according to a four-point scale. Note that both surveys share the same question: “What is your overall rating for a further expansion of loanwords in the future?” but each survey uses slightly different wordings for its rating. The survey conducted in 1995 provided the choices: “(1)Further expansion is acceptable—(2)Only moderate expansion is acceptable—(3)No more expansion is acceptable—(4)They should be decreased” (and “Not sure”) whereas the one conducted in 2003 introduced the options: “(1)Agree—(2)Somewhat Agree—(3)Somewhat Disagree—(4)Disagree” (and “Not sure”). In each case, the first two choices are grouped into one that shows a “positive attitude” to the phenomenon in contrast to the last two choices that can be grouped into one that shows a “negative attitude.” We can

then observe change in the proportion of those who display a “positive attitude” toward a further expansion of loanwords in the community roughly over a decade.

8.2.2. Corpus-based data

The Library sub-corpus (LB) in BCCWJ (the variable-length components) is used as a diachronic corpus. This corpus contains books published between 1986 and 2005, so it is possible to observe language produced in real time over the two decades.

We will have a look at the distribution of both individual loanwords and the loanword stratum as a whole. Three lexical variables are selected for real-time investigations: (i) *keesu* (< *case*), (ii) *sapooto* (< *support*) and (iii) *shinpuru* (< *simple*) and their native equivalents. I have chosen one variable from nominal, verbal and adjective usage each. These loan variants have shown a relatively consistent age differentiation in real time in Chapter 5 and are thus considered to be good candidates for those in the process of change in progress.

Native equivalents of these loanwords and criteria of data elicitation are consistent with those shown in Chapter 5 (see Table 5.6 and 5.4.1), which means that a comprehensive approach has been taken. The scope of investigation is limited to a sample written by a single author whose year of birth is available. In order to reduce irregularity of small-sized samples, birth-year groups that contribute less than 5% of the whole samples are excluded (following the discussions in 5.2.2). The majority of samples are from those born between the 1920s and 1960s. Table 8.5 shows the summary of the number of tokens for each linguistic variable by publication year. The table demonstrates a general tendency for the proportion of the loan variants to increase consistently over the two decades (except for *keesu* in 1996-2000).

Table 8.5 Corpus-based data (1986-2005) (LB, BCCWJ)

LW as opposed to non-LW	Year of Publication	LW(N)	non-LW(N)	Total(N)	LW(%)
<i>keesu</i> (< <i>case</i>)		979	8328	9307	11.8
as opposed to <i>jirei</i> , <i>baai</i> and <i>rei</i>	1986-1990	105	1301	1406	8.1
	1991-1995	222	1851	2073	12.0
	1996-2000	301	2560	2861	11.8
	2001-2005	351	2616	2967	13.4
<i>sapooto</i> (< <i>support</i>)		60	1600	1660	3.8
as opposed to <i>shien</i> , <i>sasaeru</i> and <i>shiji</i>	1986-1990	1	270	271	0.4
	1991-1995	6	382	388	1.6
	1996-2000	23	516	539	4.5
	2001-2005	30	432	462	6.9
<i>shinpuru</i> (< <i>simple</i>)		89	450	539	19.8
as opposed to <i>tanjun</i> and <i>kanso</i>	1986-1990	6	70	76	8.6
	1991-1995	19	106	125	17.9
	1996-2000	30	150	180	20.0
	2001-2005	34	124	158	27.4
The LW stratum (token)		81547	2183863	2265410	3.7
as opposed to SJ, NJ ⁷² and HB	1986-1990	11825	348792	360617	3.4
	1991-1995	19332	548541	567873	3.5
	1996-2000	24781	661547	686328	3.7
	2001-2005	25609	624983	650592	4.1

8.3. Analysis

8.3.1. Examining the variability of the community overall

In this section we will examine the variability of the community as a whole when the effect of the speakers' age at the time of the surveys is fixed. See Appendices C1a-c and C2 for the full data sets examined in this section.⁷³

Figure 8.2 shows that there is a jump in the level of diffusion of *wain* (< *wine*) between 1979 and 1996, but a further rise does not occur between 1996 and 2000 –

⁷² Particles and auxiliaries are excluded from the count.

⁷³ Since proportions by sex are not always provided in the reports of these surveys, the variable sex is not investigated in this chapter.

perhaps due to a too small interval between the two surveys. We can see that the replacement of *budooshu* by *wain* is near completion among those aged 20s and below: the age gradients in 1996 and 2000 start to level off at 90-100% among these age groups.

Figure 8.3 shows that there is a big jump in the level of diffusion of the loanword *oopun-suru* (< *open*) between 1979 and 1996. There is a rise between 1979 and 2000, though a clear rise occurs only among those aged 40s or below. The problem is that the curve in 1996 is at a level notably higher than that in 2000. This leads to a conclusion that something odd occurred, i.e., the adoption of the loanword sharply rose between 1979 and 1996 but suddenly fell over the only 4-year period after that. This fall is actually explainable by the fact that the informants were provided with a particular example sentence in making their word choice in the 1996 survey.⁷⁴ It is possible that this particular context contributed to the enhanced preference for the loanword in 1996. We will soon return to this point to discuss how to deal with the effect of this in statistical analysis (Table 8.6).

⁷⁴ *Atarashii suupaa-ga {oopun, kaiten}-suru*
New supermarket-NOM open-do
'A new supermarket will be opened.'

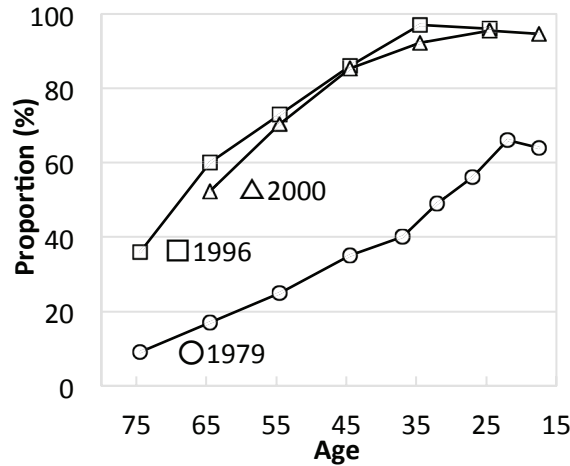


Figure 8.2 Variability of the community: *wain* (< *wine*)

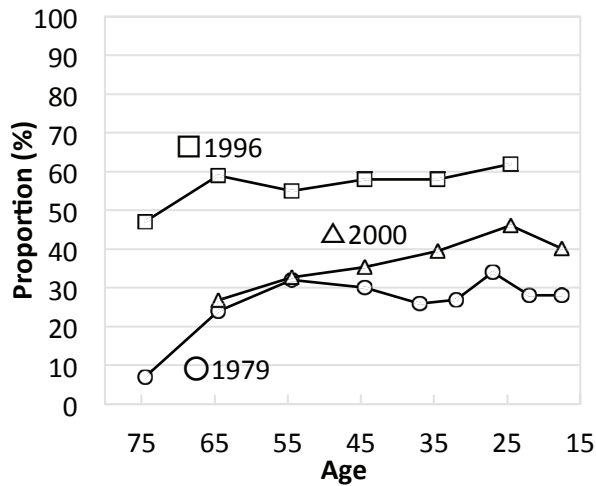


Figure 8.3 Variability of the community: *oopun-suru* (< *open*)

Figure 8.4 demonstrates an overall tendency for those who have a positive attitude toward the current use of loanwords to increase in 35 years from 1977 to 2013, although shifts between any two years are not immediately clear. Figure 8.5, however, shows a clear drop (more or less 20%) between 1995 and 2003 in the proportion of those who have a positive attitude toward a further expansion of loanwords. This is the case where, despite the fact that each age gradient indicates generational change toward more

adoption of loanwords in future, the community as a whole is moving in the opposite direction (“reversed” generational change).

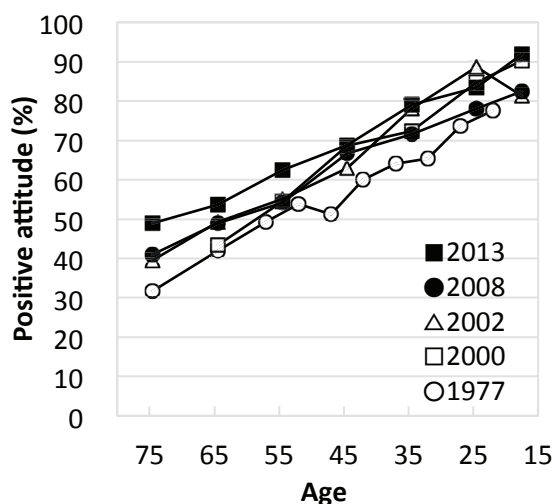


Figure 8.4 Variability of the community: The present use of LWs in general

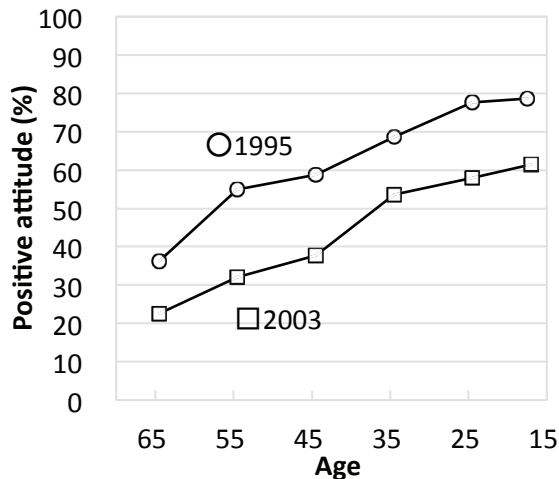


Figure 8.5 Variability of the community: A further expansion of LWs in future

Figures 8.6 – 6.9 show proportions of *keesu* (< case), *support* (< support), *shinpuru* (< simple) and the LW stratum as a whole in BCCWJ, as opposed to their native equivalents listed in Table 8.5. Due to relatively small-sized samples, the lines for individual loanwords are not as smooth as those based on the interview data above.

Diffusions of these loan variants however generally show rises along publication year (Figures 8.6 – 8.8). The loanword stratum as a whole also exhibits a jump between 1986 and 2001 (Figure 8.9).

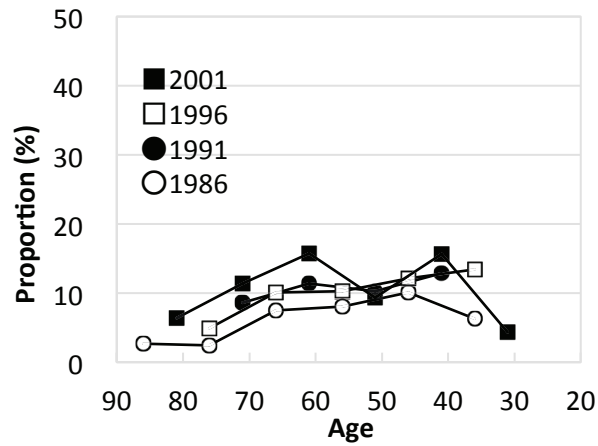


Figure 8.6 Variability of the community: *keesu* (< case)

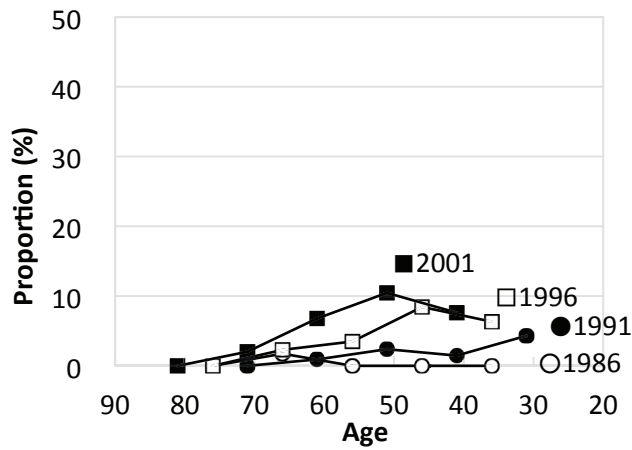


Figure 8.7 Variability of the community: *sapooto-suru* (< support)

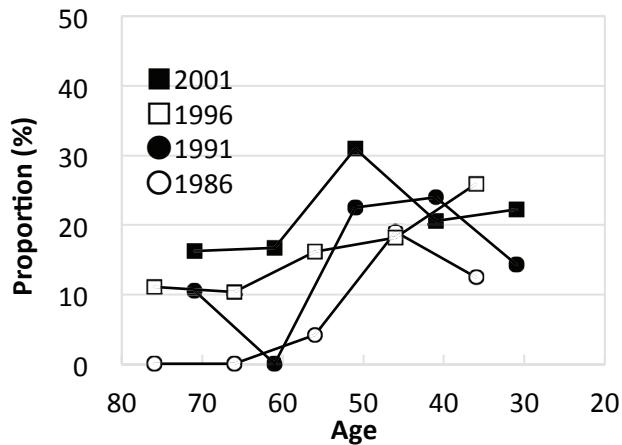


Figure 8.8 Variability of the community: *shinpuru* (< simple)

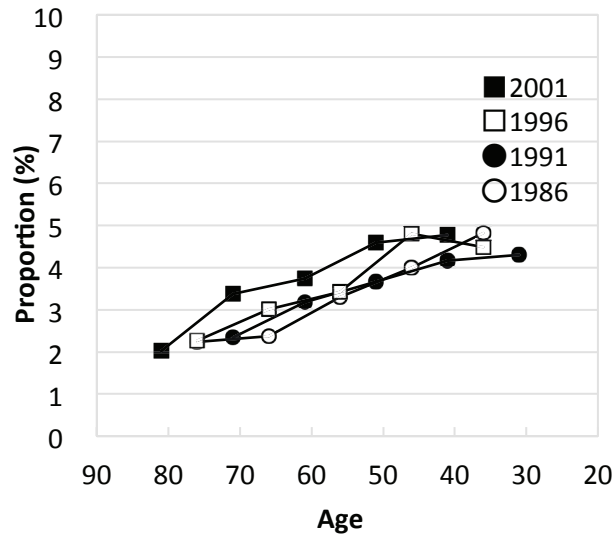


Figure 8.9 Variability of the community: The LW stratum

In order to verify the statistical significance of the variability we have observed above, a logistic regression analysis was conducted. The dependent variable is in all the cases binary: the preference for, or use of, loanwords (=1) or not (=0). Independent variables are (i) Year of Survey (Year of Publication) and (ii) Age of the speakers (at the time of survey).

The disruptive factor (iii) Context is added to the analysis of *oopun-susu*, since, as discussed above, the presence or absence of the particular context seems to have an influence on the word choice to a certain degree. The inclusion of a disruptive factor into a model helps us to elicit the effect of factors in interest (Yokoyama & Sanada 2007). (See also the discussion in 4.3.1 for the application of a disruptive factor.)

Table 8.6 shows the results of the logistic regression analysis (stepwise) with SPSS. It is shown that Year of Survey and Age are both significant (at the 5% level). The coefficient of determination (R^2) is generally high except that for *keesu*.

First, the negative coefficients for the variable Age indicate that the age effect inhibits the adoption of loanwords; in other words, preference for loanwords over native variants increases as one becomes younger.

Second, Context appears to have a positive effect on the choice of the loanword *oopun-suru*, which means that the presence of the specific context enhanced the preference for the loanword.

Third, Year of Survey is shown to have a positive influence on the preference for loanwords in most cases. This means that the receptiveness to loanwords in the community overall has been grown over time. However, this factor comes into play in the opposite direction when it comes to “Further expansion.” It indicates that the community as a whole has become less receptive to further adoption of loanwords over time. This example tells that it is possible that a change progresses in the direction opposite to what the distribution in apparent time suggests (“reversed” change).

In summary, it can be confirmed that the community as a whole is variable in all the cases.

Table 8.6 Regression coefficients of the independent variables: (i) Year of Survey and (ii) Age

	Year of Survey	Age	Context	constant	R^2
Interview-based					
<i>wain</i>	0.119189	-0.058587	NA	-233.957192	0.96
<i>oopun-suru</i>	0.022899	-0.012489	1.022264	-45.846686	0.90
Present use	0.014477	-0.038376	NA	-26.581827	0.96
Further expansion	-0.101782	-0.037763	NA	205.119082	0.97
corpus-based					
<i>keesu</i>	0.023830	-0.010205	NA	-49.121709	0.26
<i>sapooto-suru</i>	0.167374	-0.045745	NA	-334.947153	0.72
<i>shinpuru</i>	0.065205	-0.028231	NA	-130.244988	0.53
the LW stratum	0.016820	-0.017879	NA	-35.852201	0.85

8.3.2. Examining the variability of the individual speakers

This section examines the variability of individual speakers when the effect of the speakers' Year of Birth is fixed.

Figure 8.10 demonstrates a diachronic shift in the level of diffusion of the loanword *wain* (< *wine*) within the same birth-year group. There is a jump in the level of preference for *wain* between 1979 and 1996 and a small fall between 1996 and 2000. Roughly speaking, this implies that the change toward more use of the loanword is accelerated over the two decades.

Figure 8.11 shows a jump in the level of adoption of *oopun-suru* (< *open*) between 1979 and 1996 but there is no upward shift between 1979 and 2000. Considering that the effect of the context used in the 1996 survey enhances the preference for the use of the loanword, a jump between 1979 and 1996 may be disregarded. This point will be discussed further in Table 8.7. It is then implied that there is no change in the individual speakers over their lifetimes.

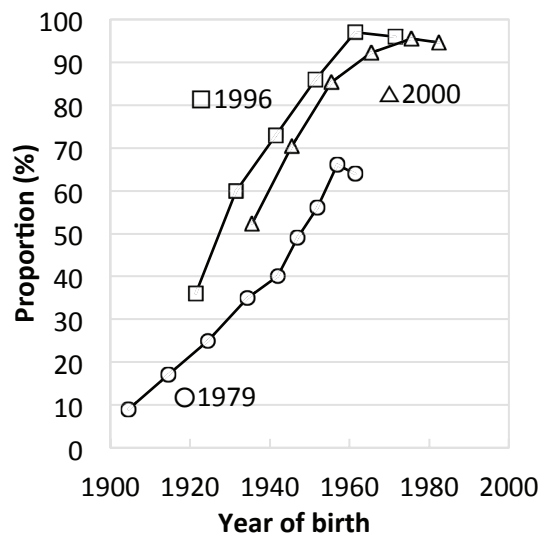


Figure 8.10 Variability of the individuals: *wain* (< wine)

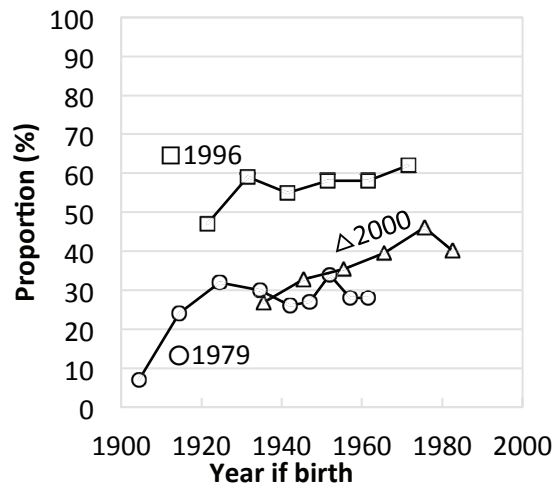


Figure 8.11 Variability of the individuals: *oopun-suru* (< open)

In Figure 8.12, the curves show downward shifts between 1977 and 2000/2002. There is also a drop between 2000/2002 and 2008/2013 among younger generations (those born in the 1960s and after), but such a tendency is not immediately clear for the older generations. This implies that the change toward more receptiveness to the current use of loanwords is decelerated over the 35 years from 1977 to 2013.

A similar downward tendency is observed in Figure 8.13, where the receptiveness to a further expansion of loanwords drops largely over about a decade. This in fact means that the *acceleration* of the change is toward *less* receptiveness to loanwords.

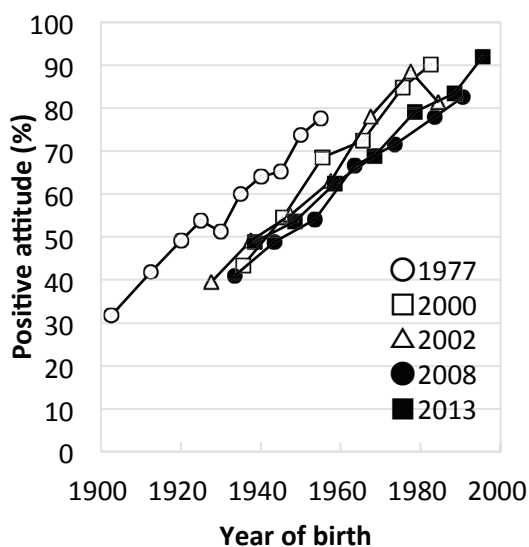


Figure 8.12 Variability of the individuals: The present use of LWs in general

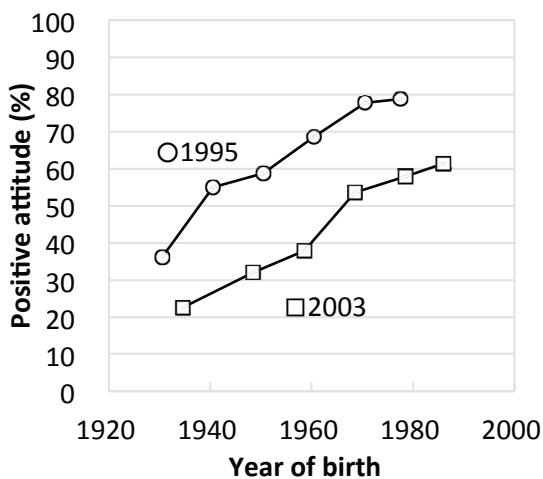


Figure 8.13 Variability of the individuals: A further expansion of LWs in future

Figures 8.14 – 8.17 show diachronic shifts in the use of individual loanwords and the LW stratum in the corpus within the same birth-year group. It appears that there is

no systematic upward shift over 20 years except for *sapooto-suru* ($< support$). *Sapooto-suru* barely shows a jump between 1986/1991 and 1996/2001 among younger generations (Figure 8.15).

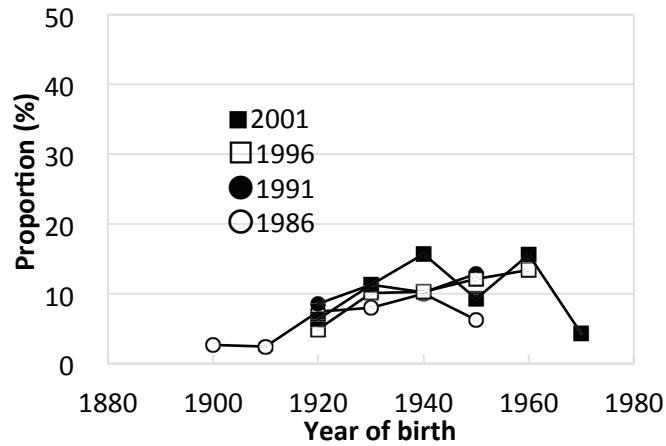


Figure 8.14 Variability of the individuals: *keesu* (< case)

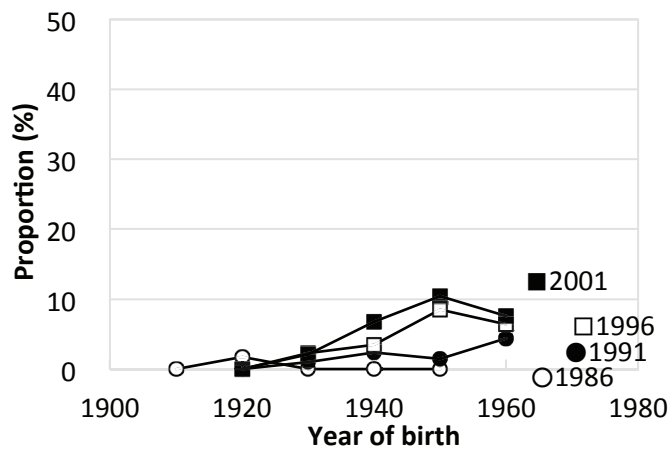


Figure 8.15 Variability of the individuals: *sapooto-suru* (< support)

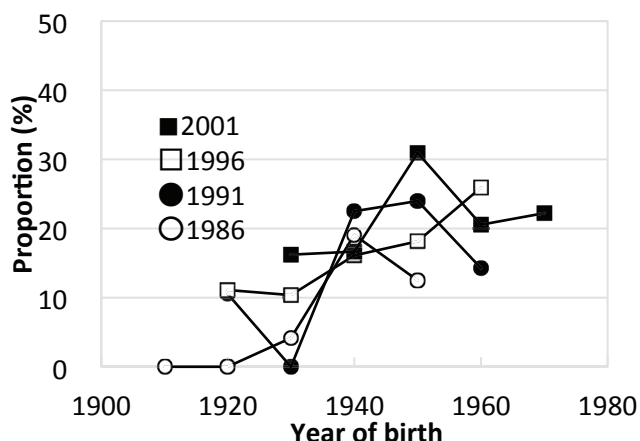


Figure 8.16 Variability of the individuals: *shinpuru* (< simple)

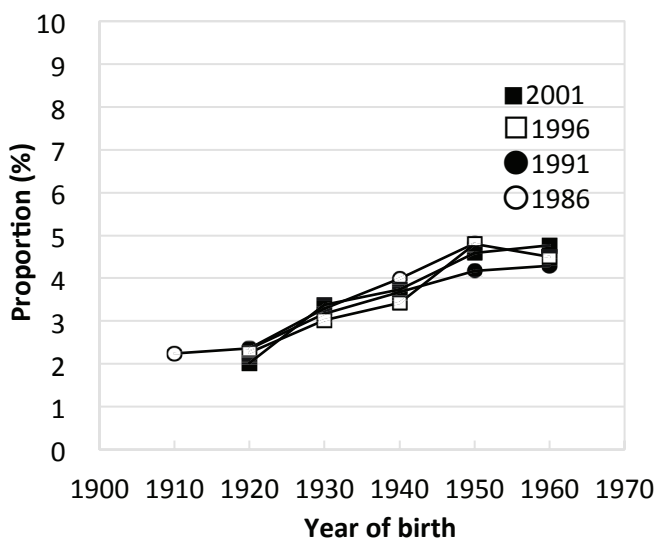


Figure 8.17 Variability of the individuals: The LW stratum

A logistic regression analysis was conducted to confirm the statistical significance of the effect of (i) Year of Survey (Year of Publication) and (ii) Year of Birth on the adoption of loanwords. Note that the disruptive factor (iii) Context is included into the analysis of *oopun-suru* this time too. The results (Table 8.7) confirm that the speakers born more recently show greater receptiveness to loanwords, which we have already seen as the age-effect in Table 8.6. The effect of Year of Survey, however, affects the

phenomenon in various ways. Positive and negative coefficients indicate that individuals are unstable over time, while 0 indicates that they are stable over time. Each case will be explained by a diagram in the following section.

Table 8.7 Regression coefficients of the independent variables: (i) Year of Survey and (ii) Year of Birth

	Year of Survey	Year of Birth	Context	constant	R^2
Interview-based					
<i>wain</i>	0.060602	0.058587	NA	-233.957192	0.96
<i>oopun-suru</i>	0	0.014613	1.110060	-29.289641	0.88
Present use	-0.023899	0.038376	NA	-26.581827	0.96
Further expansion	-0.139545	0.037763	NA	205.119082	0.97
Corpus-based					
<i>keesu</i>	0	0.012669	NA	-26.716283	0.23
<i>sapooto-suru</i>	0.121629	0.045745	NA	-334.947153	0.72
<i>shinpuru</i>	0	0.033220	NA	-66.158917	0.47
the LW stratum	0	0.017735	NA	-37.685165	0.85

8.4. Predicting change

8.4.1. Acceleration of change

The statistical analysis shows that Year of Survey has a positive effect on the occurrence of the loanwords *wain* ($R^2=0.96$) and *sapooto-suru* ($R^2=0.72$). The curves in Figure 8.18 shows predicted probabilities of people's preference for the loanword *wain* (over *budooshu*) in 1979, 1996, 2000 and, in addition, in the near future of 2020 and 2040 as a prediction of the coming 40 years. Figure 8.19 shows predicted probabilities of actual use of the loanword *sapooto-suru* in 1986, 1991, 1996 and 2001 as well as predictions for the following decade (2006 and 2011). The empty dots show the observed values.

There is an upward shift of the curves across year of survey, which indicates an acceleration of change toward more adoption of the loanword. This means that the individuals change over their lifespan in the direction of the change taking place in the community. Both examples show the cases where the apparent-time approach underestimates the actual rate of change.

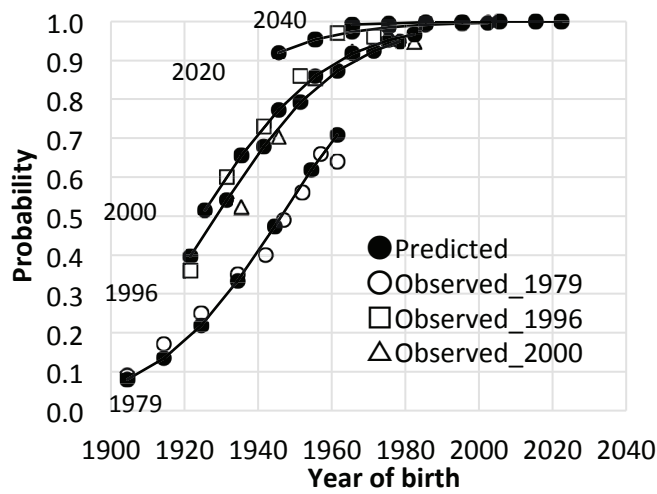


Figure 8.18 Acceleration of change: *wain* (< *wine*)

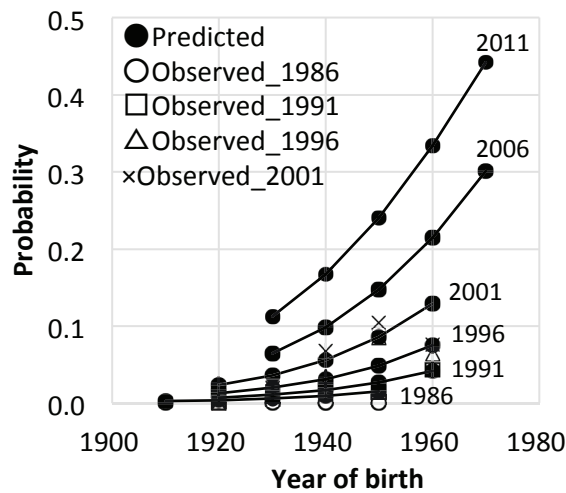


Figure 8.19 Acceleration of change: *sapooto-suru* (< *support*)

8.4.2. No acceleration/deceleration of change

The statistical analysis shows that Year of Survey is not significant for the occurrence of the loanword *oopun-suru* ($R^2=0.88$), *keesu* ($R^2=0.23$), *shinpuru* ($R^2=0.47$) and the LW stratum ($R^2=0.85$).

First, the curves in Figure 8.20 show predicted probabilities of speakers' preference for the loanword *oopun-suru* (over *kaiten-suru*) in 1979, 1996 and 2000 as well as the near future, 2020 and 2040. The empty dots show the observed values in 1979, 1996 and 2000. The upper curve predicts the case where the context is provided, whereas the lower curve predicts the case where the context is not provided.

As the coefficient of Context is positive, we can see that the presence of the context in question promotes preference for the loanword to a high degree (0.2 – 0.3). The five curves overlap each other and form one curve, which indicates that there is neither upward nor downward shift across year of survey.

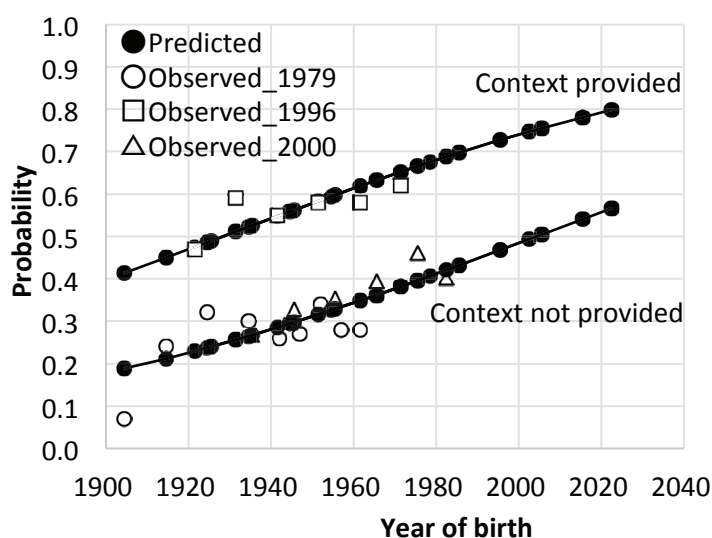


Figure 8.20 No acceleration/deceleration of change: *oopun-suru* (< *open*)

Second, the curves in Figures 8.21 – 8.23 show predicted probabilities of the occurrence of the loanword *keesu*, *shinpuru* and the LW stratum in the corpus across birth-year groups. The empty dots show the observed values in 1986, 1991, 1996 and 2001. It is evident that there are neither upward nor downward shifts across year of survey. These four linguistic variables show the cases where the individuals remain stable over time as the apparent-time hypothesis presupposes.

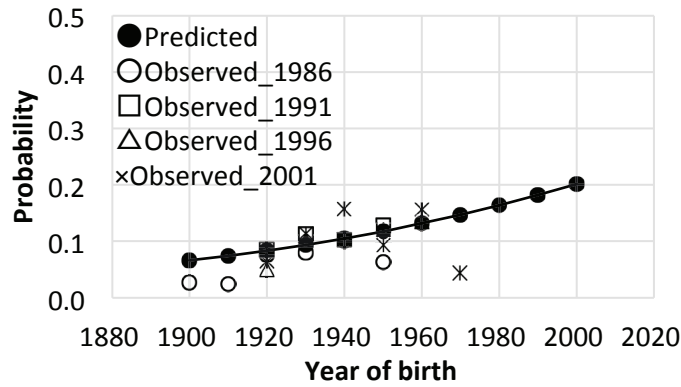


Figure 8.21 No acceleration/deceleration of change: *keesu* (< case)

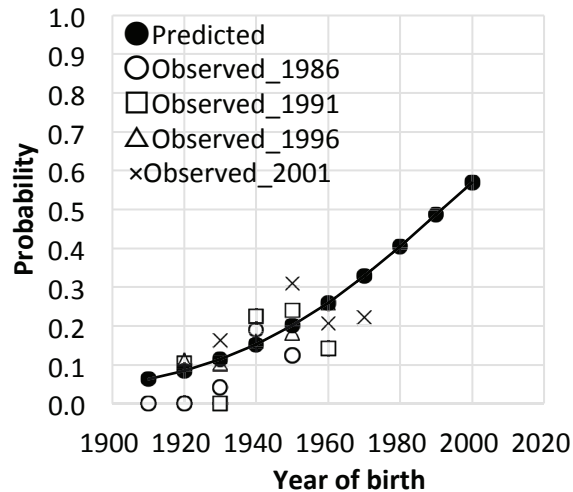


Figure 8.22 No acceleration/deceleration of change: *shinpuru* (< simple)

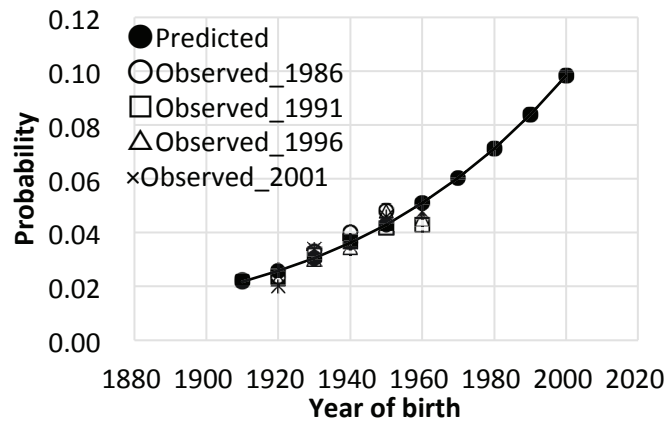


Figure 8.23 No acceleration/deceleration of change: The LW stratum

8.4.3. Deceleration of change

The statistical analysis shows that Year of Survey has a negative effect on people's receptiveness to the present use of loanwords ($R^2=0.96$) and people's receptiveness to a further expansion of loanwords ($R^2=0.97$). The curves in Figure 8.24 show predicted probabilities of people's receptiveness to the present loanword use. The prediction is made for the years 1977, 2000 and 2013 as well as the near future 2030 and 2045. The empty dots correspond to the observed values in 1977, 2000 and 2003 with observed/predicted values for the years 2002 and 2008 excluded to make the figure simpler.

The figure shows a downward shift of the curves across year of survey, which demonstrates a deceleration of change toward more receptiveness to loanwords. This means that the individual speakers change over their lifespan in the direction "opposite" to change taking place in the community. This example shows the case where the apparent-time approach overestimates the actual rate of change.

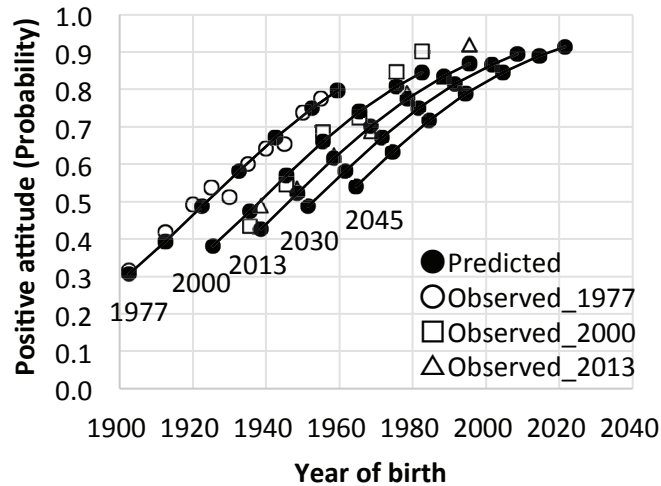


Figure 8.24 Deceleration of change: The present use of LWs in general

Figure 8.25 shows predicted probabilities of people's receptiveness to further loanword expansion. The prediction is made for the years 1995 and 2003 as well as the near future 2011 and 2019. The empty dots correspond to the observed values in 1995 and 2003.

The figure shows a downward shift of the curves across year of survey. Here, we can see that speakers change at a rate faster than that of generational change. This can be interpreted as an acceleration of "reversed" change. Therefore this example shows the case where the apparent-time approach underestimates the rate of "reversed" change.

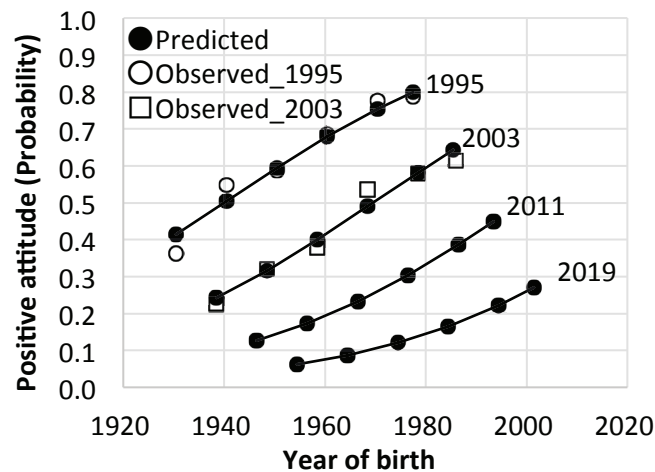


Figure 8.25 Deceleration of change: A further expansion of LWs in future

8.5. Discussion

8.5.1. Limitations on the apparent-time approach in making longitudinal predictions of change

Table 8.8 summarizes the result of the real-time approach, a real-time construct of change in preference for, or use of, loanwords. In this table we see two types of apparent-time approach's limitations in making longitudinal predictions of change: it fails to make accurate predictions of the (i) rate and (ii) direction of change.

First, the rate of change is sometimes not faithful to what the apparent-time hypothesis predicts. On the one hand, the hypothesis predicts correctly the rate of change in half of the cases under investigation, but on the other hand, it has been proven that the rate is underestimated or overestimated in the remaining half of the cases. The four cases have shown that the speakers change after critical period either in the direction of change in progress in the community or in the direction opposite to it. We have seen that the assumption that speakers do not change their overall linguistic

behavior throughout their lifespan is not always applicable to the phenomenon of lexical variation studied here.

Second, the apparent-time approach may misjudge the direction of change from time to time. Age differentiation, in which younger generations show higher receptiveness to loanwords, was attested to by all the variables. The apparent-time hypothesis then predicts that the community as a whole will become more open to loanwords over time. However, the discussed data shows that there is a case in which change is actually progressing in the direction opposite to what the hypothesis predicts (reversed change).

Table 8.8 Summary of the real-time approach

Variables: (1) Attitude (2) Behavior		Age differentiation	Variability: community	Variability: individuals
Individual items	<i>oopun-suru</i> <i>keesu, shinpuru,</i> the LW stratum	become more open (as they become younger)	becomes more open	no change
	<i>wain</i> <i>sapooto-suru</i>	become more open	becomes more open	become more open
Loanwords in general	Present use of LWs	become more open	becomes more open	become less open
	Further expansion of LWs	become more open	becomes less open	become less open

8.5.2. Mechanism of lifespan change

Table 8.9 is the revision of Table 8.1, in which *lifespan change* is subdivided into three groups: (i) “accelerated,” (ii) “decelerated” and (iii) “reversed” change. The eight linguistic variables examined here all involve the variability of the community as a whole and hence are interpreted as *generational change* as per Labov’s definition

(1994: 83). From Sankoff's point of view (2005), however, half of the cases, including *wain*, *sapooto-suru*, "Present use of LWs" and "Further expansion of LWs," can be interpreted as *lifespan change*, as there is the variability of the individuals after their critical period besides the variability of the community overall.

The original definition of lifespan change by Sankoff only mentions "change, in which individual speakers change over their lifespan in the direction of a change in progress in the rest of the community" (ibid.: 1011). This type of change, "accelerated change," is attested in the case of *wain* and *sapooto-suru*.

The present study, however, reveals that, in some cases, individual speakers change in the direction opposite to change taking place in the community overall. When individual speakers change at a rate slower than that of generational change, the community as a whole will still progress in the direction consistent with what the apparent-time hypothesis predicts (i.e. it becomes more open to the use of loanwords) although the rate of change will be decelerated over time. This type of change, "decelerated change," is attested to by the public attitude to the present use of loanwords.

On the contrary, if individual speakers change at a rate faster than that of generational change, the change is reversed and hence the community moves in the direction opposite to what the age gradient suggests. This type of change, "reversed change," is attested to by the public attitude to a further expansion of loanwords in future.

Table 8.9 Types of change (revised)

	Age differentiation	Variability: Community	Variability: Individual	Influence on change	Example	
					Attitude	Behavior
1 Stability	–	–	–			
2 Age-grading	+	–	+			
3 Lifespan change (Sankoff, G. 2005)	+ ↑	+ ↑ ↓	+ ↓ ↓	accelerated decelerated reversed	<i>wain</i> Present use Further spread	<i>sapooto</i>
4 Generational change	+	+	–		<i>oopun-suru</i>	<i>keesu</i> <i>shinpuru</i> LW stratum
5 Communal change	–	+	+			

↑: in the direction of change in progress in the community, ↓: in the direction opposite to it

A valid explanation for the mechanism of “accelerated change” would be the “Circular Model with Mere Exposure Effect” (Yokoyama 2006: 200, see also Yokoyama et al. 2008: 73). This model is based on the “Mere Exposure Effect” in psycholinguistics, in which a repeated exposure to a given stimulus is responsible for an increased preference for the stimulus. Yokoyama expands this model to account for the mechanism of life-span change in language attitudes. An increased “exposure frequency” of a given form affects one’s mental lexicon, including familiarity, preference and utility of the form, and in turn promotes increased reproduction of it. This explains the case of acceleration of diffusion of loanwords observed in the present study. That is, as the level of one’s exposure to a given loanword, e.g., *wain* or *sapooto-suru*, becomes higher, one’s preference or familiarity of the form becomes greater and in turn promotes further use of it.

In contrast, Labov (1994: 91) points out that decelerated change is attributed to age-grading. Provided that the observed age gradient is a reflection of a pure age-grading, the case of the deceleration of generational change toward more adoption

of loanwords is explainable by the fact that people potentially becomes less open to loanword use as they grow older. Moreover, an increased use of loanwords could increase people's negative impressions of loanwords as a whole (such as "a flood of loanwords") and consequently lead to regression to native forms over time, following the norm in which the native variants are considered to be more traditional or indigenous to Japanese than loan variants. This kind of change (i.e. reversed change) occurs usually, as Sankoff says, "as the effect of stigmatization from above" (2006: 113).

It is worth mentioning that interview-based data show that people become more likely to reject a further spread or the present use of loanwords overall as they age. The corpus-based data, however, confirmed that there is no deceleration in their actual adoption of the LW stratum in real time. This shows that the lifespan change on the level of people's attitude and that of their behavior are different things: people's attitude to a certain linguistic variable may be more susceptible to change than their actual use of it. It is, however, worth noting that there is also an example of lifespan change of people's behavior in the case *sapooto-suru* (Table 8.9).

8.6. Conclusion

This chapter has demonstrated that the apparent-time approach has limitations in predicting the rate and direction of change. The present data revealed the presence of lifespan change and, at the same time, provided descriptions of its subtypes. It is, in particular, notable that a change could be reversed if a speaker changes (i) in the direction opposite to the change taking place in the community and (ii) at a rate faster

than that of the generational change in progress. This highlights the importance of investigation into the variability of individual speakers over their lifetimes in making a longitudinal prediction of change.

To conclude, the present trend study has made the best use of the existing apparent-time data in order to elucidate a real-time construct of change in the diffusion of loanwords, leading to more accurate predictions of the phenomenon.

9. Conclusions

9.1. General summary and discussion

The current research attempted to describe different aspects of lexical variations between the loanword forms (that are attested to some extent) and their existing native alternatives. The empirical findings gained in the previous chapters answer the four questions set out in Chapter 1.

9.1.1. Extending variationist methodology to lexical variation

The first question concerns how one can identify environments where variation occurs on a lexical level (QUESTION I). We have witnessed that the study of lexical variation, especially a corpus-based one as opposed to an interview-based one, yields several theoretical and methodological difficulties. This comes directly from the difficulty in applying the notion of semantic equivalence to the study of variation beyond phonological level (Lavandera 1978, Labov 1978, Weiner & Labov 1983). Some researchers suggest that the criteria for defining linguistic variables in the context of phonological variation should be re-examined (Lavandera 1978, Sankoff, D. 1982). That is, functional/discourse equivalence should be employed in identifying the variable context of non-phonological variation.

Following this discussion, in Chapter 6, I carefully located environments where lexical variants are used interchangeably on a functional level, based on an in-depth observation of the function, grammar and collocational pattern of *keesu* (< *case*) and its native variants (“contextual approach”). This approach, although it gives a fine-grained analysis, poses a problem in drawing a general conclusion. To overcome this limitation,

a more comprehensive approach was taken for the investigation of groups of loanwords in Chapter 5 (“comprehensive approach”). Based on synonymy determined by the dictionary-based examination, this approach considers all environments in which the identified variants occur. Despite the fact that the analysis is less fine-grained, it succeeded to some extent in eliciting general sociolinguistic patterns in the distribution of the selected loan variants.

In summary, both types of corpus-based approach were able to capture the structured nature of lexical variation under discussion. More specific findings are summarized in the following sections.

9.1.2. Structured nature of variation

The second question asks in what manner the lexical variation in consideration is structured (QUESTION II). Four external factors including (i) age/year of birth, (ii) register, (iii) style and (iv) education, have proved to be significant in the preference for, or occurrence of, loanword variants.

First, both interview-based studies and corpus-based ones were generally able to capture age gradients in the adoption or use of loanwords, with this level getting higher as one becomes younger (Chapters 4, 5, 6, 7 and 8). The repeatedly-observed generational difference indicates that there are changes in progress in favor of loanwords as opposed to their native equivalents. There was an exception in which those in their teens show a delay in acquisition of some loanwords, but it has been shown that they catch up with the other generations in the course of time (Chapter 4).

Second, corpus-based investigations (Chapters 5 and 6) revealed that register and several stylistic factors have an impact on the occurrence of loanwords. I have pointed

out that the register difference in the written corpus (BCCWJ) such as books (PB) vs. magazines (PM) is associated with the difference in the formality of a text. The register difference between public speech (SPS) and academic presentation speech (APS) in the spoken corpus (CSJ) is also found to arise from factors such as spontaneity, formality and audience size, which characterize the sociolinguistic style of the speech settings in the registers in question. In both written and spoken corpora, there is a general tendency that loan variants are disfavored in settings with more formal or less spontaneous speech style.

Third, education was also found to be a significant variable in the occurrence of loanwords, but this variable showed a complex picture of its relation to the phenomenon. In the written corpus loanwords are disfavored by the most educated group (GRAD) (Chapter 5). The investigation of *keesu* (< *case*) also revealed a systematic decrease in the use of the loanword as one becomes more educated (HS>UNIV>GRAD) (Chapter 6). In contrast, the interview-based survey in Chapter 7 revealed that the occurrence of the loanword *sapooto* (< *support*) in conversation among friends increases as one becomes more educated (JH<HS<UNIV⁷⁵). Such an apparent discrepancy between the two studies would be explained by the interaction between educational and stylistic variables, i.e., the fact that only the most educated group UNIV shows style shifting between the two different settings: conversation with friends vs. public speech (Chapter 7). This suggests the possibility that those with higher educational backgrounds (e.g. university graduates and above) use loanwords

⁷⁵ UNIV and GRAD are not distinguished in the original data, but it is likely that those with post-graduate degrees are incorporated into the category UNIV here.

more selectively depending on settings,⁷⁶ i.e., the effect of an educational variable is modified by a stylistic variable. In a very casual setting such as the context of talking to one's friends, more educated people use loanwords more frequently than less educated people.⁷⁷ Meanwhile, a more formal setting such as public speaking or writing for publication would inhibit their use of loanwords to a degree lower than that of less educated groups.

Besides these external variables, internal ones have also proved to be significant. The in-depth study of *keesu* (< *case*) (Chapter 6) revealed that the loanword in a subject position occurs more frequently as a gap-less relative clause head (gap-less RCH). In addition, the loan variant is more strongly associated with predicates that emphasize the occurrence or quantitative evaluation of the event as opposed to its native variants that are more strongly associated with predicates reporting the mere presence/absence of the event.

9.1.3. Language attitude and language behavior

The third question inquires about whether or not corpus-based studies provide additional support for the result of previous surveys on attitudes towards loanwords (QUESTION III). We have seen that sociolinguistic asymmetry in adoption of loanwords in terms of generation and style is attested to by both public *attitudes* (Chapters 4 and 7) and their actual *behavior* (Chapters 5 and 6). An educational variable, at a glance, shows a discrepancy between attitude and behavior, but it is apparently

⁷⁶ Recall that style-shifting is more prominent among MC than WC groups when speech style switches between Formal and Casual (Trudgill 1974: 93, see also Figure 7.2).

⁷⁷ We should note that more educated people show higher rates of recognition/acceptance of the loanwords under investigation (Ishiwata 1965, Loveday 1996). Therefore it seems that people's self-reports tend to lead to a conclusion that those with higher educational backgrounds adopt loanwords more if no stylistic context is attached to the interview question.

attributable to the fact that the variable is in interaction with stylistic variable as just mentioned above.

Chapter 8 showed that there is a discrepancy between the public's attitudes and their behavior: In spite of the deceleration of changes toward more adoption of loanwords in general, the actual use of the LW stratum in the written corpus (BCCWJ) does not show such a deceleration. This suggests that the longitudinal process of the adoption of loanwords should be evaluated and predicted with meticulous care by means of examinations of people's attitudes towards, as well as their actual use of, loanwords.

9.1.4. Demonstration of change in real time

The fourth question aims to provide a picture of the process of the phenomenon in real time (QUESTION IV). In order to discuss the problems of the apparent-time approach, I have made use of existing interview-based data as well as the diachronic component of BCCWJ (the Library sub-corpus or LB) (Chapter 8). There were two questions to be tackled: (i) whether or not the observed age gradients are interpretable as changes in progress; (ii) whether or not the individuals remain stable throughout their lifetime with respect to their linguistic attitudes/behavior.

With regard to the first question, all the cases under consideration verified the variability of the community as a whole over time, which confirms that the change is in progress in real time. With respect to the second question, the apparent-time hypothesis, i.e., the presupposition that individuals are invariable during their lifespan (Labov 1994), was supported by the half of the data under discussion, but, at the same time, it was also challenged by the rest of the data. It has been shown that people changes their attitude

toward, or behavior on, the phenomenon, either in the direction consistent with that of the community as a whole moves or in the direction opposite to it.

In the former case, in which people become more open to loanwords as they age, a change is accelerated, and in turn, a change toward in favor of loanwords proceeds at a rate higher than the apparent-time approach predicts (e.g. the use, or attitude toward the use, of *wain* (< *wine*) and *sapooto* (< *support*)). On the other hand, in the latter case, in which people become less open to loanwords as they age, a change is (i) decelerated if a change proceeds at a rate lower than the apparent-time approach predicts, or (ii) reversed if the rate of deceleration is greater than a given age differentiation. In the present study, a decelerated change was attested to by people's attitude toward the present use of loanwords in general, and a reversed change was attested to by people's attitude toward a further expansion of loanwords in future. The corpus-based data, however, showed no indication of deceleration in the diffusion of the LW stratum as opposed to what the public attitude exhibits.

These discussions put emphasis on the importance of a real-time approach to variation, especially on a lexical level, which is said to be less stable compared with those on other linguistic levels such as phonology or syntax.

9.1.5. "Actuation problem" of Western loanwords

As Weinreich et al. (1968: 186) points out, "the overall process of linguistic change may involve stimuli and constraints both from society and from the structure of language." Now we turn to the actuation problem of Western loanwords, i.e., a general question of why a certain change occurred at a certain point of time in a certain language. Ultimately, it is difficult to find a single reason that applies to all the

loanwords borrowed into the language, since, at the very micro level, every loanword seems to have been borrowed for different, case-by-case, multiplex reasons. As Labov (2001a: 466) says, the beginnings of change are as mysterious as ever. We can, however, at least say something from what we find about the continuing change (ibid).

The loanwords examined in the current research point out several possible reasons why Western loanwords are borrowed and spread when the recipient language already has corresponding or synonymous native expressions.

First, we have seen in Chapter 4 that there is a case in which a borrowed item conveys a meaning slightly different from what is meant by its native equivalent (e.g. *getto-suru* (< *get*)). The reason why loanwords of this type spread is that they can deliver *specific* concepts/delicate shades of meaning that cannot be conveyed by the existing words.

Second, we have also found that a loanword often covers a broader area of meaning than its native equivalent (e.g. *suutu* (< *suit*) and *sebiro*; *oopunn-suru* (< *to open*) and *kaiten-suru*). Similar arguments are also found in Kim (2011: Ch4) and Miyata (2007). It is inferred that the use of a loan variant in this case is activated because it plays a broader function than its native variants and serves as a *general* term for them.

Third, direct borrowing from Western languages can reduce the problem of homonyms caused by the import of a large number of Sino-Japanese words in the language (Arakawa 1943: 196-202). On the one hand, SJ words are specific/concise in meaning and short (space-saving) in terms of word length, and therefore are highly favorable for formal writing. SJ words, on the other hand, usually form massive groups of homonyms, which often lead to miscommunication in speaking. For example,

according to *Kojien* 6th, the SJ word *shien* ('support') has 10 SJ homonyms; The SJ words *kikai* and *kooki* (both refer to 'chance') have 10 and 31 SJ homonyms, respectively. Their LW equivalents *sapooto* (< *support*) and *chansu* (< *chance*), in contrast, do not have homonyms; By using Western loanwords, we can avoid the difficulty of *disambiguating* SJ homonyms.

Fourth, we have seen in Chapter 7 that NJ, SJ and LW are differentiated in terms of stylistic functions. The LW *sapooto* (< *support*) is usually considered to express novelty or stylishness as opposed to NJ and SJ words (*tedasuke*, *shien*). In contrast, NJ is regarded as more colloquial, and, SJ as more intelligent or dignified, than LW. It is then not difficult to imagine that the LW *sapooto* is gaining in popularity because it plays certain *stylistic* functions that are differentiated from those of its NJ or SJ equivalents.

Finally, linguistic orientation toward English after WWII has increased *social* pressure for using loanwords. NINJAL (2005a: 139-141) shows that 73.1% of people think that English is a predominant or influential language. And the majority of those accept to some extent the idea that English is recognized as such, with 32.1% showing simply positive attitudes and 60.2% feeling that they have no choice but to accept the situation. As Chapter 8 shows, the generational change in the attitude toward the present use of loanwords implies that the whole society moves toward more frequent use of loanwords.

To sum up, the continuing spread of Western loanwords is "actuated" by multiple factors. They can be (i) linguistic (e.g. specification, generalization or disambiguation of meaning), (ii) stylistic and/or (iii) social. Word choice depends on the combination of,

or competition among, these multiple factors. This point would explain to a certain extent why the Japanese language has multiple lexical strata.

9.1.6. Sociolinguistic status of loanwords

The import of Sino-Japanese words is usually characterized as “change from above” as they are originally brought and used as learned vocabulary by a few intellectuals. And up to today SJ words have shown association with writing and formal registers in the Japanese language.

Likewise, the naturalization of early Western loanwords can be described as linguistic change from above the level of consciousness; it started in the wake of the influx of Western advanced cultures and techniques including religion, medicine and science, many of which were at the beginning imported and used by a few elites in the upper class symbolizing Westernization or modernization of the society.

Longitudinal observations made in Chapter 8 suggest that change in favor of loanwords is in progress in real time. The increasing influx of English donor loanwords suggests that the English language has been acquiring more prestige than ever before, and exerted a constant influence on the Japanese language especially since WWII. However, findings in Chapters 5 – 7 repeatedly show that loanwords are disfavored in socially formal contexts. An in-depth analysis of the mechanism of style shifting in the choice between the loan variant (*sapooto*) and its non-loan variants (*shien* and *tedasuke*) (Chapter 7) revealed that people switch to the non-loan variants (especially the SJ variant *shien*) in order to add a sense of intelligence and dignity to their speech when their speech style shift to a more formal setting such as public speaking. The most

educated people in particular regard the SJ equivalent *shien* as more appropriate to use in front of the general public than the loanword *sapooto*. This strongly suggests the non-standard nature of loanwords as opposed to their native equivalents in the Japanese language.⁷⁸ In other words, we can see that Western loanwords, especially recent borrowings that entered into the language despite the presence of corresponding native ones, have not acquired the status as standard/prestigious as SJ words have.⁷⁹

In conclusion, the synchronically asymmetric distribution of the loan variants examined here suggests that so-called well-attested loanwords examined in this thesis, e.g., *keesu*, *sapooto*, etc., still remain peripheral bound by various social contexts such as register, style, education and the interaction between these variables.

9.2. Prospects for further research

In spite of its overall success in offering a general sociolinguistic profile of the alternation between loan variants and non-loan variants, the present investigation produces several suggestions for further research.

9.2.1. Exploration of a wider range of lexical variables

First, the contextual approach should be extended to broader range of lexical variables. I have conducted a case study of one variable (*keesu* vs. its native variants) that could arguably be considered a representative of loanwords that are attested in Japanese to a certain degree. However, the findings of this case study will need further support by

⁷⁸ Yet the alternation between these variants is not merely the question of prestige/non-prestige. Speakers switch to the non-loan variants in order to make their addresses easier to understand, i.e., the switch also occurs for a communicative purpose.

⁷⁹ But we should notice that this may not apply to all the loanwords. For example, the use of technical, academic, specialized terms that are borrowed to introduce new concepts, e.g., *aidentiti* (< *identity*), *akauntabiriti* (< *accountability*), *konpuraiansu* (< *compliance*) and so on, convey a relatively strong sense of sophistication or forwardness.

empirical evidence from studies of other loanwords. In order to achieve this goal, a sufficient foundation will have to be established regarding the functions, usages and collocational patterns of a wider range of loanwords and their near-synonyms. As Mogi (2011) and Kim (2011) pointed out, the present situation is far from that phase. The further accumulation of in-depth studies of individual loanwords like Kim (2011) would promote an application of the contextual approach to lexical variation, and, in turn, the pursuit of a more comprehensive picture of the phenomenon.

9.2.2. Search for additional language-internal factors

Second, search for language-internal factors that can be applied to a wider range of words is required. Chapter 6 successfully investigated the use of the LW *keesu* (< *case*) in terms of quantitatively measurable language-internal factors such as “Usage” and “Co-occurring Predicates” since *keesu* and the other three variants examined are NPs used in various linguistic environments. But the above factors give only an ad hoc explanation applicable only to a limited case and may not be useful for every loanword and its variants. We should therefore consider more universal language-internal factors in order to explain a broader range of lexical items, including concrete nouns, abstract nouns, verbs and so on.

Loveday (1996: 81) points out “the existence of many pairs in semantic opposition where a word referring to a Western phenomenon is English-based and ‘complementary’ with a word deriving from (Sino-) Japanese and referring to a related version of the phenomenon belonging to native culture.” According to Loveday, the loanword *kusshon* (< *cushion*), for example, is semantically marked by the feature [- Japanese], as opposed to its native equivalent *zabuton* (= ‘thin cushion’) (being marked by the feature

[+Japanese]). His argument holds true for many pairs of concrete nouns such as *supuun* (< *spoon*) and *saji, hoteru* (< *hotel*) and *ryokan, kitchin* (< *kitchen*) and *daidokoro*, and so on.

Here, we extend Loveday's argument to the context in which a loanword is used in order to explain its occurrence. That is, we can hypothesize that loanwords are used more frequently as the preceding/following context becomes more westernized; in other words, the occurrence of a loanword is affected by the presence of other loanwords occurring in the surrounding context.⁸⁰ We call this the "surrounding contextual effect."

To test this hypothesis, I have looked at the surrounding context of the loanwords listed in Table 4.2.⁸¹ Table 9.1 shows the number of loanwords contained in the surrounding context of the variants under investigation, i.e. variant 1 (v1), variant 2 (v2) or variant 3 (v3). Two types of the surrounding context were examined: (i) the 10 words preceding and following each variant; (ii) the 100 words preceding and following each variant. The table suggests that all the loanwords listed except for *meritto* (< *merit*) and *akauntabiriti* (< *accountability*) have more loanwords in their surrounding context than their native equivalents do, supporting the above hypothesis. This is one of the possible language-internal factors that can be applied to a wider range of loanwords.

⁸⁰ We have seen in Chapter 8 that the context that contains the loanword *suupaa* (< *super* (*market*)) enhances the preference for the loanword *oopun-suru* (< *to open*) (see Table 8.6).

⁸¹ Tokens are counted if the variant in question occurs in the following context in the Publication sub-corpus, BCCWJ (PB and PM, the variable-length components only).

Nouns: (a word ends in hiragana writing + key word + particle/auxiliary)

Verbs: (a word ends in hiragana writing + key word + *suru* ('do')/*dekiru* ('be able to')

tootaru-de/gookei-de: (a word ends in hiragana writing + key word + *de*)

Table 9.1 Number of loanwords in the surrounding context of variants

ID	Variants			No. of collected tokens				No. of LW in the surrounding context (10 words before and after each variant)			No. of LW in the surrounding context (100 words before and after each variant)		
	variant1 (LW)	variant2 (nonLW)	variant3 (nonLW)	v1	v2	v3	Total	v1	v2	v3	v1	v2	v3
1	<i>niizu</i>	<i>hitsuyoosei</i>		579	756		1335	0.60	0.31		7.60	3.69	
2	<i>ibento</i>	<i>moyooshi</i>		416	58		474	1.35	0.62		13.78	5.21	
3	<i>meritto</i>	<i>riten</i>		546	368		914	0.69	0.79		9.00	8.70	
4	<i>konsensasuu</i>	<i>gooi</i>		65	366		431	0.28	0.19		4.12	2.43	
5	<i>sukiimu</i>	<i>keikaku</i>		36	1260		1296	0.33	0.32		4.47	3.77	
6	<i>risuku</i>	<i>kikensei</i>		888	487		1375	0.65	0.39		8.48	5.02	
7	<i>bijon</i>	<i>tenboo</i>		171	160		331	0.47	0.30		6.23	4.14	
8	<i>akauntabiri</i>	<i>setsumei</i>		7	31		38	0.14	0.48		5.14	5.48	
	<i>ti</i>	<i>-sekinin</i>											
9	<i>wain</i>	<i>budooshu</i>		445	63		508	1.10	0.56		11.15	3.19	
10	<i>suutsu</i>	<i>sebiro</i>		251	47		298	1.35	0.55		9.38	3.47	
11	<i>kitchin</i>	<i>daidokoro</i>		297	268		565	1.05	0.44		11.04	4.53	
12	<i>oopun-suru</i>	<i>kaiten-suru</i>		196	51		247	1.48	1.20		13.10	8.96	
13	<i>getto-suru</i>	<i>kakutoku-suru</i>		81	942		1023	1.81	0.65		17.04	6.23	
14	<i>tootaru-de</i>	<i>gookei-de</i>		24	26		50	1.88	0.46		14.29	7.50	
15	<i>kyanseru</i>	<i>kaiyaku</i>	<i>torikeshi</i>	28	40	141	209	0.75	0.15	0.13	8.11	2.43	1.65
16	<i>happii</i>	<i>shiawase</i>	<i>koofuku</i>	46	694	194	934	1.11	0.21	0.18	11.37	3.31	2.12
17	<i>sapooto</i>	<i>shien</i>	<i>tedasuke</i>	120	668	105	893	0.96	0.34	0.32	10.49	4.01	4.65

9.2.3. Exploration of other sources of data

Third, language in a more casual setting should be examined. This thesis consulted written text (BCCWJ) and recordings of public speaking (CSJ), both of which are characterized as language that is relatively formal in style (and produced by those with relatively high educational backgrounds). In light of the non-standard nature of loanwords in Japanese, it can be expected that one will witness heavier use of loanwords when observing language that is more casual in style, such as chatting with friends/family.

Meanwhile, although it has not yet been released, NINJAL plans to launch the Corpus of Everyday Japanese Conversation (CEJC) (Koiso et al. 2016a; 2016b), which

will be a collection of 200 hours of recordings of daily conversation between 100 – 250 ordinary speakers. This corpus is designed to collect speech produced by people in a wider range of registers that are associated with everyday life, e.g., conversations observed in one's home, community, workplace, school and so on. The informants record themselves, so that the effect of an observer (the problem of the observer's paradox) can be minimized.

In addition, NINJAL has been launching the NINJAL Web Japanese Corpus (NWJC) (Asahara et al. 2014), which is a collection of ten billion words from the Internet. One of the advantages of using such an immense sum of data is that it enables us to investigate linguistic phenomena that are relatively rare in frequency (such as lexical variation) more easily.

Moreover, it is a good opportunity for us to incorporate a new type of register to the current analysis. The nature of writing on the Web is considered to be something between formal writing and natural speech. It is expected that we will obtain more tokens of loanwords in such a register than in formal text like books. As already mentioned in 5.4.6, Tanaka (2012: 229) says that text from online is more strongly associated with the “private” and “daily” domains of life and actually shows that the proportion of loanwords in online text and magazines is relatively higher than that in books and newspapers.

However, attention needs to be paid to the question of whether we can characterize a register like “online conversation” within the framework of the traditional definition of style: can it simply said to be something between formal and casual? The most characteristic feature of online text is that addressees are not known/unpredictable to the

author. Such “unpredictability” can lead the author either to pay more or, on the contrary, less attention to speech. The investigation of a register with “unspecified addressee” would help us to understand a new and unavoidable type of communication in the current age.

Although the two corpora investigated in this thesis are among the best corpora in terms of the volume (reasonable for the quantitative analysis) and quality (rich annotation for sociolinguistic study), it is possible that an additional investigation into speech closer to vernacular or speech on the Web would add a new insight to the present research into the sociolinguistic picture of the phenomenon.

9.2.4. Application of a multinomial logistic regression analysis

Fourth, there is room for improvement in the analytical method. Since the investigations of *keesu* (and its native variants *jirei*, *baai* and *rei*) in Chapter 6 and *sapooto* (and its native variants *shien* and *tedasuke*) in Chapter 7 deal with four- or three-way distinction, a multinomial logistic regression (see Maeda 2005) could be used instead of binary one in order to understand the competition among the non-LW variants.

The question, however, arises as to which variant should be a reference category. One of the possibilities is that we use the most unmarked variant as a reference category. We can choose the stylistically most neutral lexical variant as a reference – that is, the variant that occurs most frequently in natural speech settings. We can expect that a NJ word would be that variant in most cases, e.g., *baai* (‘case’) in the case of Chapter 6 and *tedasuke* (‘support’) in the case of Chapter 7. In this way we can examine the possibility of the occurrence of “stylistically (more) marked” variants including LW and SJ as opposed to a “stylistically (more) unmarked” variant. That will make our understanding

of the stylistic difference between LW and SJ clearer. Another possibility is that we choose LW as a reference category so that we can see the probability of the occurrence of a certain loanword in contrast to every other variant.

These points need to be considered well before the analyst conducts a multinomial logistic regression analysis. In addition, we should make sure that the sample size of each lexical item does not become too small when we create a multi-way distinction for a dependent variable in order to make the results more reliable.

9.2.5. Towards cross-linguistic comparison

Lastly, the methodology employed in this thesis could be extended to other languages to verify the universality of the findings gained here. For example, a language like Korean also exhibits lexical strata that are very similar to Japanese. People recognize the differentiation among *ko-yu-e* ('native Korean words'), *han-ca-e* (Sino-Korean words, lit. 'Chinese character words') and *oy-lay-e* (loanwords, lit. 'words from abroad') (Sohn 1999: 87-88). Some *oy-lay-e* are said to function as innovative alternatives to existing native Korean words or Sino-Korean words in a given context (Yang 2007); however, Korean speakers exhibit a stronger resistance to adoption of such loanwords than Japanese speakers do, probably due to the national language policy to promote linguistic purism (醇化運動) by the Korean government since the latter half of the 20th century (Yang 2005, 2012: 150).⁸² Furthermore, since the Korean language does not distinguish each lexical stratum in their writing system as the Japanese language does, people may be less aware of the distinction among loanwords, native Korean words and

⁸² It should be noted that the primary interest of this movement has been removing the influence of Japanese loanwords. The emphasis on native Korean words does, however, seem to have an effect on the adoption of Western loanwords to a certain extent.

Sino-Korean words. Such social/linguistic differences between the two languages may then exhibit different sociolinguistic patterns of the phenomenon. A contrastive study of the two would encourage us to pursue study of the universality and diversity of lexical variation and change with respect to the diffusion of Western loanwords.

Appendix

A: Data set for the apparent-time study, interview-based (Ch. 4)

ID	LW	Year of Survey	Sex	Age	Year of Birth median	Acquisition	LW(%)	No. of informant		
1	<i>niizu</i>	1999	M	0	16-19	1981.5	2	29.0	62	
1		1999	M	0	20-29	1974.5	1	47.2	144	
1		1999	M	0	30-39	1964.5	0	51.3	154	
1		1999	M	0	40-49	1954.5	0	45.7	173	
1		1999	M	0	50-59	1944.5	0	42.5	193	
1		1999	M	0	60-69	1934.5	0	22.2	189	
1		1999	M	0	70+	1924.5	0	17.8	107	
1		1999	F	1	16-19	1981.5	2	32.3	62	
1		1999	F	1	20-29	1974.5	1	46.8	158	
1		1999	F	1	30-39	1964.5	0	48.0	223	
1		1999	F	1	40-49	1954.5	0	44.3	221	
1		1999	F	1	50-59	1944.5	0	31.9	210	
1		1999	F	1	60-69	1934.5	0	21.3	183	
1		1999	F	1	70+	1924.5	0	6.6	121	
2		<i>ibento</i>	1999	M	0	16-19	1981.5	1	71.0	62
2			1999	M	0	20-29	1974.5	0	79.2	144
2	1999		M	0	30-39	1964.5	0	78.6	154	
2	1999		M	0	40-49	1954.5	0	65.9	173	
2	1999		M	0	50-59	1944.5	0	58.0	193	
2	1999		M	0	60-69	1934.5	0	35.4	189	
2	1999		M	0	70+	1924.5	0	22.4	107	
2	1999		F	1	16-19	1981.5	0	90.3	62	
2	1999		F	1	20-29	1974.5	0	81.0	158	
2	1999		F	1	30-39	1964.5	0	72.2	223	
2	1999		F	1	40-49	1954.5	0	67.4	221	
2	1999		F	1	50-59	1944.5	0	54.8	210	
2	1999		F	1	60-69	1934.5	0	31.7	183	
2	1999		F	1	70+	1924.5	0	14.0	121	
3	<i>meritto</i>		1999	M	0	16-19	1981.5	2	62.9	62
3			1999	M	0	20-29	1974.5	1	68.1	144
3		1999	M	0	30-39	1964.5	0	72.1	154	
3		1999	M	0	40-49	1954.5	0	68.2	173	
3		1999	M	0	50-59	1944.5	0	64.8	193	
3		1999	M	0	60-69	1934.5	0	42.3	189	
3		1999	M	0	70+	1924.5	0	22.4	107	
3		1999	F	1	16-19	1981.5	1	69.4	62	
3		1999	F	1	20-29	1974.5	0	75.9	158	
3		1999	F	1	30-39	1964.5	0	69.5	223	
3		1999	F	1	40-49	1954.5	0	67.0	221	
3		1999	F	1	50-59	1944.5	0	57.1	210	
3		1999	F	1	60-69	1934.5	0	37.2	183	
3		1999	F	1	70+	1924.5	0	19.8	121	
4		<i>konsensasu</i>	1999	M	0	16-19	1981.5	0	3.2	62
4			1999	M	0	20-29	1974.5	0	6.3	144
4	1999		M	0	30-39	1964.5	0	11.7	154	
4	1999		M	0	40-49	1954.5	0	17.3	173	
4	1999		M	0	50-59	1944.5	0	17.6	193	
4	1999		M	0	60-69	1934.5	0	13.8	189	
4	1999		M	0	70+	1924.5	0	12.1	107	
4	1999		F	1	16-19	1981.5	0	3.2	62	
4	1999		F	1	20-29	1974.5	0	8.2	158	

ID	LW	Year of Survey	Sex	Age	Year of Birth median	Acquisition	LW(%)	No. of informant
4		1999	F 1	30-39	1964.5	0	4.9	223
4		1999	F 1	40-49	1954.5	0	9.5	221
4		1999	F 1	50-59	1944.5	0	4.3	210
4		1999	F 1	60-69	1934.5	0	3.8	183
4		1999	F 1	70+	1924.5	0	2.5	121
5		1999	M 0	16-19	1981.5	0	0.0	62
5		1999	M 0	20-29	1974.5	0	3.5	144
5		1999	M 0	30-39	1964.5	0	4.5	154
5		1999	M 0	40-49	1954.5	0	4.0	173
5		1999	M 0	50-59	1944.5	0	6.2	193
5		1999	M 0	60-69	1934.5	0	5.3	189
5		1999	M 0	70+	1924.5	0	1.9	107
5	<i>sukiimu</i>	1999	F 1	16-19	1981.5	0	0.0	62
5		1999	F 1	20-29	1974.5	0	4.4	158
5		1999	F 1	30-39	1964.5	0	2.2	223
5		1999	F 1	40-49	1954.5	0	2.3	221
5		1999	F 1	50-59	1944.5	0	1.9	210
5		1999	F 1	60-69	1934.5	0	4.4	183
5		1999	F 1	70+	1924.5	0	3.3	121
6		1999	M 0	16-19	1981.5	1	43.5	62
6		1999	M 0	20-29	1974.5	0	52.8	144
6		1999	M 0	30-39	1964.5	0	49.4	154
6		1999	M 0	40-49	1954.5	0	49.7	173
6		1999	M 0	50-59	1944.5	0	37.8	193
6		1999	M 0	60-69	1934.5	0	21.2	189
6		1999	M 0	70+	1924.5	0	12.1	107
6	<i>risuku</i>	1999	F 1	16-19	1981.5	1	45.2	62
6		1999	F 1	20-29	1974.5	0	51.9	158
6		1999	F 1	30-39	1964.5	0	43.9	223
6		1999	F 1	40-49	1954.5	0	39.4	221
6		1999	F 1	50-59	1944.5	0	28.1	210
6		1999	F 1	60-69	1934.5	0	16.4	183
6		1999	F 1	70+	1924.5	0	5.8	121
7		1999	M 0	16-19	1981.5	1	38.7	62
7		1999	M 0	20-29	1974.5	0	45.8	144
7		1999	M 0	30-39	1964.5	0	44.8	154
7		1999	M 0	40-49	1954.5	0	42.2	173
7		1999	M 0	50-59	1944.5	0	31.1	193
7		1999	M 0	60-69	1934.5	0	23.8	189
7		1999	M 0	70+	1924.5	0	14.0	107
7	<i>bijon</i>	1999	F 1	16-19	1981.5	0	46.8	62
7		1999	F 1	20-29	1974.5	0	42.4	158
7		1999	F 1	30-39	1964.5	0	32.7	223
7		1999	F 1	40-49	1954.5	0	29.9	221
7		1999	F 1	50-59	1944.5	0	25.2	210
7		1999	F 1	60-69	1934.5	0	13.7	183
7		1999	F 1	70+	1924.5	0	8.3	121
8		1999	M 0	16-19	1981.5	0	3.2	62
8		1999	M 0	20-29	1974.5	0	4.2	144
8		1999	M 0	30-39	1964.5	0	4.5	154
8		1999	M 0	40-49	1954.5	0	7.5	173
8		1999	M 0	50-59	1944.5	0	2.6	193
8		1999	M 0	60-69	1934.5	0	5.8	189
8		1999	M 0	70+	1924.5	0	0.0	107
8	<i>akauntabiriti</i>	1999	F 1	16-19	1981.5	0	0.0	62
8		1999	F 1	20-29	1974.5	0	5.7	158
8		1999	F 1	30-39	1964.5	0	4.5	223
8		1999	F 1	40-49	1954.5	0	3.2	221
8		1999	F 1	50-59	1944.5	0	0.5	210

ID	LW	Year of Survey	Sex	Age	Year of Birth median	Acquisition	LW(%)	No. of informant
8		1999	F 1	60-69	1934.5	0	2.7	183
8		1999	F 1	70+	1924.5	0	2.5	121
9	<i>wain</i>	2000	M 0	16-19	1982.5	0	95.9	49
9		2000	M 0	20-29	1975.5	0	94.4	108
9		2000	M 0	30-39	1965.5	0	91.7	144
9		2000	M 0	40-49	1955.5	0	83.7	178
9		2000	M 0	50-59	1945.5	0	69.9	193
9		2000	M 0	60+	1935.5	0	52.8	309
9		2000	F 1	16-19	1982.5	1	93.7	63
9		2000	F 1	20-29	1975.5	0	96.3	161
9		2000	F 1	30-39	1965.5	0	92.6	203
9		2000	F 1	40-49	1955.5	0	86.7	210
9		2000	F 1	50-59	1945.5	0	70.7	256
9		2000	F 1	60+	1935.5	0	51.9	322
10	<i>suutsu</i>	2000	M 0	16-19	1982.5	0	85.7	49
10		2000	M 0	20-29	1975.5	0	80.6	108
10		2000	M 0	30-39	1965.5	0	70.8	144
10		2000	M 0	40-49	1955.5	0	53.4	178
10		2000	M 0	50-59	1945.5	0	31.6	193
10		2000	M 0	60+	1935.5	0	21.7	309
10		2000	F 1	16-19	1982.5	1	87.3	63
10		2000	F 1	20-29	1975.5	0	88.2	161
10		2000	F 1	30-39	1965.5	0	71.4	203
10		2000	F 1	40-49	1955.5	0	61.9	210
10		2000	F 1	50-59	1945.5	0	44.1	256
10		2000	F 1	60+	1935.5	0	37.6	322
11	<i>kitchin</i>	2000	M 0	16-19	1982.5	0	12.2	49
11		2000	M 0	20-29	1975.5	0	10.2	108
11		2000	M 0	30-39	1965.5	0	11.1	144
11		2000	M 0	40-49	1955.5	0	6.2	178
11		2000	M 0	50-59	1945.5	0	6.2	193
11		2000	M 0	60+	1935.5	0	8.4	309
11		2000	F 1	16-19	1982.5	0	15.9	63
11		2000	F 1	20-29	1975.5	0	19.3	161
11		2000	F 1	30-39	1965.5	0	13.8	203
11		2000	F 1	40-49	1955.5	0	12.9	210
11		2000	F 1	50-59	1945.5	0	16.4	256
11		2000	F 1	60+	1935.5	0	14.3	322
12	<i>oopun-suru</i>	2000	M 0	16-19	1982.5	1	34.7	49
12		2000	M 0	20-29	1975.5	0	42.6	108
12		2000	M 0	30-39	1965.5	0	35.4	144
12		2000	M 0	40-49	1955.5	0	28.7	178
12		2000	M 0	50-59	1945.5	0	25.4	193
12		2000	M 0	60+	1935.5	0	22.0	309
12		2000	F 1	16-19	1982.5	1	44.4	63
12		2000	F 1	20-29	1975.5	0	48.4	161
12		2000	F 1	30-39	1965.5	0	42.4	203
12		2000	F 1	40-49	1955.5	0	41.0	210
12		2000	F 1	50-59	1945.5	0	38.3	256
12		2000	F 1	60+	1935.5	0	31.4	322
13	<i>getto-suru</i>	2000	M 0	16-19	1982.5	0	44.9	49
13		2000	M 0	20-29	1975.5	0	36.1	108
13		2000	M 0	30-39	1965.5	0	22.2	144
13		2000	M 0	40-49	1955.5	0	6.2	178
13		2000	M 0	50-59	1945.5	0	3.6	193
13		2000	M 0	60+	1935.5	0	1.3	309
13		2000	F 1	16-19	1982.5	0	46.0	63
13		2000	F 1	20-29	1975.5	0	41.0	161
13		2000	F 1	30-39	1965.5	0	25.1	203
13		2000	F 1	40-49	1955.5	0	11.0	210
13		2000	F 1	50-59	1945.5	0	4.7	256
13		2000	F 1	60+	1935.5	0	3.4	322

ID	LW	Year of Survey	Sex	Age	Year of Birth median	Acquisition	LW(%)	No. of informant	
14	<i>tootaru-de</i>	2000	M	0	16-19	1982.5	1	18.4	49
14		2000	M	0	20-29	1975.5	0	24.1	108
14		2000	M	0	30-39	1965.5	0	21.5	144
14		2000	M	0	40-49	1955.5	0	10.7	178
14		2000	M	0	50-59	1945.5	0	13.0	193
14		2000	M	0	60+	1935.5	0	8.7	309
14		2000	F	1	16-19	1982.5	1	17.5	63
14		2000	F	1	20-29	1975.5	0	18.6	161
14		2000	F	1	30-39	1965.5	0	13.3	203
14		2000	F	1	40-49	1955.5	0	13.3	210
14		2000	F	1	50-59	1945.5	0	11.3	256
14		2000	F	1	60+	1935.5	0	9.0	322
15	<i>kyanseru</i>	2003	M	0	15-19	1986.0	1	60.2	88
15		2003	M	0	20-29	1978.5	0	70.4	135
15		2003	M	0	30-39	1968.5	0	65.9	176
15		2003	M	0	40-49	1958.5	0	65.8	184
15		2003	M	0	50-59	1948.5	0	57.9	259
15		2003	M	0	60+	1938.5	0	42.9	545
15		2003	F	1	15-19	1986.0	1	67.3	101
15		2003	F	1	20-29	1978.5	0	78.7	183
15		2003	F	1	30-39	1968.5	0	72.0	271
15		2003	F	1	40-49	1958.5	0	63.3	237
15		2003	F	1	50-59	1948.5	0	55.7	368
15		2003	F	1	60+	1938.5	0	45.0	540
16	<i>happii</i>	2003	M	0	15-19	1986.0	1	20.5	88
16		2003	M	0	20-29	1978.5	0	23.0	135
16		2003	M	0	30-39	1968.5	0	22.7	176
16		2003	M	0	40-49	1958.5	0	14.7	184
16		2003	M	0	50-59	1948.5	0	16.6	259
16		2003	M	0	60+	1938.5	0	13.6	545
16		2003	F	1	15-19	1986.0	0	41.6	101
16		2003	F	1	20-29	1978.5	0	23.0	183
16		2003	F	1	30-39	1968.5	0	19.2	271
16		2003	F	1	40-49	1958.5	0	19.0	237
16		2003	F	1	50-59	1948.5	0	17.9	368
16		2003	F	1	60+	1938.5	0	12.0	540
17	<i>sapooto</i>	2004	M	0	15-19	1987.0	1	37.0	73
17		2004	M	0	20-29	1979.5	0	49.2	128
17		2004	M	0	30-39	1969.5	0	47.8	205
17		2004	M	0	40-49	1959.5	0	42.4	210
17		2004	M	0	50-59	1949.5	0	29.5	258
17		2004	M	0	60-69	1939.5	0	23.5	311
17		2004	M	0	70+	1929.5	0	15.3	236
17		2004	F	1	15-19	1987.0	1	45.3	64
17		2004	F	1	20-29	1979.5	0	58.9	146
17		2004	F	1	30-39	1969.5	0	50.8	303
17		2004	F	1	40-49	1959.5	0	39.1	266
17		2004	F	1	50-59	1949.5	0	32.1	361
17		2004	F	1	60-69	1939.5	0	20.7	314
17		2004	F	1	70+	1929.5	0	7.9	215

B1: Frequency lists by usage, BCCWJ (Ch.5)

a. Nominal use, BCCWJ

	lemmaID	lemma_JP	lemma	meaning	frequency	Is there native single-word equivalents?
1	25819	データ	<i>deeta</i>	data	557	
2	2503	イメージ	<i>imeeji</i>	image	519	
3	11462	ケース	<i>keesu</i>	case	493	1
4	21906	タイプ	<i>taipu</i>	type	407	1
5	35287	ポイント	<i>pointo</i>	point	368	
6	3965	エネルギー	<i>enerugii</i>	energy	339	
7	25641	デザイン	<i>dezain</i>	design	330	
8	14744	サービス	<i>saabisu</i>	service	328	
9	26951	ドア	<i>doa</i>	door	325	
10	15342	システム	<i>sisutemu</i>	system	313	
11	25515	テーマ	<i>teema</i>	theme	308	1
12	32308	ファイル	<i>fairu</i>	file	288	
13	37817	モデル	<i>moderu</i>	model	267	
14	30757	バランス	<i>baransu</i>	balance	261	1
15	40734	レベル	<i>reberu</i>	level	250	1
16	34179	ベッド	<i>beddo</i>	bed	247	
17	25512	テーブル	<i>deeburu</i>	table	243	
18	39979	リスク	<i>risuku</i>	risk	239	1
19	25317	テレビ	<i>terebi</i>	television	234	
20	30953	パソコン	<i>pasokon</i>	personal computer	233	
21	23722	チャンス	<i>chansu</i>	chance	223	1
22	10796	グループ	<i>guruupu</i>	group	222	
23	35281	ボール	<i>booru</i>	ball	218	
24	24311	チーム	<i>chiimu</i>	team	210	
25	19209	スタイル	<i>sutairu</i>	style	209	
26	34761	ホテル	<i>hoteru</i>	hotel	204	
27	37444	メッセージ	<i>messeeji</i>	message	199	
28	37544	メンバー	<i>menbaa</i>	member	197	
29	33767	プログラム	<i>puroguramu</i>	program	195	
30	34353	ページ	<i>peeji</i>	page	190	
31	189	アイディア	<i>aidia</i>	idea	186	
32	13014	コミュニケーション	<i>komyunikeeshon</i>	communication	184	
33	22590	タバコ	<i>tabako</i>	tabaco	177	
34	19384	スピード	<i>supiido</i>	speed	176	1
35	26733	トラブル	<i>toraburu</i>	trouble	169	
36	19214	スタッフ	<i>sutaffu</i>	staff	163	
37	40548	ルール	<i>ruuru</i>	rule	163	1
38	30956	パターン	<i>pataan</i>	pattern	161	
39	39646	ライン	<i>rain</i>	line	159	
40	4216	エンジン	<i>enjin</i>	engine	158	
41	12723	コスト	<i>kosuto</i>	cost	158	1
42	28860	ネットワーク	<i>nettowaaku</i>	network	157	
43	40836	レース	<i>reesu</i>	race	155	
44	7169	カメラ	<i>kamera</i>	camera	154	
45	17670	シーン	<i>shin</i>	scene	153	
46	19308	ストレス	<i>sutoresu</i>	stress	152	
47	33774	プロセス	<i>purosesu</i>	process	152	1
48	34288	ベース	<i>beesu</i>	base	152	
49	13259	コンピューター	<i>konpyuutaa</i>	computer	151	
50	35173	ボタン	<i>botan</i>	button	151	
51	37486	メリット	<i>meritto</i>	merit	149	1
52	11778	ゲーム	<i>geemu</i>	game	147	
53	28673	ニーズ	<i>niizu</i>	needs	147	1

	lemmaID	lemma_JP	lemma	meaning	frequency	Is there native single-word equivalents?
54	37556	メール	<i>meeru</i>	mail	146	
55	10431	クラス	<i>kurasu</i>	class	140	
56	21876	タイトル	<i>taitoru</i>	title	139	
57	37464	メニュー	<i>menyuu</i>	menu	137	
58	38965	ユーザー	<i>yuuzaa</i>	user	137	
59	83696	バス	<i>basu</i>	bus	134	
60	32332	ファン	<i>fan</i>	fan (enthusiast)	133	
61	2776	インターネット	<i>intaanetto</i>	internet	132	
62	13299	コース	<i>koosu</i>	course	130	
63	103238	プロ	<i>puro</i>	pro	130	
64	2451	イベント	<i>ibento</i>	event	129	1
65	31044	パワー	<i>pawaa</i>	power	129	
66	13746	サイズ	<i>saizu</i>	size	126	1
67	32011	ビジネス	<i>bijinesu</i>	business	124	
68	36149	マンション	<i>manshon</i>	mansion	124	
69	13315	コーヒー	<i>kooonii</i>	coffee	123	
70	26491	トップ	<i>toppu</i>	top	120	
71	41132	ワイン	<i>wain</i>	wine	118	1
72	7871	カード	<i>kaado</i>	card	113	
73	31046	パン	<i>pan</i>	pao	113	
74	32164	ビール	<i>biiru</i>	beer	113	
75	10449	クラブ	<i>kurabu</i>	club	111	
76	17025	ショック	<i>shokku</i>	shock	111	
77	19421	スペース	<i>supeesu</i>	space	111	
78	40684	レストラン	<i>resutoran</i>	restaurant	111	
79	20229	セット	<i>setto</i>	set	110	
80	187	アイテム	<i>aitemu</i>	item	109	
81	21928	タイミング	<i>taimingu</i>	timing	109	
82	33772	プロジェクト	<i>purojekuto</i>	project	109	
83	281548	グラス	<i>gurasu</i>	glass	109	
84	17657	シート	<i>shiito</i>	sheet	108	
85	32120	ビル	<i>biru</i>	building	108	
86	32229	ピンク	<i>pinku</i>	pink	108	
87	21416	ソフト	<i>sofuto</i>	soft	107	
88	2931	ウィンドー	<i>uindoo</i>	window	106	
89	28590	ニュース	<i>nyuusu</i>	news	104	
90	39990	リズム	<i>rizumu</i>	rhythm	103	
91	31081	パーティー	<i>paatii</i>	party	102	
92	30678	バッグ	<i>baggu</i>	bag	100	

b. Verbal use, BCCWJ

	lemmaID	lemma_JP	lemma	meaning	frequency	Is there native single-word equivalents?
1	10494	クリック	<i>kurikku</i>	click	898	
2	23405	チェック	<i>chekku</i>	check	267	
3	19233	スタート	<i>sutaato</i>	start	145	1
4	2503	イメージ	<i>imeeji</i>	image	130	1
5	13232	コントロール	<i>kontorooru</i>	control	114	
6	6986	カバー	<i>kabaa</i>	cover	92	
7	116693	ドラッグ	<i>doraggu</i>	drag	89	
8	795	アップ	<i>appu</i>	up	85	
9	458	アクセス	<i>akusesu</i>	access	84	
10	20229	セット	<i>setto</i>	set	80	
11	25641	デザイン	<i>dezain</i>	design	73	
12	939	アピール	<i>apiiru</i>	appeal	71	
13	14371	サポート	<i>sapooto</i>	support	69	1
14	6756	カット	<i>katto</i>	cut	66	
15	5554	オープン	<i>oopun</i>	open	62	1
16	10471	クリア	<i>kuria</i>	clear	58	
17	40345	リラックス	<i>rirakkusu</i>	relax	58	1
18	2758	インストール	<i>insutooru</i>	install	57	
19	33702	プラス	<i>purasu</i>	plus	55	
20	40443	リード	<i>riido</i>	lead	48	
21	12939	コピー	<i>kopii</i>	copy	47	
22	23716	チャレンジ	<i>charenji</i>	challenge	47	1
23	33749	プレゼント	<i>purezento</i>	present	47	
24	33757	プレー	<i>puree</i>	play	44	
25	35712	マスター	<i>masutaa</i>	master	44	1
26	23101	ダウンロード	<i>daunroodo</i>	download	42	
27	9743	キープ	<i>kiipu</i>	keep	40	1
28	15683	シフト	<i>shifuto</i>	shift	40	1
29	8884	キャッチ	<i>kyatchi</i>	catch	39	
30	40352	リリース	<i>ririisu</i>	release	39	
31	35818	マッチ	<i>matchi</i>	match	36	
32	1284	アレンジ	<i>arenji</i>	arrange	35	
33	878	アドバイス	<i>adobaisu</i>	advice	30	1

c. Adjectival use, BCCWJ

	lemmaID	lemma_JP	lemma	meaning	frequency	Is there native single-word equivalents?
1	17538	シンプル	<i>shinpuru</i>	simple	125	1
2	19482	スムーズ	<i>sumuuzu</i>	smooth	108	1
3	38878	ユニーク	<i>yuniiku</i>	unique	59	1
4	15484	シック	<i>shikku</i>	chic	47	
5	37739	モダン	<i>modan</i>	modern	46	
6	39877	リアル	<i>riaru</i>	real	46	
7	10683	クール	<i>kuru</i>	cool	44	
8	9114	キュート	<i>kyuuto</i>	cute	38	
9	21416	ソフト	<i>sofuto</i>	soft	35	
10	34141	ベスト	<i>besuto</i>	best	34	
11	41022	ロマンチック	<i>romanchikku</i>	romantic	34	
12	17025	ショック	<i>shokku</i>	shock	31	

B2: Distribution of LWs and non-LWs by Year of Birth, BCCWJ (Ch. 5)

a. Frequency by Year of Birth

	variable	1=LW 0=nonLW	1830-90	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	total
1	advice	1				2	9	6	17	5	1			40
		0		2	1	8	24	25	21	5	3	1		90
2	balance	1	1	19	35	47	120	146	110	18				496
		0	3	1	11	17	25	56	14	4				131
3	case	1	3	11	68	176	277	309	236	33	1			1114
		0	77	52	235	838	2110	3191	3266	2459	628	7		12863
4	challenge	1		1	3	17	34	15	19	2				94
		0	1	1	22	40	61	64	33	17	1			240
5	chance	1		3	19	72	98	95	62	21	2	1		375
		0	12	11	15	114	217	304	222	172	46	6		1119
6	event	1			2	16	39	41	52	10		3		163
		0	10	7	14	54	135	196	239	130	30	3	1	819
7	image	1			9	20	42	57	53	13	1			195
		0	1	5	5	27	56	57	67	42	19	2		281
8	keep	1			1	3	2	10	23	5				45
		0	12	1	19	94	236	427	411	296	42	1		1539
9	level	1			29	84	188	157	123	44	1			626
		0	7	1	17	34	128	115	79	23				404
10	master	1			1	5	4	12	23	17	11	1		74
		0			2	3	4	4						13
11	merit	1			8	16	60	77	78	12	1			253
		0		8	43	81	106	96	75	14	1			424
12	needs	1		2	3	51	92	101	53	4				306
		0	8	5	24	73	144	188	216	99	16			773
13	open	1			1	5	18	23	17	3	1			69
		0	70	40	79	394	767	1079	1107	957	388	47	3	4931
14	process	1		1	2	21	48	105	107	73	30	1		388
		0	27	13	94	419	723	1179	1250	753	200	17	1	4676
15	relax	1		3	23	6	8	28	24	4				96
		0	3	1	9	13	20	20	21	5				92
16	risk	1			15	26	97	119	118	18				393
		0	5	6	15	84	172	194	257	128	57	9		927
17	rule	1			1	16	43	81	98	78	13		1	331
		0			1	23	28	58	22	22	7			161
18	shift	1			7	7	16	30	9	3				72
		0	1	2	29	51	78	85	48	7				301
19	simple	1			2	15	29	45	53	19				164
		0	3	1	5	32	74	124	115	95	23	2		474
20	smooth	1			8	35	47	68	33	11				204
		0	2	1	5	10	28	22	20	7	1			96
21	start	1		1	10	40	53	89	47	10				252
		0	84	85	161	816	1783	2315	2143	1602	384	48	1	9422
22	support	1			6	17	28	48	10					112
		0	9	9	18	152	280	474	443	300	79	6		1770
23	theme	1		2	30	105	183	170	83	14	4			598
		0	1	1	14	43	35	35	31	3				165
24	type	1	3	4	10	27	87	174	202	155	30	2		694
		0	9	4	4	29	58	92	57	67	10			330
25	unique	1			8	16	21	21	24	2				93
		0	1	2	1	21	33	51	39	39	15	1		203
	total	N	353	256	781	3682	8086	12258	12448	9080	2368	168	11	*49491
		%	0.7%	0.5%	1.6%	7.4%	16.3%	24.8%	25.2%	18.3%	4.8%	0.3%	0.0%	100.0%

The total count is not equal to that in Table 5.7 since the author's year of birth is unknown for 25672 out of 75163 tokens.

b. Result of the regression analysis

	% of LW as opposed to its non-LW variants					Regression analysis					
	variable	1920	1930	1940	1950	1960	B	intercept	R ²	p(t-test)	
1	advice	20.0	27.3	19.4	44.7	50.0	0.7746	-1470.5309	0.74	0.062	+
2	balance	76.1	73.4	82.8	72.3	88.7	0.2409	-388.5983	0.30	0.338	ns
3	case	7.5	7.7	8.0	8.6	8.8	0.0345	-58.7572	0.95	0.005	**
4	challenge	12.0	29.8	35.8	19.0	36.5	0.3824	-715.2223	0.31	0.325	ns
5	chance	14.3	24.9	24.4	30.0	26.5	0.2947	-547.8064	0.63	0.107	ns
6	event	3.6	10.6	16.6	14.6	28.6	0.5405	-1033.7130	0.87	0.021	*
7	image	25.0	26.3	42.4	46.0	55.8	0.8123	-1536.7800	0.94	0.006	**
8	keep	1.1	1.3	0.5	2.4	7.2	0.1343	-258.1645	0.60	0.123	ns
9	level	63.0	71.2	59.5	57.7	60.9	-0.1777	407.2173	0.29	0.353	ns
10	master	71.4	57.1	75.0	85.2	100.0	0.8519	-1574.8413	0.71	0.073	+
11	merit	15.7	16.5	36.1	44.5	51.0	0.9860	-1880.1170	0.94	0.007	**
12	needs	3.9	26.2	32.9	31.9	34.9	0.6755	-1284.5219	0.71	0.075	+
13	open	0.3	0.6	1.6	2.0	1.7	0.0437	-83.5554	0.81	0.038	*
14	process	4.8	6.2	8.2	7.9	8.8	0.0979	-182.7356	0.88	0.019	*
15	relax	71.9	31.6	28.6	58.3	53.3	-0.1033	249.1200	0.01	0.887	ns
16	risk	15.2	13.1	33.3	31.6	48.0	0.8415	-1604.2548	0.86	0.024	*
17	rule	41.0	60.6	58.3	81.7	78.0	0.9505	-1780.1031	0.84	0.030	*
18	shift	19.4	12.1	17.0	26.1	15.8	0.0671	-112.0539	0.04	0.742	ns
19	simple	5.9	16.9	19.0	28.1	35.8	0.7113	-1358.7576	0.97	0.002	**
20	smooth	61.5	77.8	62.7	75.6	62.3	-0.0077	82.9149	0.0002	0.981	ns
21	start	1.2	2.2	2.2	4.0	2.9	0.0507	-95.9082	0.62	0.115	ns
22	support	0.0	2.1	3.5	5.9	13.8	0.3143	-604.7426	0.87	0.022	*
23	theme	68.2	70.9	83.9	82.9	72.8	0.2123	-336.1256	0.22	0.429	ns
24	type	48.2	60.0	65.4	78.0	69.8	0.6120	-1123.0569	0.75	0.056	+
25	unique	27.6	32.7	29.2	35.0	38.1	0.2337	-420.7808	0.75	0.058	+

B3: Distribution of LWs and non-LWs by Sex, BCCWJ (Ch.5)

	variable	1=LW 0=non-LW	F	M	total	X ² (1)	p-value	
1	advice	1 0	6 10	33 72	39 82	.234	.628	ns
2	balance	1 0	102 7	346 113	448 120	17.503	<.001	**
3	case	1 0	106 1004	943 10681	1049 11685	2.768	.096	+
4	challenge	1 0	11 30	75 185	86 215	.071	.790	ns
5	chance	1 0	58 169	283 844	341 1013	.019	.889	ns
6	event	1 0	31 124	114 617	145 741	1.813	.178	ns
7	image	1 0	23 46	157 200	180 246	2.685	.101	ns
8	keep	1 0	6 109	32 1328	38 1437	3.466	.063	-
9	level	1 0	43 24	520 343	563 367	.401	.527	ns
10	master	1 0	11 2	49 11	60 13	.063	.801	-
11	merit	1 0	25 40	209 356	234 396	.054	.816	ns
12	needs	1 0	35 78	265 635	300 713	.113	.737	ns
13	open	1 0	14 874	45 3334	59 4208	.309	.578	ns
14	process	1 0	49 517	301 3749	350 4266	1.064	.302	ns
15	relax	1 0	16 34	71 46	87 80	11.549	.001	**
16	risk	1 0	29 133	344 686	373 819	15.636	<.001	**
17	rule	1 0	34 12	277 139	311 151	1.011	.315	ns
18	shift	1 0	3 20	64 268	67 288	.546	.460	-
19	simple	1 0	42 50	94 382	136 432	28.409	<.001	**
20	smooth	1 0	36 16	153 67	189 83	.002	.965	ns
21	start	1 0	33 1447	204 7044	237 8491	1.591	.207	ns
22	support	1 0	23 273	69 1359	92 1632	4.190	.041	*
23	theme	1 0	100 14	463 144	563 158	7.344	.007	**
24	type	1 0	106 36	527 266	633 302	3.696	.055	+
25	unique	1 0	16 25	73 156	89 181	.804	.370	ns

“-” indicates that more than 1 cell has expected count less than 5.

B4: Distribution of LWs and non-LWs by Education, BCCWJ (Ch.5)

	variable	1=LW 0=non-LW	up to UNIV	GRAD	total	X ² (1)	p-value	
1	advice	1 0	17 33	11 26	28 59	.178	.673	ns
2	balance	1 0	168 36	124 45	292 81	4.385	.036	*
3	case	1 0	429 4351	337 4001	766 8352	4.301	.038	*
4	challenge	1 0	52 92	12 42	64 134	3.463	.063	+
5	chance	1 0	171 409	36 307	207 716	44.664	<.001	**
6	event	1 0	76 320	29 238	105 558	8.303	.004	**
7	image	1 0	63 130	48 47	111 177	8.596	.003	**
8	keep	1 0	7 503	2 564	9 1067	3.360	.067	-
9	level	1 0	201 108	211 187	412 295	10.359	.001	**
10	master	1 0	22 4	14 4	36 8	.334	.563	-
11	merit	1 0	97 164	66 122	163 286	.200	.655	ns
12	needs	1 0	106 273	107 257	213 530	.185	.667	ns
13	open	1 0	29 2122	2 662	31 2784	5.107	.024	*
14	process	1 0	121 1559	136 1498	257 3057	1.454	.228	ns
15	relax	1 0	39 35	14 9	53 44	.472	.492	ns
16	risk	1 0	127 425	134 187	261 612	34.000	<.001	**
17	rule	1 0	102 45	121 64	223 109	.589	.443	ns
18	shift	1 0	20 121	39 113	59 234	5.988	.014	*
19	simple	1 0	55 174	23 154	78 328	7.816	.005	**
20	smooth	1 0	81 36	35 15	116 51	.010	.921	ns
21	start	1 0	110 4025	37 1981	147 6006	3.975	.046	*
22	support	1 0	45 543	28 607	73 1150	5.723	.017	*
23	theme	1 0	218 42	181 81	399 123	15.790	<.001	**
24	type	1 0	253 116	169 98	422 214	1.925	.165	ns
25	unique	1 0	48 57	12 68	60 125	19.547	<.001	**

“-” indicates that more than 1 cell has expected count less than 5.

B5: Distribution of LWs and non-LWs by Register, BCCWJ (Ch.5)

	variable	1=LW 0=non-LW	PB	PM	total	X ² (1)	p-value	
1	advice	1 0	33 79	6 4	39 83	3.936	.047	-
2	balance	1 0	397 117	61 6	458 123	6.771	.009	**
3	case	1 0	973 11340	93 524	1066 11864	39.938	<.001	**
4	challenge	1 0	83 198	5 22	88 220	1.466	.226	ns
5	chance	1 0	320 964	26 65	346 1029	.601	.438	ns
6	event	1 0	128 701	22 53	150 754	9.592	.002	**
7	image	1 0	159 220	22 29	181 249	.026	.872	ns
8	keep	1 0	23 1377	16 87	39 1464	73.248	<.001	-
9	level	1 0	537 347	44 26	581 373	.121	.728	ns
10	master	1 0	50 11	11 2	61 13	.052	.820	-
11	merit	1 0	207 379	32 22	239 401	12.106	.001	**
12	needs	1 0	289 687	11 33	300 720	.431	.511	ns
13	open	1 0	52 3900	12 404	64 4304	6.416	.011	*
14	process	1 0	345 4127	9 197	354 4324	3.152	.076	+
15	relax	1 0	76 69	13 14	89 83	.166	.684	ns
16	risk	1 0	342 775	33 60	375 835	.951	.330	ns
17	rule	1 0	293 150	23 3	316 153	5.567	.018	*
18	shift	1 0	64 266	5 25	69 291	.132	.716	ns
19	simple	1 0	111 406	33 34	144 440	24.646	<.001	**
20	smooth	1 0	169 71	22 14	191 85	1.272	.259	ns
21	start	1 0	200 7920	39 739	239 8659	17.659	<.001	**
22	support	1 0	90 1539	9 110	99 1649	.862	.353	ns
23	theme	1 0	508 152	63 6	571 158	7.561	.006	**
24	type	1 0	591 286	54 17	645 303	2.269	.132	ns
25	unique	1 0	75 167	15 16	90 183	3.763	.052	+

“-” indicates that more than 1 cell has expected count less than 5.

B6: Frequency lists by usage, CSJ (Ch.5)

a. Nominal use, CSJ

	lemmaID	lemma_JP	lemma	meaning	frequency	Investigated with BCCWJ?
1	25819	データ	<i>deeta</i>	data	1879	
2	37817	モデル	<i>moderu</i>	model	1348	
3	15342	システム	<i>shisutemu</i>	system	1139	
4	30956	パターン	<i>pataan</i>	pattern	656	
5	25317	テレビ	<i>terebi</i>	television	610	
6	10761	グラフ	<i>gurafu</i>	graph	567	
7	21876	タイトル	<i>taitoru</i>	title	558	
8	38965	ユーザー	<i>yuuzaa</i>	user	531	
9	25515	テーマ	<i>teema</i>	theme	522	1
10	2503	イメージ	<i>imeeji</i>	image	505	
11	10796	グループ	<i>guruupu</i>	group	476	
12	463	アクセント	<i>akusento</i>	accent	470	
13	20754	ゼロ	<i>zero</i>	zero	459	
14	40548	ルール	<i>ruuru</i>	rule	458	1
15	31017	パラメーター	<i>parameetaa</i>	parameter	453	
16	40734	レベル	<i>reberu</i>	level	446	1
17	25097	テキスト	<i>tekisuto</i>	text	442	
18	21906	タイプ	<i>taipu</i>	type	390	1
19	10431	クラス	<i>kurasu</i>	class	387	
20	34761	ホテル	<i>hoteru</i>	hotel	369	
21	13014	コミュニケーション	<i>komyunikeeshon</i>	communication	353	
22	28590	ニュース	<i>nyuusu</i>	news	342	
23	13314	コーパス	<i>koopasu</i>	corpus	290	
24	22299	タスク	<i>tasuku</i>	task	288	
25	25820	データベース	<i>deetaabeesu</i>	database	283	
26	83696	バス	<i>basu</i>	bus	282	
27	30953	パソコン	<i>pasokon</i>	personal computer	274	
28	6838	カテゴリー	<i>kategorii</i>	category	266	
29	32202	ピッチ	<i>pitchi</i>	pitch	263	
30	4318	エージェント	<i>eejento</i>	agent	254	
31	13259	コンピューター	<i>konpyuutaa</i>	computer	253	
32	35287	ポイント	<i>pointo</i>	point	247	
33	22590	タバコ	<i>tabako</i>	tabaco	243	
34	2776	インターネット	<i>intanetto</i>	internet	239	
35	19308	ストレス	<i>sutoresu</i>	stress	234	
36	10434	クラスター	<i>kurasutaa</i>	cluster	229	
37	25149	テスト	<i>tesuto</i>	test	226	
38	19429	スポーツ	<i>supootsu</i>	sport	224	
39	1226	アルゴリズム	<i>arugorizumu</i>	algorithm	218	
40	36149	マンション	<i>manshon</i>	mansion	215	
41	37556	メール	<i>meeru</i>	mail	215	
42	32172	ピアノ	<i>piano</i>	piano	210	
43	19147	スコア	<i>sukoa</i>	score	209	
44	35281	ボール	<i>booru</i>	ball	207	
45	29314	ノード	<i>noodo</i>	node	205	
46	11462	ケース	<i>keesu</i>	case	185	1
47	973	アプローチ	<i>apuroochi</i>	approach	184	
48	32370	フィラー	<i>firaa</i>	filler	182	
49	24311	チーム	<i>chiimu</i>	team	180	
50	19413	スペクトル	<i>supekutoru</i>	spectre	177	
51	31046	パン	<i>pan</i>	pao	170	
52	38078	モーラ	<i>moora</i>	mora	166	
53	19084	スキー	<i>sukii</i>	ski	164	
54	25844	トイレ	<i>toire</i>	toilet	161	

	lemmaID	lemma_JP	lemma	meaning	frequency	Investigated with BCCWJ?
55	27505	ナイフ	<i>naifu</i>	knife	161	
56	33767	プログラム	<i>puroguramu</i>	program	161	
57	35360	ポーズ	<i>poozu</i>	pause	161	
58	39827	ランダム	<i>randamu</i>	random	161	
59	37544	メンバー	<i>menbaa</i>	member	160	
60	2451	イベント	<i>ibento</i>	event	159	1
61	32258	ピーク	<i>piiku</i>	peak	158	
62	19581	スーパー	<i>suupaa</i>	super	156	
63	34288	ベース	<i>beesu</i>	base	155	
64	37486	メリット	<i>meritto</i>	merit	152	1
65	33206	フレーム	<i>fureemu</i>	frame	150	
66	190045	ベクトル	<i>bekutoru</i>	vector	145	
67	33772	プロジェクト	<i>purojekuto</i>	project	143	
68	41132	ワイン	<i>wain</i>	wine	141	
69	1245	アルバイト	<i>arubaito</i>	Arbeit	140	
70	33774	プロセス	<i>purosesu</i>	process	139	1
71	39982	リスト	<i>risuto</i>	list	136	
72	19380	スピーカー	<i>supiikaa</i>	speaker	135	
73	103238	プロ	<i>puro</i>	pro	135	
74	32042	ビデオ	<i>bideo</i>	video	133	
75	120760	バイト	<i>baito</i>	Arbeit	131	
76	11778	ゲーム	<i>geemu</i>	game	129	
77	17025	ショック	<i>shokku</i>	shock	129	
78	15768	シミュレーション	<i>shumireeshon</i>	simulation	125	
79	13746	サイズ	<i>saizu</i>	size	124	
80	39710	ラジオ	<i>rajio</i>	radio	123	
81	41016	ロボット	<i>robotto</i>	robot	122	
82	2817	イントネーション	<i>intoneeshon</i>	intonation	121	
83	27357	ドラマ	<i>dorama</i>	drama	119	
84	14744	サービス	<i>saabisu</i>	service	118	
85	25231	テニス	<i>tenisu</i>	tennis	118	
86	21416	ソフト	<i>sofuto</i>	soft	117	
87	24325	ツアー	<i>tsuaa</i>	tour	117	
88	39990	リズム	<i>rizumu</i>	rhythm	117	
89	3965	エネルギー	<i>enerugii</i>	energy	115	
90	7169	カメラ	<i>kamera</i>	camera	112	
91	34296	ペア	<i>pea</i>	pair	112	
92	7871	カード	<i>kaado</i>	card	111	
93	30860	バンド	<i>bando</i>	band (group)	111	
94	37453	メディア	<i>media</i>	media	111	
95	40684	レストラン	<i>resutoran</i>	restaurant	111	
96	30757	バランス	<i>baransu</i>	balance	110	1
97	22900	ターゲット	<i>taagetto</i>	target	109	
98	37558	メイン	<i>meen</i>	main	109	
99	13315	コーヒー	<i>koohee</i>	coffee	108	
100	19384	スピード	<i>supiido</i>	speed	108	
101	20138	セグメント	<i>segumento</i>	segment	108	
102	30534	バイク	<i>baiku</i>	bike	107	
103	888	アナウンス	<i>anaansu</i>	announce	106	
104	14741	サークル	<i>saakuru</i>	circle	106	
105	39960	リサイクル	<i>risaikuru</i>	recycle	105	
106	14689	サンプル	<i>sanpuru</i>	sample	104	
107	34316	ペット	<i>petto</i>	pet	104	
108	19382	スピーチ	<i>supiichi</i>	speech	103	
109	2949	ウェブ	<i>uebu</i>	web	102	
110	26882	トレーニング	<i>toreeningu</i>	training	102	
111	25683	デパート	<i>depaato</i>	department	101	
112	33702	プラス	<i>purasu</i>	plus	101	
113	26951	ドア	<i>doa</i>	door	100	

b. Verbal use, CSJ

	lemmaID	lemma_JP	lemma	meaning	frequency	Investigated with BCCWJ ?
1	23405	チェック	<i>chekku</i>	check	141	
2	33783	プロット	<i>purotto</i>	plot	91	
3	19233	スタート	<i>sutaato</i>	start	82	1
4	2503	イメージ	<i>imeeji</i>	image	68	1
5	458	アクセス	<i>akusesu</i>	access	65	
6	10433	クラスタリング	<i>kurasutaringu</i>	clustering	61	
7	40345	リラックス	<i>rirakkusu</i>	relax	61	1
8	6986	カバー	<i>kabaa</i>	cover	58	
9	13232	コントロール	<i>kontooru</i>	control	55	
10	36202	マーク	<i>maaku</i>	mark	53	
11	6756	カット	<i>katto</i>	cut	50	

c. Adjectival use, CSJ

	lemmaID	lemma_JP	lemma	meaning	frequency	Investigated with BCCWJ ?
1	39827	ランダム	<i>randamu</i>	random	154	
2	17025	ショック	<i>shokku</i>	shock	66	
3	19482	スムーズ	<i>sumuuzu</i>	smooth	54	1
4	17538	シンプル	<i>sinpuru</i>	simple	52	1

B7: Distribution of LWs and non-LWs by Year of Birth, CSJ (Ch.5)

a. Frequency by Year of Birth

	variable	1=LW 0=nonLW	1910	1920	1930	1940	1950	1960	1970	1980	total
1	balance	1 0			9	15	24	26	34	1	109
2	case	1 0	20	11	12	16	44	57	56	9	185
3	event	1 0			6	6	10	48	78	9	157
4	image	1 0			5	6	10	32	15	4	68
5	level	1 0	1		7	58	84	141	154	2	446
6	merit	1 0			8	15	16	47	64	2	152
7	process	1 0	1	3	6	23	18	33	57	13	138
8	relax	1 0			1	3	21	19	16	1	61
9	rule	1 0			14	8	37	105	287	4	455
10	simple	1 0		1	5	4	8	20	15		52
11	smooth	1 0			2	6	11	14	20	1	54
12	start	1 0	7	2	9	26	9	22	15	1	82
13	theme	1 0	3		60	61	110	146	138	4	522
14	type	1 0	1	1	8	45	51	133	149	1	389
	total	N %	43 0.3%	18 0.1%	776 4.6%	1828 10.9%	2349 14.0%	4498 26.8%	7148 42.6%	103 0.6%	*16763 100.0%

* The total count is not equal to that in Table 5.9 since the author's year of birth is unknown for 131 out of 16894 tokens

b. Result of the regression analysis, CSJ

	variable	% of LW as opposed to its non-LW variants				Regression analysis				
		1940	1950	1960	1970	B	intercept	R ²	p(t-test)	
1	balance	100.0	100.0	96.3	100.0	-0.0370	171.4815	0.07	0.742	ns
2	case	1.9	4.3	2.7	1.6	-0.0262	53.8550	0.08	0.713	ns
3	event	10.9	20.8	42.5	42.6	1.1679	-2253.9579	0.90	0.053	+
4	image	46.2	45.5	74.4	34.1	-0.0722	191.2734	0.003	0.946	ns
5	level	84.1	91.3	91.6	88.0	0.1208	-147.4373	0.20	0.556	ns
6	merit	46.9	50.0	50.5	39.5	-0.2157	468.4007	0.30	0.452	ns
7	process	9.8	5.1	4.9	4.9	-0.1481	295.7170	0.63	0.208	ns
8	relax	100.0	95.5	100.0	84.2	-0.4282	932.1053	0.55	0.258	ns
9	rule	32.0	74.0	65.6	62.4	0.8280	-1560.2147	0.34	0.417	ns
10	simple	14.3	11.3	12.3	9.3	-0.1383	282.1562	0.74	0.140	ns
11	smooth	66.7	52.4	66.7	51.3	-0.3187	682.2711	0.23	0.520	ns
12	start	6.5	2.4	3.9	1.9	-0.1214	241.0030	0.57	0.243	ns
13	theme	85.9	89.4	84.4	59.5	-0.8434	1728.5750	0.63	0.206	ns
14	type	88.2	67.1	84.7	70.0	-0.3724	805.5167	0.21	0.543	ns

B8: Distribution of LWs and non-LWs by Sex, CSJ (Ch.5)

	variable	1=LW 0=non-LW	F	M	total	X ² (1)	p-value	
1	balance	1 0	52	47 1	99 1	1.094	.296	-
2	case	1 0	51 1399	122 5923	173 7322	11.654	.001	**
3	event	1 0	58 147	84 110	142 257	9.792	.002	**
4	image	1 0	24 26	39 33	63 59	.449	.503	ns
5	level	1 0	131 7	306 46	437 53	6.570	.0104	*
6	merit	1 0	47 44	95 133	142 177	2.624	.105	ns
7	process	1 0	12 373	119 1920	131 2293	4.684	.030	*
8	relax	1 0	29 2	30 2	59 4	.001	.974	-
9	rule	1 0	31 21	406 237	437 258	.256	.613	ns
10	simple	1 0	16 66	31 309	47 375	7.213	.007	**
11	smooth	1 0	21 12	30 27	51 39	1.031	.310	ns
12	start	1 0	26 969	46 1067	72 2036	3.679	.055	+
13	theme	1 0	180 39	275 105	455 144	7.342	.007	**
14	type	1 0	159 52	219 67	378 119	.099	.753	ns

“-” indicates that more than 1 cell has expected count less than 5.

B9: Distribution of LWs and non-LWs by Education, CSJ (Ch.5)

	variable	1=LW 0=non-LW	UNIV	GRAD	total	X ² (1)	p-value	
1	balance	1 0	66 1	33	99 1	.498	.481	-
2	case	1 0	79 2861	94 4454	173 7315	3.044	.081	+
3	event	1 0	110 205	28 51	138 256	.008	.931	ns
4	image	1 0	47 42	16 17	63 59	.180	.671	ns
5	level	1 0	186 14	251 39	437 53	5.102	.024	*
6	merit	1 0	105 110	36 67	141 177	5.441	.020	*
7	process	1 0	35 852	96 1438	131 2290	5.871	.015	*
8	relax	1 0	44 4	15	59 4	1.335	.248	-
9	rule	1 0	156 99	281 159	437 258	.499	.480	ns
10	simple	1 0	29 122	18 253	47 375	15.464	<.001	**
11	smooth	1 0	39 23	12 16	51 39	3.157	.076	+
12	start	1 0	60 1542	11 494	71 2036	2.896	.089	+
13	theme	1 0	322 74	130 70	452 144	19.299	<.001	**
14	type	1 0	183 47	195 72	378 119	2.895	.089	+

“-” indicates that more than 1 cell has expected count less than 5.

B10: Distribution of LWs and non-LWs by Register, CSJ (Ch.5)

	variable	1=LW 0=non-LW	APS	SPS	total	X ² (1)	p-value	
1	balance	1 0	33	66 1	99 1	.498	.481	-
2	case	1 0	103 6361	70 961	173 7322	106.465	<.001	**
3	event	1 0	32 47	110 210	142 257	1.039	.308	ns
4	image	1 0	19 16	44 43	63 59	.138	.711	ns
5	level	1 0	308 45	129 8	437 53	4.883	.027	*
6	merit	1 0	54 115	88 62	142 177	22.961	<.001	**
7	process	1 0	120 1956	11 337	131 2293	4.000	.045	*
8	relax	1 0	13	46 4	59 4	1.111	.292	-
9	rule	1 0	405 242	32 16	437 258	.317	.573	ns
10	simple	1 0	28 292	19 83	47 375	7.625	.006	**
11	smooth	1 0	19 24	32 15	51 39	5.223	.022	*
12	start	1 0	14 535	58 1501	72 2036	1.686	.194	ns
13	theme	1 0	138 141	317 3	455 144	200.804	<.001	**
14	type	1 0	232 84	146 35	378 119	3.317	.069	+

“-” indicates that more than 1 cell has expected count less than 5.

B11: Distribution of LWs and non-LWs by Spontaneity, CSJ (Ch.5)

	variable	1=LW 0=non-LW	others	Spontaneous	total	X ² (1)	p-value	
1	balance	1 0	39	59 1	98 1	.657	.418	-
2	case	1 0	88 3122	85 4155	173 7277	4.371	.037	*
3	event	1 0	22 109	114 148	136 257	27.548	<.001	**
4	image	1 0	14 12	49 47	63 59	.064	.800	ns
5	level	1 0	151 26	285 27	436 53	4.257	.039	*
6	merit	1 0	46 88	94 88	140 176	9.382	.002	**
7	process	1 0	50 935	79 1341	129 2276	.272	.602	ns
8	relax	1 0	21 2	38 2	59 4	.335	.562	-
9	rule	1 0	233 154	199 99	432 253	3.122	.077	+
10	simple	1 0	5 126	42 246	47 372	10.481	.001	**
11	smooth	1 0	15 17	36 22	51 39	1.939	.164	ns
12	start	1 0	25 624	47 1408	72 2032	.525	.469	ns
13	theme	1 0	154 105	296 39	450 144	66.422	<.001	**
14	type	1 0	145 70	232 48	377 118	15.918	<.001	**

“-” indicates that more than 1 cell has expected count less than 5.

B12: Distribution of LWs and non-LWs by Formality, CSJ (Ch.5)

	variable	1=LW 0=non-LW	others	Formal	total	X ² (1)	p-value	
1	balance	1 0	67	30 1	97 1	2.184	.139	-
2	case	1 0	111 4187	57 2354	168 6541	.302	.583	ns
3	event	1 0	118 172	17 82	135 254	18.014	<.001	**
4	image	1 0	48 42	14 16	62 58	.400	.527	ns
5	level	1 0	279 25	128 20	407 45	3.107	.078	+
6	merit	1 0	105 107	33 57	138 164	4.212	.040	*
7	process	1 0	69 1441	56 651	125 2092	10.165	.001	**
8	relax	1 0	50 4	9	59 4	.712	.399	-
9	rule	1 0	244 117	168 123	412 240	6.732	.009	**
10	simple	1 0	42 270	4 82	46 352	5.119	.024	*
11	smooth	1 0	33 20	16 7	49 27	.373	.541	ns
12	start	1 0	52 1511	19 480	71 1991	.263	.608	ns
13	theme	1 0	324 41	116 100	440 141	90.774	<.001	**
14	type	1 0	237 56	128 61	365 117	10.829	.001	**

“-” indicates that more than 1 cell has expected count less than 5.

B13: Distribution of LWs and non-LWs by Audience Size, CSJ (Ch.5)

	variable	1=LW 0=non-LW	1to10	20to50	100to300	total	X ² (2)	p-value	
1	balance	1 0	66 1	18	15	99 1	.498	.780	-
2	case	1 0	70 958	45 2598	54 3618	169 7174	108.394	<.001	**
3	event	1 0	110 209	17 14	14 32	141 255	5.713	.057	+
4	image	1 0	44 43	11 8	8 8	63 59	.354	.838	ns
5	level	1 0	129 8	133 19	162 23	424 50	4.529	.104	ns
6	merit	1 0	88 62	33 43	21 61	142 166	23.608	<.001	**
7	process	1 0	11 337	56 801	57 1087	124 2225	6.025	.049	*
8	relax	1 0	46 4	7	6	59 4	1.111	.574	-
9	rule	1 0	32 16	233 81	147 160	412 257	46.022	<.001	**
10	simple	1 0	19 83	16 133	9 152	44 368	11.127	.004	**
11	smooth	1 0	32 15	11 6	8 18	51 39	10.044	.007	**
12	start	1 0	58 1501	5 242	9 285	72 2028	1.992	.369	ns
13	theme	1 0	314 3	77 29	58 112	449 144	254.440	<.001	**
14	type	1 0	146 35	91 19	137 63	374 117	11.101	.004	**

“-” indicates that more than 1 cell has expected count less than 5.

C1: Data set for the real-time study, interview-based (Ch. 8)

a. Distribution of *wain* (< *wine*) and *oopun-suru* (< *open*)

Year of Survey	Age range	Age (median)	Year of Birth (median)	No. of informant	Diffusion of <i>wain</i> (%)	Diffusion of <i>oopun-suru</i> (%)
1979	16-19	17.5	1961.5	201	64.0	28.0
1979	20-24	22.0	1957.0	171	66.0	28.0
1979	25-29	27.0	1952.0	262	56.0	34.0
1979	30-34	32.0	1947.0	309	49.0	27.0
1979	35-39	37.0	1942.0	348	40.0	26.0
1979	40-49	44.5	1934.5	557	35.0	30.0
1979	50-59	54.5	1924.5	412	25.0	32.0
1979	60-69	64.5	1914.5	213	17.0	24.0
1979	70+	74.5	1904.5	166	9.0	7.0
1996	20-29	24.5	1971.5	191	96.0	62.0
1996	30-39	34.5	1961.5	212	97.0	58.0
1996	40-49	44.5	1951.5	274	86.0	58.0
1996	50-59	54.5	1941.5	237	73.0	55.0
1996	60-69	64.5	1931.5	202	60.0	59.0
1996	70+	74.5	1921.5	135	36.0	47.0
2000	16-19	17.5	1982.5	112	94.7	40.2
2000	20-29	24.5	1975.5	269	95.5	46.1
2000	30-39	34.5	1965.5	347	92.2	39.5
2000	40-49	44.5	1955.5	388	85.3	35.4
2000	50-59	54.5	1945.5	449	70.4	32.8
2000	60+	64.5	1935.5	631	52.3	26.8

b. Like and dislike of the present use of loanwords in general

Year of Survey	Age range	Age (median)	Year of Birth (median)	No. of informant	Positive (%)			Negative (%)	
					Agree	Neither	Disagree		
1977	20-24	22.0	1955.0	619	77.6	16.5	61.1	17.8	
1977	25-29	27.0	1950.0	1069	73.7	11.8	61.9	22.0	
1977	30-34	32.0	1945.0	1064	65.3	10.2	55.1	29.7	
1977	35-39	37.0	1940.0	952	64.1	10.8	53.3	30.1	
1977	40-44	42.0	1935.0	975	60.0	9.9	50.1	34.6	
1977	45-49	47.0	1930.0	896	51.2	7.6	43.6	41.5	
1977	50-54	52.0	1925.0	721	53.8	9.6	44.2	37.7	
1977	55-59	57.0	1920.0	541	49.2	8.9	40.3	40.5	
1977	60-69	64.5	1912.5	888	41.9	6.5	35.4	46.8	
1977	70+	74.5	1902.5	445	31.7	4.5	27.2	41.1	
2000	16-19	17.5	1982.5	112	90.2	33.0	57.2	8.9	
2000	20-29	24.5	1975.5	269	84.8	20.8	64.0	13.8	
2000	30-39	34.5	1965.5	347	72.4	13.8	58.5	25.7	
2000	40-49	44.5	1955.5	388	68.6	13.4	55.2	29.9	
2000	50-59	54.5	1945.5	449	54.5	9.8	44.7	43.0	
2000	60+	64.5	1935.5	631	43.4	8.9	34.5	52.9	
2002	16-19	17.5	1984.5	112	81.3	26.8	54.5	17.9	
2002	20-29	24.5	1977.5	227	88.6	23.8	64.8	11.0	
2002	30-39	34.5	1967.5	338	78.1	20.1	58.0	21.3	
2002	40-49	44.5	1957.5	340	63.0	15.6	47.4	35.3	
2002	50-59	54.5	1947.5	462	55.2	11.7	43.5	43.5	
2002	60-69	64.5	1937.5	417	49.2	14.4	34.8	48.9	
2002	70+	74.5	1927.5	304	39.5	12.5	27.0	53.9	
2008	16-19	17.5	1990.5	80	82.5	20.0	62.5	17.5	
2008	20-29	24.5	1983.5	177	78.0	19.2	58.8	20.9	
2008	30-39	34.5	1973.5	274	71.5	15.7	55.8	28.1	
2008	40-49	44.5	1963.5	306	66.6	12.4	54.2	33.3	
2008	50-59	54.5	1953.5	353	54.1	14.7	39.4	44.5	
2008	60-69	64.5	1943.5	417	48.9	14.6	34.3	48.2	
2008	70+	74.5	1933.5	368	41.0	11.7	29.3	53.8	
2013	16-19	17.5	1995.5	74	91.9	16.2	75.7	6.8	
2013	20-29	24.5	1988.5	175	83.4	13.1	70.3	16.6	
2013	30-39	34.5	1978.5	291	79.1	12.4	66.7	20.6	
2013	40-49	44.5	1968.5	327	68.8	8.6	60.2	31.2	
2013	50-59	54.5	1958.5	323	62.5	9.6	52.9	36.8	
2013	60-69	64.5	1948.5	446	53.6	7.0	46.6	45.1	
2013	70+	74.5	1938.5	517	48.9	7.5	41.4	47.2	

c. Like and dislike of a further expansion of loanwords in future

Year of Survey	Age range	Age (median)	Year of Birth (median)	No. of informant	Positive (%)	Negative (%)
1995	16-19	17.5	1977.5	131	78.6	19.1
1995	20-29	24.5	1970.5	277	77.6	19.5
1995	30-39	34.5	1960.5	382	68.6	27.8
1995	40-49	44.5	1950.5	492	58.8	37.2
1995	50-59	54.5	1940.5	410	54.9	40.5
1995	60+	64.5	1930.5	520	36.2	54.8
2003	15-19	17.0	1986.0	189	61.4	25.4
2003	20-29	24.5	1978.5	318	57.9	33.3
2003	30-39	34.5	1968.5	447	53.5	38.0
2003	40-49	44.5	1958.5	421	37.8	53.9
2003	50-59	54.5	1948.5	627	32.1	61.9
2003	60+	64.5	1938.5	1085	22.5	70.8

C2: Data set for the real-time study, corpus-based (Ch. 8)

Variable	Year of Publication (oldest)	Year of Birth (oldest)	Age (oldest)	LW (N)	non-LW (N)	Total (N)	LW (%)
<i>keesu</i>	1986	1900	86	3	109	112	2.7
	1986	1910	76	3	120	123	2.4
	1986	1920	66	23	284	307	7.5
	1986	1930	56	31	355	386	8.0
	1986	1940	46	40	358	398	10.1
	1986	1950	36	5	75	80	6.3
	1991	1920	71	36	384	420	8.6
	1991	1930	61	71	555	626	11.3
	1991	1940	51	67	588	655	10.2
	1991	1950	41	48	324	372	12.9
	1996	1920	76	13	252	265	4.9
	1996	1930	66	65	578	643	10.1
	1996	1940	56	94	822	916	10.3
	1996	1950	46	92	669	761	12.1
	1996	1960	36	37	239	276	13.4
	2001	1920	81	10	147	157	6.4
	2001	1930	71	59	461	520	11.4
	2001	1940	61	103	552	655	15.7
	2001	1950	51	86	836	922	9.3
	2001	1960	41	86	465	551	15.6
	2001	1970	31	7	155	162	4.3
<i>sapooto-suru</i>	1986	1910	76		18	18	0.0
	1986	1920	66	1	57	58	1.7
	1986	1930	56		70	70	0.0
	1986	1940	46		101	101	0.0
	1986	1950	36		24	24	0.0
	1991	1920	71		66	66	0.0
	1991	1930	61	1	103	104	1.0
	1991	1940	51	3	123	126	2.4
	1991	1950	41	1	68	69	1.5
	1991	1960	31	1	22	23	4.4
	1996	1920	76		41	41	0.0
	1996	1930	66	3	128	131	2.3
	1996	1940	56	7	195	202	3.5
	1996	1950	46	10	108	118	8.5
	1996	1960	36	3	44	47	6.4
	2001	1920	81		34	34	0.0
	2001	1930	71	2	95	97	2.1
	2001	1940	61	8	110	118	6.8
	2001	1950	51	14	120	134	10.5
	2001	1960	41	6	73	79	7.6
	<i>shinpuru</i>	1986	1910	76		7	7
1986		1920	66		16	16	0.0
1986		1930	56	1	23	24	4.2
1986		1940	46	4	17	21	19.1
1986		1950	36	1	7	8	12.5
1991		1920	71	2	17	19	10.5
1991		1930	61		27	27	0.0
1991		1940	51	9	31	40	22.5
1991		1950	41	6	19	25	24.0
1991		1960	31	2	12	14	14.3
1996		1920	76	2	16	18	11.1
1996		1930	66	3	26	29	10.3
1996		1940	56	10	52	62	16.1
1996		1950	46	8	36	44	18.2
1996		1960	36	7	20	27	25.9
2001		1930	71	6	31	37	16.2
2001		1940	61	6	30	36	16.7
2001		1950	51	13	29	42	31.0

Variable	Year of Publication (oldest)	Year of Birth (oldest)	Age (oldest)	LW (N)	non-LW (N)	Total (N)	LW (%)
	2001	1960	41	7	27	34	20.6
	2001	1970	31	2	7	9	22.2
The LW stratum	1986	1910	76	598	26151	26749	2.2
	1986	1920	66	2280	94023	96303	2.4
	1986	1930	56	3833	112546	116379	3.3
	1986	1940	46	3531	84783	88314	4.0
	1986	1950	36	1583	31289	32872	4.8
	1991	1920	71	2477	103170	105647	2.3
	1991	1930	61	5403	164798	170201	3.2
	1991	1940	51	5734	150390	156124	3.7
	1991	1950	41	4002	91963	95965	4.2
	1991	1960	31	1716	38220	39936	4.3
	1996	1920	76	1986	85840	87826	2.3
	1996	1930	66	4993	160733	165726	3.0
	1996	1940	56	6782	191057	197839	3.4
	1996	1950	46	7245	143678	150923	4.8
	1996	1960	36	3775	80239	84014	4.5
	2001	1920	81	1178	57196	58374	2.0
	2001	1930	71	4453	127389	131842	3.4
	2001	1940	61	6184	159054	165238	3.7
	2001	1950	51	7657	158972	166629	4.6
	2001	1960	41	6137	122372	128509	4.8

Bibliography

- Aitchison, J. (2001). *Language change: Progress or decay?* (3rd ed.). Cambridge: Cambridge University Press.
- Aizawa, M. (2012). Gairaigo iikae teian to wa nan de atta ka [What was the "suggestions for paraphrasing loanwords"?]. In M. Jinnouchi, M. Tanaka & M. Aizawa (Eds.), *Gairaigo kenkyuu no shintenkai [New developments in Japanese loanword studies]* (pp. 133-147). Tokyo: Ohfu.
- Altmann, G., von Buttlar, H., Rott, W., & Strauss, U. (1983). A law of change in language. *Quantitative Linguistics, 18 (Historical Linguistics)*, 104-115.
- Arakawa, S. (1943). *Gairaigo gaisetsu [Outline of loanwords]*. Tokyo: Sanseido.
- Asahara, M., Maekawa, K., Imada, M., Yasuda, S., Konishi, H. (2014). Archiving and analysing techniques of the ultra-large-scale web-based corpus project in NINJAL, *Alexandria, 25(1-2)*, 129-148.
- Bailey, G., Wikle, T., Tillery, J., & Sand, L. (1991). The apparent time construct. *Language Variation and Change, 3(3)*, 241-264.
- BBK. (1995). *Kokugo ni kansuru yoron choosa: Heisei 7 [Opinion poll on the national language: Heisei 7]*. Tokyo: BBK.
- BBK. (1999). *Kokugo ni kansuru yoron choosa: Heisei 10 [Opinion poll on the national language: Heisei 10]*. Tokyo: BBK.
- BBK. (2000). *Kokugo ni kansuru yoron choosa: Heisei 11 [Opinion poll on the national language: Heisei 11]*. Tokyo: BBK.
- BBK. (2003). *Kokugo ni kansuru yoron choosa: Heisei 14 [Opinion poll on the national language: Heisei 14]*. Tokyo: BBK.

- BBK. (2008). *Kokugo ni kansuru yoron choosa: Heisei 19 [Opinion poll on the national language: Heisei 19]*. Tokyo: BBK.
- BBK. (2009). *Kokugo ni kansuru yoron choosa: Heisei 20 [Opinion poll on the national language: Heisei 20]*. Tokyo: BBK.
- BBK. (2013). *Kokugo ni kansuru yoron choosa: Heisei 24 [Opinion poll on the national language: Heisei 24]*. Tokyo: BBK.
- BBK. (2014). *Kokugo ni kansuru yoron choosa: Heisei 25 [Opinion poll on the national language: Heisei 25]*. Tokyo: BBK.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145-204.
- Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50(2), 333-355.
- Chambers, J. K. (1990). The Canada-US border as a vanishing isogloss: The evidence of *chesterfield*. *Journal of English Linguistics*, 23(1-2), 155-166.
- Chambers, J. K., & Trudgill, P. (1980). *Dialectology*. Cambridge: Cambridge University Press.
- D'Arcy, A. (2006). Lexical replacement and the like(s). *American Speech*, 81(4), 339-357.
- Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society*, 17(2), 183-207.
- Hashimoto, W. (2010). *Gendai nihongo ni okeru gairaigo no ryooteki-suii ni kansuru kenkyuu [The increasing process of loanwords in modern Japanese]*. Tokyo: Hituzi Shobo.

- Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the world's languages: A comparative handbook* (pp. 35-54). Berlin: Walter de Gruyter.
- Haugen, E. (1953). *The Norwegian language in America. vol. 2*. Philadelphia: University of Pennsylvania Press.
- Hibiya, J. (1988a). A quantitative study of Tokyo Japanese. Doctoral Dissertation, University of Pennsylvania.
- Hibiya, J. (1988b). Variations riron [Variationist theory]. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]*, 93, 155-171.
- Hibiya, J. (1995). The velar nasal in Tokyo Japanese: A case of diffusion from above. *Language Variation and Change*, 7(2), 139-152.
- Inoue, F. (2000). *Tohoku hoogen no hensen [Change of Tohoku dialects]*. Tokyo: Akiyama Shoten.
- Irie, S. (2010). "Chūo-Kōron 101 nen no goi [Research on vocabulary of "Chūo-kōron" in the last 101 years]. *Dodai Goi Kenkyuu [Doshisha Lexicology]*, (13), 9-14.
- Irwin, M. (2011). *Loanwords in Japanese*. Amsterdam: John Benjamins Publishing.
- Ishii, H. (1990). "Chūo-kōron" 1986 nen no yoogo [Words used in the magazine "Chūo-kōron" in 1986]. In NINJAL (Ed.), *Kenkyuu Hookoku Shuu [Occasional Papers 11]* (pp. 1-40). Tokyo: NINJAL.
- Ishino, H. (1983). *Gendai gairaigo koo [Discussions on contemporary loanwords]*. Tokyo: Taishukan Shoten.

- Ishiwata, T. (1965). Gairaigo no fukyuudo: Matsue-shi de no choosa kara. *Gengo Seikatsu*, 161, 55-62.
- Ishiwata, T. (2001). *Gairaigo no soogooteki kenkyuu [Comprehensive study of loanwords]*. Tokyo: Tokyodo Shuppan.
- ISM & NINJAL. (2014). *Dai 4 kai Tsuruoka-shi ni okeru gengo choosa vol.1: Onsei/On'in hen [The 4th linguistic survey in Tsuruoka City, Japan vol.1: Phonetics & phonology]*. Tokyo: ISM & NINJAL.
- ISM & NINJAL. (2015). *Dai 4 kai Tsuruoka-shi ni okeru gengo choosa vol.2: Goi/Bunpoo hen [The 4th linguistic survey in Tsuruoka City, Japan vol.2: Vocabulary & grammar]*. Tokyo: ISM & NINJAL.
- Jinnouchi, M. (1993). *Saji to supuun: Gairaigo to meimei no yure [Saji or supuun: A study on variation in denomination caused by loan words]*. *Gengo Bunka Ronkyu*, 4, 47-54.
- Jinnouchi, M. (2007). *Gairaigo no shakaigengogaku: Nihongo no gurookaru na kangae kata [Sociolinguistic studies of Japanese foreign words: From the global point of view]*. Kyoto: Sekaishiso-sha.
- Kageyama, T., & Saito, M. (2016). Handbook of Japanese lexicon and word formation. In T. Kageyama, & H. Kishimoto (Eds.), (pp. 11-50). Berlin: Walter de Gruyter.
- Kajiki, M., Iwaya, A., Kashiwakura, Y., & Asamatsu, A. (1995). Kokka ni yoru gairaigo kisei [State control over the use of words of foreign origin]. In NHK (Ed.), *NHK hoosoo bunka choosa kenkyuu nenpoo 40 [The NHK annual bulletin of broadcasting culture research 40]* (pp. 191-249). Tokyo: NHK.

- Kim, E. (2011). *Nijusseiki koohan no shinbun goi ni okeru gairaigo no kihongoka* [Shift of the loanwords to basic words in the Japanese newspaper vocabulary in the second half of the 20th century]. Osaka: Osaka University.
- Kim, E. (2013). Gairaigo doomeishi “*chekku*” no kihongoka [Shift of the loan verbal noun “*chekku*” to basic words]. In M. Aizawa (Ed.), *Gendai nihongo no dootai kenkyuu* [Study on dynamics of contemporary Japanese] (pp. 29-45). Tokyo: Ohfu.
- Koiso, H., Tanaka, Y., Watanabe, R., & Den, Y. (2016a). A large-scale corpus of everyday Japanese conversation: On methodology for recording naturally occurring conversations. *Proceedings of LREC 2016 Workshop on Casual Talk among Humans and Machines*, 10, 9-12. Retrieved from http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-Just%20Talking_Proceedings.pdf (on 10/Augst/2016)
- Koiso, H., Tsuchida, T., Watanabe, R., Yokomori, D., Aizawa, M., & Den, Y. (2016b). Kinkoo kaiwa koopasu sekkei no tame no ichinichi no kaiwa koodoo ni kansuru kiso choosa [Survey of conversational behavior: Towards the design of a balanced corpus of conversational Japanese]. *Kokuritsu Kokugo Kenkyuujo Ronshuu* [NINJAL Research Papers], 10, 85-106.
- Kunihiro, T., Inoue, F., & Long, D. (Eds.). (1999). *Takeshi Shibata: Sociolinguistics in Japanese contexts*. Berlin: Mouton de Gruyter.
- Kuya, A. (2013). Gendai nihongo ni okeru gairaigo no shintoodo: Gairaigo shiyoo ni kansuru ishiki to koodoo no shakaigengogaku-teki koosatsu [Loanword diffusion in contemporary Japanese: A sociolinguistic discussion on linguistic

- attitude toward, and behavior on, the use of loanwords]. *Shakaigengokagakukai Dai 32 Kai Taikai Happyoo Ronbunshuu [Proceedings of the 32nd Meeting of the Japanese Association of Sociolinguistic Science (JASS 32)]*, 136-139.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3), 273-309.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1978). Where does the linguistic variable stop?: A response to Beatriz Lavandera. *Working Papers in Sociolinguistics*, (44), 1-17.
- Labov, W. (1982). Building on empirical foundations. *Perspectives on Historical Linguistics*, 24, 17-92.
- Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh, & J. Sherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28-53). Englewood Cliffs, NJ: Prentice-Hall.
- Labov, W. (1994). *Principles of linguistic change (vol.1): Internal factors*. Oxford: Blackwell.
- Labov, W. (2001a). *Principles of linguistic change (vol.2): Social factors*. Oxford: Blackwell.
- Labov, W. (2001b). The anatomy of style-shifting. In P. Eckert, & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 85-108). Cambridge: Cambridge University Press.

- Labov, W. (2004). Quantitative analysis of linguistic variation. In U. Ammon, N. Dittmar & K. J. Mattheier (Eds.), *Sociolinguistics: An international handbook of the science of language and society (vol.1)* (pp. 6-21). Berlin: De Gruyter Mouton.
- Lavandera, B. (1978). Where does the sociolinguistic variable stop? *Language in Society*, 7, 171-183.
- Loveday, L. J. (1996). *Language contact in Japan: A socio-linguistic history*. Oxford: Oxford University Press.
- Maeda, T. (2005). Yuusoo choosa no tokuchoo ni kansuru ichi kenkyuu: Mensetsu choosa hoo to no hikaku o chuushin to shite [A study on the characteristics of mail survey: A comparison with face-to-face interviewing], *Tookei Suuri [Proceedings of the Institute of Statistical Mathematics]*, 53(1), 57-81.
- Maekawa, K. (2003). Corpus of spontaneous Japanese: Its design and evaluation. *Proceedings ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR 2003)*, 7-12.
- Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W. (2014). Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation*, 48(2), 345-371.
- Maruyama, T. (2012). Dai kibo koopasu no riyoo to meta deeta no yakuwari [The role of metadata in the analysis of large-scale corpora]. *Dai 1 Kai Koopasu Nihongogaku Waakushoppu Yokooshuu [Proceedings of the 1st Workshop on Corpus-Based Japanese Linguistics]*, 203-210.

- Matsuda, K. (1993). Dissecting analogical leveling quantitatively: The case of the innovative potential suffix in Tokyo Japanese. *Language Variation and Change*, 5(1), 1-34.
- Matsuda, K. (1995). Variable zero-marking of (o) in Tokyo Japanese. Doctoral Dissertation, University of Pennsylvania.
- Matsuda, K. (2013). Sahen dooshi no godan katsuyooka, kamiichidan katsuyooka no henka [Change in conjugation of *sahen* verbs – shift to *godan* or *kamiichidan* conjugation pattern]. In M. Aizawa (Ed.), *Gendai nihongo no dootai kenkyuu [Study on dynamics of contemporary Japanese]* (pp. 69-89). Tokyo: Ohfu.
- Meyerhoff, M. (2011). *Introducing sociolinguistics* (2nd ed.). Oxford: Routledge.
- Miyata, K. (2007). Gairaigo “meritto” to sono ruigigo no imi hikaku [Comparison in meaning between the loanword “merit” and its synonyms]. In NINJAL (Ed.), *Kookyoo baitai no gairaigo: “gairaigo iikae teian” o sasaeru choosa kenkyuu [Loanwords in the public media: basic researches for “suggestions for paraphrasing loanwords”]* (pp. 402-409). Tokyo: NINJAL.
- Miyata, K., & Tanaka, M. (2006). Gairaigo “risuku” to sono ruigigo no imi hikaku [Comparison in meaning between the loanword “risk” and its synonyms]. *Gengo Shori Gakkai Dai 12 Kai Nenji Taikai Happyoo Ronbunshuu [Proceedings NLP 2006]*, pp. 600-603.
- Mogi, T. (2011). Koopasu o mochiita gairaigo sahen dooshi no bunseki: “katto-suru” o rei to shite [A corpus-based study on loanword verbs in Japanese: A case study of *katto-suru* (< *cut*)]. *Tokutei Ryooiki Kenkyuu*

- “Nihongo Koopasu” Heisei 22 Nendo Kookai Waakushoppu Yokooshuu, 103-110.
- Mogi, T. (2015). Koopasu o mochiita gairaigo sahen dooshi no bunseki: “*maaku-suru*” o rei to shite [A corpus-based study on loanword verbs in Japanese: A case study of *maaku-suru* (< *mark*)]. *Kumamoto Daigaku Bungakubu Ronsoo*, 106, 83-95.
- Mori, H. (2013). Gendai nihongo kakiketoba kinkoo koopasu “toshokan shoseki” no seinendai betsu bunpu wa nani o imi shite iru no ka: “*denai*”, “*dewanai*” and “*janai*” no shiyoo wariai kara mita ichi koosatsu [What does the birth year distribution in the library sub-corpus in the BCCWJ represent; A study based on the percentage of “*denai*”, “*dewanai*” and “*janai*”]. *Dai 4 Kai Koopasu Nihongogaku Waakushoppu Yokooshuu [Proceedings of the 4th Workshop on Corpus-Based Japanese Linguistics]*, 275-284.
- Mougeon, R., & Beniak, E. (1991). *Linguistic consequences of language contact and restriction: The case of French in Ontario, Canada*. Oxford: Oxford University Press.
- Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Oxford University Press.
- Myers-Scotton, C. (2006). *Multiple voices: An introduction to bilingualism*. Oxford: Wiley-Blackwell.
- Nambu, S. (2007). Teiryooteki bunseki ni motozuku “*ga/no*” kootai saikoo [Reconsideration of Ga/No conversion based on a quantitative analysis]. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]*, 131, 115-149.

- Nambu, S. (2014). *Koopasu gengogaku oyobi jikken gengogaku ni motozuku kakujoshi kootai no bunseki* [On the use of case particles in Japanese: Corpus and experimental studies]. Doctoral Dissertation, Osaka University.
- NHK. (1980). *NHK bunken geppoo* [Monthly bulletin of the radio & television culture research institute], No. 30-2. Tokyo: NHK.
- NHK. (1986). *Hoosoo kenkyuu to choosa* [The NHK monthly report on broadcast research], No. 36-7. Tokyo: NHK.
- NHK. (1991). *Hoosoo kenkyuu to choosa* [The NHK monthly report on broadcast research], No. 41-8. Tokyo: NHK.
- NHK. (1995). *Hoosoo kenkyuu to choosa* [The NHK monthly report on broadcast research], No. 45-9. Tokyo: NHK.
- NHK. (1996). *Hoosoo kenkyuu to choosa* [The NHK monthly report on broadcast research], No. 46-9. Tokyo: NHK.
- NINJAL. (1957). *Keigo to keigo ishiki* [Socio-psychological survey on Japanese polite expression]. Tokyo: Shuei Shuppan.
- NINJAL. (1962). *Gendai zasshi 90 shu no yoogo yooji I* [Vocabulary and Chinese characters in ninety magazines of today I]. Tokyo: NINJAL.
- NINJAL. (1963). *Gendai zasshi 90 shu no yoogo yooji II* [Vocabulary and Chinese characters in ninety magazines of today II]. Tokyo: NINJAL.
- NINJAL. (1964). *Gendai zasshi 90 shu no yoogo yooji III* [Vocabulary and Chinese characters in ninety magazines of today III]. Tokyo: NINJAL.
- NINJAL. (1983). *Keigo to keigo ishiki: Okazaki ni okeru 20 nen mae to no hikaku* [Socio-psychological survey on Japanese polite expression: After 20

- years from the preceding survey in Okazaki City, Aichi Pref.]. Tokyo: Sanseido.
- NINJAL. (1987). *Zasshi yoogo no hensen [Changes in the language of a magazine]*. Tokyo: NINJAL.
- NINJAL. (1995). *Terebi hoosoo no goi choosa 1 [Vocabulary survey of television broadcasts vol. 1]*. Tokyo: Shuei Shuppan.
- NINJAL. (1997). *Terebi hoosoo no goi choosa 2 [Vocabulary survey of television broadcasts vol. 2]*. Dainippon Tosho.
- NINJAL. (1999). *Terebi hoosoo no goi choosa 3 [Vocabulary survey of television broadcasts vol. 3]*. Dainippon Tosho.
- NINJAL. (2004). *Gairaigo ni kansuru ishiki choosa: Zenkoku choosa [The national survey on attitudes to loanwords]*. Tokyo: NINJAL.
- NINJAL. (2005a). *Gairaigo ni kansuru zenkoku ishiki choosa II [The national survey on attitudes to loanwords II]*. Tokyo: NINJAL.
- NINJAL. (2005b). *Gendai zasshi no goi choosa: 1994 nen hakkoo 70 shi [A survey of vocabulary in contemporary magazines (1994)]*. Tokyo: NINJAL.
- NINJAL (2006). *Nihongo hanashikotoba koopasu no koochikuhoo [Construction of the Corpus of Spontaneous Japanese]*. Tokyo: NINJAL.
- NINJAL. (2007a). *Chiiki shakai no gengo seikatsu [Social aspects of language in a local community]*. Tokyo: NINJAL.
- NINJAL. (2007b). *Koogyoo baitai no gairaigo: "Gairaigo iikae teian" o sasaeru choosa kenkyuu [Loanwords in the public media: Basic researches for "suggestions for paraphrasing loanwords"]*. Tokyo: NINJAL.

- NINJAL. (2010a). *Keigo to keigo ishiki: Aicikken Okazaki-shi ni okeru dai 3 ji choosa [Socio-psychological survey on Japanese polite expression: The 3rd survey in Okazaki City, Aichi Pref.]* No. 1. Tokyo: NINJAL.
- NINJAL. (2010b). *Keigo to keigo ishiki: Aicikken Okazaki-shi ni okeru dai 3 ji choosa [Socio-psychological survey on Japanese polite expression: The 3rd survey in Okazaki City, Aichi Pref.]* No. 2. Tokyo: NINJAL.
- NINJAL. (2010c). *Keigo to keigo ishiki: Aicikken Okazaki-shi ni okeru dai 3 ji choosa [Socio-psychological survey on Japanese polite expression: The 3rd survey in Okazaki City, Aichi Pref.]* No. 3. Tokyo: NINJAL.
- NINJAL. (2010d). *Keigo to keigo ishiki: Aicikken Okazaki-shi ni okeru dai 3 ji choosa [Socio-psychological survey on Japanese polite expression: The 3rd survey in Okazaki City, Aichi Pref.]* No. 4. Tokyo: NINJAL.
- NINJAL. (2011). *BCCWJ riyoo no tebiki 1.0 [BCCWJ user's manual: Version 1.0]*. Tokyo: NINJAL.
- NSDKK. (1977). *Kokugo ni kansuru yoron choosa: Showa 52 [Opinion poll on the national language: Showa 52]*. Tokyo: NSDKK.
- Pope, J., Meyerhoff, M., & Ladd, D. R. (2007). Forty years of language change on Martha's Vineyard. *Language*, 83(3), 615-627.
- Rickford, J. R., & McNair-Knox, F. (1994). Addressee-and topic-influenced style shift: A quantitative sociolinguistic study. In D. Biber, & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 235-276). Oxford: Oxford University Press.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

- Sanada, H. (2002). *Kindai nihongo ni okeru gakujutsu yoogo no seiritsu to teichaku [A quantitative study of the popularization of Meiji era scholarly terms]*. Tokyo: Junbunsha.
- Sanada, H. (2008). Gengo henka no S-ji kaabu: Kaiseki shuhoo no hikaku to sono tekiyoo jirei [S-curve model of language change: Methods and examples]. *Saitama Gakuen Daigaku Kiyoo: Ningen Gakubu-Hen [Bulletin of Saitama Gakuen University: Faculty of Humanities]*, 8, 1-11.
- Sankoff, D. (1982). Sociolinguistic method and linguistic theory. *Logic, Methodology, Philosophy of Science VI: Proceedings of the Sixth International Congress of Logic, Methodology and Philosophy of Science*, 677-689.
- Sankoff, D., & Laberge, S. (1978). The linguistic market and the statistical explanation of variability. In D. Sankoff (Ed.), *Linguistic variation: Models and methods* (pp. 239-250). New York: Academic Press.
- Sankoff, G. (1973). Above and beyond phonology in variable rules. In C. N. Bailey, & R. W. Shuy (Eds.), *New ways of analyzing variation in English* (pp. 44-62). Washington: Georgetown University Press.
- Sankoff, G. (2005). Cross-sectional and longitudinal studies in sociolinguistics. In N. Dittmar, K. J. Mattheier & U. Ammon (Eds.), *Sociolinguistics: An international handbook of the science of language and society, vol. 2* (pp. 1003-1013). Berlin: Mouton de Gruyter.
- Sankoff, G. (2006). Age: Apparent time and real time. In K. Brown (Ed.), *Encyclopedia of language and linguistics vol.1* (2nd ed., pp. 110-116). Amsterdam: Elsevier.

- Sankoff, G., & Thibault, P. (1980). The alternation between the auxiliaries *avoir* and *être* in Montreal French. In G. Sankoff (Ed.), *The social life of language* (pp. 311-345). Philadelphia: University of Pennsylvania Press.
- Sano, S. (2008). Nihongo hanashikotoba koopasu ni arawareru “*sa-ire kotoba*” ni kansuru suuryooteki bunseki [A quantitative analysis of *sa*-insertion in Corpus of Spontaneous Japanese]. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]*, 133, 77-106.
- Sano, S. (2009). The roles of internal and external factors and the mechanism of analogical leveling: Variationist-and probabilistic OT approach to ongoing language change in Japanese voice system. Doctoral Dissertation, Sophia University.
- Sano, S. (2011). Real-time demonstration of the interaction among internal and external factors in language change: A corpus study. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]*, 139, 1-27.
- Sano, S., & Kawahara, S. (2013). A corpus-based study of geminate devoicing in Japanese: The role of the OCP and external factors. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]*, 144, 103-118.
- Schilling-Estes, N. (2002). Investigating stylistic variation. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 375-401). Oxford: Wiley-Blackwell.
- Schwenter, S. A., & Torres Cacoullos, R. (2010). Grammaticalization paths as variable contexts in weak complementarity in Spanish. In J. A. Walker (Ed.), *Aspect in grammatical variation* (pp. 13-26). Amsterdam: John Benjamins.

- Shibata, T. (1978). *Shakai gengogaku no kadai [Topics for sociolinguistics]*. Tokyo: Sanseido.
- Shimooka, K. (2013). “Kutsurogu” and “rirakkusu-suru” [The native Japanese “relax” and the loanword “relax”]. *Nihon Gengo Bunka Kenkyuu*, 17, 10-24.
- Sohn, H. (1999). *The Korean language*. Cambridge: Cambridge University Press.
- Soomushoo Tookeikyoku (Statistics Bureau, Ministry of Internal Affairs and Communications, Japan). *Heisei 12 nen kokusei choosa no shuyoo kekka [Results of the concensus in heisei 12]* Soomushoo Tookeikyoku. Retrieved from http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000000030587&cycleCode=0&requestSender=search (on 29/March/2016)
- Tagliamonte, S. (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Takano, S. (2012). Kotoba no sutairu o rikai shi ooyoo suru [Understanding and making use of language style]. In J. Hibiya (Ed.), *Hajimete manabu shakai gengogaku [Introduction to sociolinguistics]* (pp. 248-269). Kyoto: Minerva Shobo.
- Tanaka, M. (2007). Kango, wago to hikaku shita gairaigo ni taisuru ishiki [The public attitude to loanwords compared with that to Sino-Japanese words and native-Japanese words]. In NINJAL (Ed.), *Kookyoo baitai no gairaigo: “gairaigo iikae teian” o sasaeru choosa kenkyuu [Loanwords in the public media: basic researches for “suggestions for paraphrasing loanwords”]* (pp. 302-310). Tokyo: NINJAL.

- Tanaka, M. (2012). Kokugo kyooiku ni okeru gairaigo: Koopasu ni yoru ruikeika o tooshite [Loanwords in Japanese language teaching: A corpus-based typological approach]. In M. Jinnouchi, M. Tanaka & M. Aizawa (Eds.), *Gairaigo kenkyuu no sintenkai [New developments in Japanese loanword studies]* (pp. 224-242). Tokyo: Ohfu.
- Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Watanabe, R., Tanaka, Y., & Koiso, H. (2015). Nihongo hanashikotoba koopasu UniDic ban keitairon joofoo no koochiku [Constructing the UniDic version of the morphological information of *Corpus of Spontaneous Japanese*]. *Dai 8 Kai Koopasu Nihongogaku Waakushoppu Yokooshuu [Proceedings of the 8th Workshop on Corpus-Based Japanese Linguistics]*, 279-288.
- Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics*, 19(1), 29-58.
- Weinreich, U. (1968). *Languages in contact: Findings and problems* (6th ed.). The Hague: Mouton.
- Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann, & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 95-195). Austin: University of Texas Press.
- Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103-129.

- Yamaguchi, M., Takada, T., Kitamura, M., Mabuchi, Y., Oshima, H., Kobayashi, M., et al. (2011). “*Gendai nihongo kakikotoba kinkoo koopasu*” ni okeru denshika foomatto ver. 2.2 [*Electronization of data in BCCWJ ver. 2.2*] (No. JC-D-10-04). Tokyo: NINJAL.
- Yang, M. (2005). Nikkan daigakusei no ankeeto kara mita keiyooshi-kei gairaigo no juyoo ishiki [A study on the awareness of foreign origin adjectives through the survey on Japanese and Korean university students]. *Gengo Kagaku Ronshuu*, 9, 83-93.
- Yang, M. (2007). Gairaigo o meguru ishiki ni kansuru nikkans taishoo kenkyuu [A contrastive study between Japanese and Korean concerning consciousness of loan words]. *Kokugogaku Kenkyuu*, 46, 73-85.
- Yang, M. (2012). Nihongo to kankokugo no gairaigo no juyoo ishiki: Imeeji choosa no bunseki [Receptiveness to loanwords in Japanese and Korean: An analysis of an impression survey on loanwords]. In M. Jinnouchi, M. Tanaka & M. Aizawa (Eds.), *Gairaigo kenkyuu no shintenkai [New developments in Japanese loanword studies]* (pp. 148-167). Tokyo: Ohfu.
- Yokoyama, S. (2006). Itaiji senkoo ni okeru tanjun sesshoku kooka to ippan taioo hoosoku no kankei [Mere exposure effect and generalized matching law for preference of *kanji* form]. *Keiryoo Kokugogaku [Mathematical Linguistics]*, 25(5), 199-214.
- Yokoyama, S., Asahi, Y., & Sanada, H. (2008). Kioku moderu ni yoru keigo ishiki no henka yosoku [New method of statistical analysis for longitudinal survey

data on honorifics use by psychological model]. *Shakai Gengo Kagaku*, 11(1), 64-75.

Yokoyama, S., & Sanada, H. (2007). Tahenryoo S-ji kaabu ni yoru gengo henka no kaiseki [Multiple logistic regression analysis for formulating a change in language]. *Keiryoo Kokugogaku [Mathematical Linguistics]*, 26(3), 79-93.

Yokoyama, S., & Sanada, H. (2010). Gengo no shoogai shuutoku moderu ni yoru kyootsuugoka yosoku [Predictions of dialect standardization by the “life-long assimilation of language change” model]. *Nihongo no Kenkyuu*, 6(2), 31-45.

Yokoyama, S., & Wada, Y. (2006). A logistic regression model of variant preference in Japanese *kanji*: An integration of mere exposure effect and the generalized matching law. *Glottometrics*, 12, 63-74.

Dictionaries & Corpora

- Horiuchi, K. (Ed.). (2011). *Gendai Yoogo-no Kisochishiki Katakana Gairaigo/Ruigo Jiten* (4th ed.). Tokyo: Jiyu Kokumin Sha.
- Kenbo, H., Ichikawa, T., Hida, Y., Yamazaki, M., Iima, H., & Shioda, T. (Eds.). (2014). *Sanseido Kokugo Jiten* (7th ed.). Tokyo: Sanseido.
- Masuda, K. (Ed.). (1974). *Kenkyusha's New Japanese-English Dictionary* (4th ed.). Tokyo: Kenkyusha.
- Nihon Kokugo Daijiten dai 2 han henshuu iinkai, & Shogakukan kokugo jiten henshubu (Eds.). (2000-2). *Nihon Kokugo Daijiten* (2nd ed.). Tokyo: Shogakukan.
- NINJAL. (2011). *Balanced Corpus of Contemporary Written Japanese (BCCWJ)*. Tokyo: NINJAL.
- NINJAL & NICT. (2004). *Corpus of Spontaneous Japanese (CSJ)* NINJAL;NICT.
- Nishio, M., Iwabuchi, E., & Mizutani, S. (Eds.). (2009). *Iwanami kokugo jiten* (7th ed.). Tokyo: Iwanami Shoten.
- Sanseido Henshujo (Ed.). (2010). *Concise Katakana Jiten [Sanseido's concise dictionary of katakana words] (web)* (4th ed.). Tokyo: Sanseido. Retrieved from <http://www.sanseido.net/> (on 4-16/May/2015 and 7-17/June/2016)
- Shinmura, I. (Ed.). (2008). *Kojien* (6th ed.). Tokyo: Iwanami Shoten.
- Shinozaki, K., Aizawa, M., Oshima, M., & Hayashi, S. (Eds.). (2012). *Reikai Shin Kokugo Jiten* (8th ed.). Tokyo: Sanseido.
- Yamada, T., Shibata, T., Sakai, K., Kuramochi, Y., Yamada, A., Uwano, Z., et al (Eds.). (2011). *Shinmeikai Kokugo Jiten (web)* (7th ed.). Tokyo: Sanseido.

Retrieved from <http://www.sanseido.net/> (on 4-16/May/2015 and 7-17/June/2016)