The Japanese Mental Lexicon:

The Lexical Retrieval and Representation of Two-Kanji Compound Words from a Morphological Perspective

Terry Joyce

975058

Thesis Submitted to the

Institute of Psychology, University of Tsukuba

in Part Fulfillment of Requirements for the Degree of

Doctor of Philosophy (Psychology)

2001 Academic Year

At its most general level, this research takes a psycholinguistic approach in seeking an acceptable answer to the question *how do kanji function as a writing system?* In pursuing this answer, the thesis takes the two-kanji compound word as its principle frame of reference. The emphasis on the two-kanji compound word, rather than on the single kanji character, follows from a belief that to understand how a writing system functions it is essential to identify the principles that govern the combination of graphic units.

In searching for an answer to the research question, this thesis explores and traces the connections between four broad areas; namely, the classification of the Japanese writing system, Japanese word formation, visual word recognition, and, in particular, the Japanese mental lexicon. Given that the nature of a writing system will be reflected in the structure of the literate mental lexicon, the transition from writing system to mental lexicon is simply a shift in perspective—from external form to internal representation. The answer that emerges from the exploration of these areas is that *kanji function as a morphographic writing system*, and, moreover, this is reflected in the Japanese mental lexicon.

Part 1 of this thesis is linguistic in nature, focusing primarily on the external form. Based on discussion concerning the appropriate classification of the Japanese writing system, and, especially of kanji, within a linguistic taxonomy of writing systems, it is argued that kanji may be most accurately characterized as a morphographic writing system. Evidence for this is provided by looking at the word formation processes that underlie two-kanji compound words, which are primarily morphological. Compounding is the combining of abstract linguistic elements—morphemes—to form larger word units, with the representation of these larger word units being a product of combining the constituent representations. The discussion of two-kanji compound word morphology also provides background for the experiments presented in this thesis.

Part 2 of the thesis is psycholinguistic in nature, focusing primarily on the internal representation. After arguing in Part 1 that kanji function as a morphographic writing system, Part 2 investigates the implications of this for the organization of the Japanese mental lexicon. Specifically, a lemma unit model version (Joyce, 1999, in press) of the multilevel interactive-activation framework is proposed as a model for the Japanese mental lexicon. The major feature of the lemma unit model is the incorporation of lemma units that mediate the links between access representations and semantic units. This model is examined through a series of constituent-morpheme priming experiments, which investigated the lexical retrieval and representation of two-kanji compound words within the Japanese mental lexicon from a morphological perspective by controlling for the word-formation principle of compound word stimuli as an experimental condition.

As background to the conducted experiments, important tasks and effects are discussed in a brief review of visual word recognition research, which focuses on the two central issues for the mental lexicon—lexical retrieval and lexical representation. A short review of models of lexical representation, which all allow for some form of morphological involvement in the lexical retrieval of polymorphemic words, also highlights the language-universal importance of morphology for the organization of the mental lexicon.

As any model of the Japanese mental lexicon must in some way capture the morphological relations that exist between polymorphemic words, two proposals for models of the Japanese mental lexicon—Hirose's (1982, 1994, 1996) hypotheses and Joyce's (1999, in press) suggestion for a Japanese lemma unit model—are investigated in terms of their ability to cope with the diversity in the morphology of two-kanji compound words. Because these proposals make different predictions concerning constituent-morpheme priming, the patterns of facilitation are examined in two experiments varying stimulus onset synchronicity (SOA) with five word-formation principles as experimental conditions. The results from these experiments, of similar levels of priming for both constituents in all but one word-formation condition, are more consistent with the predictions from the Japanese lemma-unit model.

The lemma unit model is investigated further through a series of seven experiments that seek to examine various aspects of this model of the Japanese mental lexicon. While these experiments retain a focus on the lexical retrieval and representation of two-kanji compound words from a morphological perspective, they also investigate the extent to which the results in the first two experiments might be task specific, the possibility that the lemma unit model offers to integrate both kana and kanji processing within a single model, and the potential to model the dual reading system of on-readings and kun-readings. In addition, by employing recently compiled constituent-morpheme frequency data (Joyce & Ohta, in press), the issue of positional sensitivity is examined for verbal constituent morphemes of two-kanji compounds words. Although the results from this series of experiments raise further questions, they are generally consistent with the lemma unit model, which is undoubtedly an extremely appealing way of thinking about the relations that exist between semantic representations, and access representations for orthography and phonology, as well as accounting for the morphological information that underlies two-kanji compound words in the Japanese mental lexicon.

This investigation of the lexical retrieval and representation of two-kanji compound words within the Japanese mental lexicon from a morphological perspective has serious implications for our understanding of kanji, as well as for our perception of writing systems and of language itself.

Acknowledgements

I would like to thank many of the wonderful people I have been privileged to meet and know in the course of my studies. Without their kind help, in various ways, it simply would not have been possible for me to undertake this research and to write this thesis. For failing to make better use of all this kind help, much to the detriment of this thesis, I can only offer my sincere apologizes.

I shall be eternally indebted to Professor Ohta Nobuo, who has supervised and supported my research over the past five years with so much patience, kindness, and consideration.

I shall also always remember the advice, support, kindness, and friendship of all associated with Professor Ohta's laboratory during that time. The *visiting foreign lecturers*: Firstly, Dr. Bradford Challis, for advice on the preparation of my first experiments, and, later, Dr. Ulrich Olofsson, for help with analyzing those initial results. The *post-doctoral research fellow*: Dr. Eleanor Olds Batchelder for beneficial 'prodding' and discussion as I struggled to write my master's thesis. The *senpai*: Dr. Terasawa Takafumi, Dr. Kawano Rie, and Yoshida Tetsuya for the benefit of their experience and kindness. The *dōkyūsei*: Gotoh Fumiko, Hayashi Mitsuko, and Hoshino Satoshi for support throughout.

Beyond Professor Ohta's laboratory, I have also benefited greatly from the advice, help, and kindness from many at the Institute of Psychology during this time, and I would like to thank everyone for that. In particular, I would like to mention the following people: Professor Kaiho Hiroyuki for his kind concern and encouragement. Professor Kikuchi Tadashi for the loan of equipment used in conducted naming experiments. Dr. Aoyama Masahiko, for advice on the preparation of a number of experiments, as well as helpful comments on a draft manuscript. Dr. Nanbu Misako, for her generous time and help with so many of the conducted experiments. Dr. Yugawa Shintaro, for various help. Dr. Kobayashi Takefumi, for help with a translation. Shimada Hideaki for helpful discussion of ratios. Nakajima Masato for advice on waveform data. My special thanks also to Yagi Yoshihiko for his technical knowledge in making the naming experiments possible. And, to Simon Downes for his friendship and for sharing the experience of being a foreign graduate student in the department.

Beyond the Institute of Psychology, I have received help from many people in other departments at the University of Tsukuba. I would like to express my thanks to Dr. Stefan Kaiser for his advice and help as supervisor during my first two years in Tsukuba and since then, and to all the Japanese language teachers at the Foreign Student Center. For allowing me to recruit participants for my experiments at their classes, I am grateful to the following lecturers: in the Area Studies program, Dr. Philip Hansen, and at the Foreign Language Center, Andrew Barfield, Dr. Joe Johnson, Alan Milne, Sakuma Yasuyuki, and Garth Warries. In addition to their help with recruiting, I also wish to thank Dr. Yvonne Stapp for all her kindness and words of encouragement, and at the Institute of Literature and Linguistics, Dr. John Shillaw for much help and advice, especially on computing matters, and Dr. Scott Saft for helpful comments on a couple of draft chapters. For his technical expertise in fabricating the luminance sensor used in the naming experiments, I would like to thank Toda Hideki at the Institute of Engineering Mechanics and Systems. I am also grateful to all the many undergraduate and post-graduate students at the University of Tsukuba who either completed various survey questionnaires related to the experiments or participated in the experiments themselves.

Then, beyond the University of Tsukuba, I would like to thank a number of people for kindnesses at various stages along the way. For the encouragement to start post-graduate research in the first place, I would like to express my appreciation to lecturers and good friends at the School of East Asian Studies, University of Sheffield, England at that time, including Graham Healey, Dr. Miriam Jelinek, Roger Lancefield, Katia Lancelotti, Sue Preston and Dr. Nicolas Tranter. I wish to thank to Dr. Jinmin Wang for his advice while at the University of Library and Information Science, Tsukuba. I am extremely grateful to Dr. Marcus Taft, University of New South Wales, for his time and insightful comments. I appreciate also the advice and kindness of Dr. Saito Hirofumi, Nagoya University, as well as that of Dr. Masuda Hisashi, Nagoya University.

Finally, on a personal note, I would like to express my special thanks to my wife, Sachiko, and to our three sons, Kazunari, Yoshinari, and Tomonari, for all their support and for so much happiness.

This research was supported by a scholarship from the Japanese Ministry of Education, Culture, Sports, Science and Technology, for which I am truly grateful.

Table of Contents

Abstract ii

Acknowledgments vi

Table of contents ix

List of tables xix

List of figures xxii

Part 1
How do kanji function as a writing system?

Chapter 1 2
Introduction:
From external form to internal representation

1.1 The Japanese writing system 4

1.2 Japanese word formation 5

1.3 Visual word recognition 7

1.4 The Japanese mental lexicon 8

Chapter 2 12
The Japanese writing system:
Classification and terminology issues

2.1 Descriptions of the Japanese writing system: A collection of superlatives 12

2.2 A multi-script writing system 15

2.2.1 Introduction of writing to Japan 15

2.2.2	*Development of kana*	*16*
2.2.3	*Kanji limitations and the present Jōyō kanji list*	*17*
2.2.4	*Orthographic features of the Japanese writing system*	*19*
2.3	*The dual-reading system*	*20*
2.3.1	*Adapting Chinese characters for the Japanese language*	*20*
2.3.2	*Multiple on-readings*	*21*
2.3.3	*Multiple readings for Jōyō kanji*	*23*
2.4	*Issues in classifying the Japanese writing system*	*24*
2.4.1	*Creation and formation of kanji*	*25*
2.4.2	*Classifications of writing systems*	*27*
2.4.2.1	*Taylor's (1883) classification*	*28*
2.4.2.2	*Gelb's (1952) classification*	*29*
2.4.2.3	*Diringer's (1962) classification*	*30*
2.4.2.4	*Hill's (1967) classification*	*31*
2.4.2.5	*Haas' (1976) classification*	*32*
2.4.2.6	*Halliday's (1985) classification*	*32*
2.4.2.7	*Sampson's (1985) classification*	*33*
2.4.2.8	*DeFrancis' (1989) classification*	*35*
2.4.2.9	*Daniels' (1996, 2001) classification*	*37*
2.4.2.10	*Sproat's (2000) classification*	*38*
2.4.3	*Terminology issues*	*39*
2.4.3.1	*Language and mediums of expression*	*40*
2.4.3.2	*Pictographs and the ideographic myth*	*46*
2.4.3.3	*Logographic versus morphographic*	*47*

2.5 The Japanese writing system: Summary 49

Chapter 3 51
Japanese word formation:
Two-kanji compound word morphology

3.1 The Japanese lexicon 52

3.1.1 Lexical stratum 53

3.1.2 Simplex and complex words 55

3.2 Japanese compound words 57

3.2.1 Syntactic compounds 58

3.2.1.1 Modifier + modified 58

3.2.1.2 Verb + complement 60

3.2.1.3 Complement + verb 61

3.2.2 Semantic compounds 65

3.2.2.1 Associative pairs 65

3.2.2.2 Synonymous pairs 66

3.3 The morphology of two-kanji compound words 67

3.4 Japanese word formation: Summary 73

Part 2

The Japanese mental lexicon

Chapter 4 76
Visual word recognition:
Tasks, effects, issues and models

4.1 Tasks 78

4.1.1	*Lexical decision*	*79*
4.1.2	*Naming*	*79*
4.2	*Effects*	*81*
4.2.1	*Word frequency and word familiarity effects*	*81*
4.2.2	*Non-word legality effect*	*82*
4.2.3	*Neighborhood effects*	*83*
4.2.4	*Morphological family effects*	*85*
4.2.5	*Semantic priming effect*	*86*
4.2.6	*Constituent morpheme priming effects*	*89*
4.3	*Issues and models*	*92*
4.3.1	*Lexical retrieval: From print to meaning*	*93*
4.3.1.1	*Serial search model: Forster (1976, 1989)*	*94*
4.3.1.2	*Interactive-activation and competition (IAC) model: McClelland & Rumelhart (1981), Rumelhart & McClelland (1982)*	*96*
4.3.1.3	*Distributed activation model: Seidenberg & McClelland, (1989)*	*98*
4.3.2	*Lexical representation: Representing polymorphemic words*	*101*
4.3.2.1	*Prefix-stripping model: Taft & Forster (1975, 1976)*	*103*
4.3.2.2	*Augmented addressed morphology model: Caramazza, Laudanna, & Romani (1988)*	*105*
4.3.2.3	*Multilevel interactive-activation framework: Taft (1991, 1994)*	*106*

4.3.2.4 *Parallel dual-route model of morphological processing: Schreuder & Baayen (1995), Baayen, Dijkstra and Schreuder (1997), and Baayen & Schreuder (1999)* *108*

4.4 *Visual word recognition research: Summary* *111*

Chapter 5 *112*
The Japanese mental lexicon:
The lemma unit model

5.1 *Psycholinguistic studies of kana and kanji processing: A brief review* *113*

5.1.1 *Kana studies* *113*

5.1.2 *Kanji studies: Single characters* *115*

5.1.3 *Kanji studies: Two-kanji compound words* *118*

5.1.3.1 *Clusters and serial search: Hirose (1992)* *118*

5.1.3.2 *Distributed activation model for Japanese: Ijuin, Fushimi, Patterson, and Tatsumi (1999)* *123*

5.1.3.3 *Decomposed storage: Zhang & Peng (1992)* *124*

5.1.3.4 *Dual-route model of morphological processing for Chinese: Liu & Peng (1997)* *126*

5.1.3.5 *Multilevel interactive-activation model for Chinese: Taft & Zhu (1997)* *128*

5.1.3.6 *Companion-activation model: Saito (1997)* *134*

5.1.3.7 *Multilevel interactive-activation model for Japanese: Tamaoka & Hatsuzuka (1998)* *135*

5.1.3.8 *Lemma unit version for Chinese: Taft, Liu, & Zhu (1999)* *137*

5.1.3.9	*Lemma unit model for Japanese: Joyce (1999; in press)*	139
5.2	*Constituent morpheme priming for two-kanji compound words Joyce (1999, in press)*	140
5.2.1	*Purpose*	141
5.2.2	*Survey 1: Two-kanji compound word classification survey Joyce and Ohta (1999)*	144
5.2.2.1	*Purpose*	144
5.2.2.2	*Method*	145
5.2.2.3	*Results*	147
5.2.2.4	*Discussion*	148
5.2.3	*Experiment 1: Primed lexical decision task*	149
5.2.3.1	*Purpose*	149
5.2.3.2	*Method*	149
5.2.3.3	*Results*	153
5.2.3.4	*Discussion*	156
5.2.4	*Experiment 2: Automatic primed lexical decision task*	157
5.2.4.1	*Purpose*	159
5.2.4.2	*Method*	159
5.2.4.3	*Results*	162
5.2.4.4	*Discussion*	166
5.3	*The Japanese mental lexicon: Summary*	166

Chapter 6 *172*
The lemma unit model:
Testing the model

6.1 *Experiment 3:* *173*
Primed naming task

6.1.1 *Purpose* *174*

6.1.2 *Method* *174*

6.1.3 *Results* *180*

6.1.4 *Discussion* *183*

6.2 *Cross-script priming (Kana orthography version)* *184*
Joyce (2001)

6.2.1 *Experiment 4:* *188*
On-reading primes

6.2.1.1 *Purpose* *188*

6.2.1.2 *Method* *188*

6.2.1.3 *Results* *192*

6.2.1.4 *Discussion* *195*

6.2.2 *Experiment 5:* *196*
Kun-reading primes

6.2.2.1 *Purpose* *196*

6.2.2.2 *Method* *196*

6.2.2.3 *Results* *199*

6.2.2.4 *Discussion* *203*

6.3 *Cross-modality priming (Auditory priming)* *204*
Joyce (2001)

6.3.1 *Experiment 6:* *205*
On-reading primes

6.3.1.1 *Purpose* *205*

6.3.1.2	*Method*	*205*
6.3.1.3	*Results*	*207*
6.3.1.4	*Discussion*	*210*
6.3.2	*Experiment 7: Kun-reading primes*	*211*
6.3.2.1	*Purpose*	*211*
6.3.2.2	*Method*	*211*
6.3.2.3	*Results*	*213*
6.3.2.4	*Discussion*	*216*
6.4	*Positional sensitivity*	*217*
6.4.1	*Survey 2: Constituent morpheme frequency data Joyce and Ohta (in press)*	*219*
6.4.1.1	*Purpose*	*219*
6.4.1.2	*Method*	*219*
6.4.1.3	*Results*	*220*
6.4.1.4	*Discussion*	*222*
6.4.2	*Experiment 8: Positional sensitivity experiment*	*226*
6.4.2.1	*Purpose*	*226*
6.4.2.2	*Method*	*227*
6.4.2.3	*Results*	*229*
6.4.2.4	*Discussion*	*233*
6.5	*Experiment 9: Compound-word priming*	*234*
6.5.1	*Purpose*	*234*
6.5.2	*Method*	*235*

6.5.3 Results 236

6.5.4 Discussion 240

6.6 The lemma unit model: Summary 241

Part 3
Conclusion and future research

Chapter 7 246
Conclusion:
Morphology and the mental lexicon

7.1 The Japanese writing system 247

7.2 Japanese word formation 249

7.3 Visual word recognition 249

7.4 The Japanese mental lexicon 251

Chapter 8 258
Future research:
Different procedures, different words, and different tasks

8.1 Different procedures 259

8.2 Different words 260

8.3 Different tasks 262

Notes 265

References 293

Appendices 327

Appendix A: Experiment 1 stimuli 327

Appendix B: Experiment 1 instructions *329*

Appendix C: Experiment 2 stimuli *330*

Appendix D: Experiment 2 instructions *333*

Appendix E: Experiment 3 stimuli *334*

Appendix F: Experiment 3 apparatus *336*

Appendix G: Experiment 3 instructions *337*

Appendix H: Experiment 4 stimuli *338*

Appendix I: Experiment 4 instructions *340*

Appendix J: Experiment 5 stimuli *341*

Appendix K: Experiment 5 instructions *343*

Appendix L: Experiment 6 instructions *344*

Appendix M: Experiment 7 instructions *345*

Appendix N: Experiment 8 stimuli *346*

Appendix O: Experiment 9 instructions *347*

List of Tables

2.1	Examples of on-readings	22
2.2	Distribution of on-readings and kun-readings among the 1945 Jôyô kanji	24
2.3	Examples of formation principles for kanji	26
2.4	Halliday's (1985) classification of writing systems	33
2.5	Daniels' (1996, 2001) classification of writing systems	37
2.6	Terms used for kanji in various classifications of writing systems	40
3.1	Examples of similar words at different lexical stratums	55
3.2	Examples of adjective + noun and verb + noun modifier + modified compounds	58
3.3	Examples of noun + noun modifier + modified compounds	59
3.4	Examples of verb + complement compounds	60
3.5	Examples of Sino-Japanese complement + verb compounds	61
3.6	Examples of Native Japanese complement + verb compounds	63
3.7	Examples of referents of Native Japanese complement + verb compounds	64
3.8	Examples of associative pairs	65
3.9	Examples of synonymous pairs	66
3.10	Examples of orthographic modifications	68
3.11	Nomura's (1988a) classification of word bases and affixes	69
3.12	Nomura's (1988a) classification of the combination patterns for two-kanji compound words	71
3.13	Word-formation principles underlying two-kanji compound	73

words

5.1	*Prime-target relation conditions in Hirose (1992) Experiment 1*	*119*
5.2	*Prime-target relation conditions in Hirose (1992) Experiment 2*	*121*
5.3	*Prime-target relation conditions in Hirose (1992) Experiment 3*	*122*
5.4	*Examples of the five word-formation principles for two-kanji compound words examined in this study*	*145*
5.5	*Prime-target conditions in Experiment 1*	*149*
5.6	*Mean classification scores and familiarity scores for target compound words used in Experiment 1*	*150*
5.7	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 1*	*154*
5.8	*Mean classification scores and familiarity scores for target compound words used in Experiment 2*	*160*
5.9	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 2*	*163*
6.1	*Mean classification scores and familiarity scores (NTT database) for target compound words used in Experiment 3*	*176*
6.2	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 3*	*181*
6.3	*Mean classification scores and familiarity scores (NTT database) for target compound words used in Experiment 4*	*190*
6.4	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 4*	*193*
6.5	*Mean classification scores and familiarity scores (NTT*	*198*

	database) for target compound words used in Experiment 5	
6.6	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 5*	*201*
6.7	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 6*	*209*
6.8	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 7*	*214*
6.9	*Annual type and token counts, together with total type and token counts for the six-year period (1993-1998)*	*221*
6.10	*Examples kanji with cumulative type and token frequency data*	*222*
6.11	*Word-formation principle conditions in Experiment 8*	*227*
6.12	*Mean Classification Scores and Familiarity Scores (NTT database), as well as Mean Positional Ratios for Target Compound Word Used in Experiment 8*	*229*
6.13	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 8*	*231*
6.14	*Target conditions in Experiment 9, compound-word priming*	*235*
6.15	*Error rates (as percentages) and mean reaction times (in milliseconds) as a function of word formation principle and prime-target relation in Experiment 9*	*238*

List of Figures

2.1	*Gelb's (1952) stages of the development of writing*	*29*
2.2	*Sampson's (1985) classification of writing systems*	*34*
2.3	*DeFrancis' (1989) classification of writing systems*	*36*
2.4	*Sproat's (2000) non-arboreal classification of writing systems*	*39*
2.5	*Schematic representation of the 'language is speech' view and the abstract entity view*	*43*
2.6	*Logographic versus morphographic*	*48*
3.1	*Structural classification of Japanese words*	*56*
4.1	*Forster's (1976) serial search model, showing the organization of peripheral access files and the master file*	*94*
4.2	*The IAC model, showing an outline of some of the processing levels involved in visual and auditory word perception, with interconnections*	*97*
4.3	*General framework for lexical processing in the Seidenberg and McClelland (1989) distributed activation model*	*99*
4.4	*The multilevel interactive-activation model incorporating a morpheme level, depicting the representation for INVENT*	*106*
4.5	*General outline of the Schreuder and Baayen (1995) model of morphological processing*	*109*
5.1	*Mean reaction times (in milliseconds) as a function of prime-target relation in Hirose's (1992) Experiment 1*	*119*
5.2	*The network architecture of Ijuin, et al.'s (1999) connectionist model of word naming for two-kanji compound words*	*123*
5.3	*Taft and Zhu's (1997) multilevel interactive-activation framework for conceptualizing the lexical processing of Chinese words*	*129*

5.4	*Saito's (1997) companion-activation model (CAM)*	134
5.5	*Tamaoka and Hatsuzuka's (1998) version of the multilevel interactive-activation framework for Japanese*	136
5.6	*Results of classification survey plotted as cumulative percentage curves for word-formation principles*	147
5.7	*Presentation sequence during a trial in Experiment 1*	152
5.8	*Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 1*	155
5.9	*Presentation sequence during a trial in Experiment 2*	161
5.10	*Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 2*	164
5.11	*Lemma-unit-mediated connections in the Japanese lemma unit model (Joyce, 1999; in press)*	169
6.1	*Screen print of Chart showing waveform of participant response in Channel 1 pane (upper section) and luminance-sensor signal in Channel 2 pane (lower section)*	178
6.2	*Presentation sequence during a trial in Experiment 3*	179
6.3	*Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 3*	182
6.4	*Presentation sequence during a trial in Experiment 4*	191
6.5	*Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 4*	194
6.6	*Presentation sequence during a trial in Experiment 5*	198
6.7	*Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 5*	202
6.8	*Presentation sequence during a trial in Experiment 6*	206
6.9	*Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 6*	210
6.10	*Presentation sequence during a trial in Experiment 7*	213

6.11 *Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 7* *215*

6.12 *Distribution of positional frequency ratios for 2,743 kanji based on (a) cumulative type counts and (b) cumulative token counts* *225*

6.13 *Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 8* *232*

6.14 *Presentation sequence during a trial in Experiment 9* *236*

6.15 *Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 9* *239*

7.1 *Tsuzuki, Kawamoto, and Yukihiro's (1999) parallel distributed processing model of lexical ambiguity resolution* *255*

7.2 *Levelt's (1989) notion of a lexical entry consisting of a lemma and morpho-phonological form* *256*

Part 1

How do kanji function as a writing system?

Chapter 1

Introduction:
From external form to internal representation

The central question addressed in this thesis is *how do kanji function as a writing system?* The answer proposed in this thesis is *kanji function as a morphographic writing system*—a writing system where the orthographic units represent morphemes. The rest of the thesis is devoted to understanding the significance of this question, and to presenting the linguistic discussion of the classification of writing systems and the psycholinguistic investigation of the Japanese mental lexicon that inform this answer.

In seeking to understand this question and its answer, the thesis explores and traces the connections between four broad areas; namely, the classification of the Japanese writing system, Japanese word formation, visual word recognition, and, in particular, the Japanese mental lexicon. Given that the nature of a writing system will be reflected to some extent in the structure of the literate mental lexicon, the transition from writing system to mental lexicon is simply a shift in perspective—from external form to internal representation.

Before turning, however, to briefly map out these four areas and the connections between them in this introductory chapter, a few words of explanation are required concerning *two-kanji compound words*, which this thesis takes as its principle frame of reference. In general, when we speak of two-kanji compound words, we shall be referring to words that more accurately would be described as words consisting of two morphemes, each of which is represented orthographically by a kanji character.[1] And, for the sake of completeness, morphemes, in turn, will be defined as the smallest elements of linguistic meaning in a language (Anderson, 1992; Bauer, 1983; Matthews, 1991; Singleton, 2000).[2] As the distinction will bear on our discussion later, we may also note that morphemes can be either free morphemes, which can stand alone as simplex words, or bound morphemes, which only appear as elements of complex words.

The emphasis on the two-kanji compound word, rather than on the single kanji character, follows from a belief that in seeking to understand how a writing system functions, it is essential to identity the principles that govern the combination of graphic units. Given the central claim of this thesis that kanji are a morphographic writing system, the relations between morphemes, kanji and kanji readings are at the heart of this thesis. It is extremely important, however, not to confuse elements of meaning, orthographic form, and phonological form. This is because word structure and word formation, as we shall discuss in Chapter 3, are primarily issues of morphology dealing with abstract linguistic units, rather than issues of orthography or phonology, which are concerned with units of representation. The term two-kanji compound word is not intended to imply otherwise, but is merely a shorthand convention to refer to such compound words from an orthographic perspective as we discuss how kanji function as a writing system.

With these points in mind, we can now proceed with outlining the broad areas of this thesis and the connections between them that will be the focuses of our discussions along our journey from external form to internal representation.

1.1 The Japanese writing system

The Japanese writing system is often portrayed as being extremely complex. While its multi-script nature and the dual-reading system undoubtedly contribute to the complexity of the Japanese writing system, I would argue, however, that this portrayal is also a reflection of the controversy and heated debate that has surrounded the appropriate classification of the Japanese writing system, and, especially of kanji, within a linguistic taxonomy of writing systems. Our central question of how kanji function as a writing system turns out to be quite a serious matter, which depends not only on how we perceive the creation and historical development of writing systems, but also on how we believe speech and writing to be related, and even on what we comprehend language to be (Henderson, 1982).

After briefly describing the Japanese writing system in terms of its multi-script nature and the dual-reading system, Chapter 2 of the thesis examines in some detail the issues in appropriately classifying kanji as a writing system. Following a review of a number of important classifications of writing systems that seeks to highlight some of their underlying assumptions about writing, we discuss some of the typological labels, reflecting these assumptions, that have been applied to kanji as used in both the Chinese writing system and the Japanese writing system.

The understanding that emerges from this discussion is that while we must be careful to remember that there are no pure writing systems (Gelb, 1952), still the distinctions we apply in differentiating different systems should as far as possible seek to capture the complex relations between abstract linguistic elements and elements of representation, namely, orthographic and phonological units for different classes of systems. Although, the exact form of these three-way relations vary with the language and the writing system employed, this thesis argues that for Japanese, we can most accurately characterize these relations by recognizing that kanji function as a morphographic writing system.

1.2 Japanese word formation

In the discussions over the appropriateness of the terminology used for kanji, much attention has been given to the principles of their formation, and to the relative importance of these principles. As we shall note in Chapter 2, DeFrancis (1989), for instance, has written at length on this, and in line with his view that writing is simply the visual representation of speech has emphasized the phonological aspect of Chinese characters, even to the rather extreme point of suggesting that they are 100 percent syllabic (p. 102). It is unfortunate that these debates should have been so polarized, for simply pointing to the unpredictable nature of the radicals, or the unreliable character of the phonetic markers only highlights the fact that kanji form a natural system that utilizes both phonological and semantic elements in various combinations to represent individual morphemes. It is unfortunate also that this debate has continued for so long, and has focused so exclusively on the nature of the single character, for in the process very little attention has been given to providing an adequate explanation of the different ways in which kanji are combined in representing polymorphemic (compound) words in Japanese. Compounding is extremely productive in Japanese (Kageyama, 1999), and the two-kanji compound word is the most common word structure in the Japanese language (Nomura, 1988), according to one estimate, accounting for 70 percent of all Japanese words (Yokosawa & Umeda, 1988).[3] Given that the question of how orthographic units are combined to represent language would seem to be a fundamental one for a satisfactory account of how a writing system functions, it is somewhat surprising to find that this has not been given greater attention in the discussions about classifying how kanji function as a writing system.

Chapter 3, which discusses Japanese word formation and the morphology of two-kanji compounds, is precisely about identifying the principles that underlie the combination of kanji in two-kanji compound words. The important insight to be gained from this discussion is that in the relations between abstract linguistic elements, morphemes, and elements of expression—orthographic and phonological units, which in this context are primarily kanji and their readings—priority is at the abstract linguistic level, and the combination of elements is primarily morphologically motivated, with the representation of the resulting polymorphemic units being based on the representations of component elements.

The central claim that kanji function as a morphographic writing system is being made in this thesis precisely because, in contrast to the commonly used term logographic, the term morphographic most accurately reflects and emphasizes this basic reality—the fact that the principles underlying the combination of graphic units in two-kanji compound words (and other polymorphemic Japanese words) are primarily morphological.

The discussion of two-kanji compound word morphology in this chapter also provides important background for the experiments presented in Part 2 of this thesis where the lexical retrieval and representation of two-kanji compound words within the Japanese mental lexicon is examined from a morphological perspective by controlling for the word-formation principle of compound word stimuli as an experimental condition.

1.3 Visual word recognition

The transition from writing system to mental lexicon in this thesis—the shift in perspective from external form to internal representation—takes us directly through the area of visual word recognition research. Hopefully, the reader will find the move from the linguistic study of writing systems to psycholinguistic research into visual word recognition to be an obvious and natural one. We would, however, do well to remember Harley's (2001) observation about the limitations of what linguistics can tell us about the psychological processes involved in speaking and understanding language (p. 12), as we seek to investigate the extent to which the broad distinctions we might make in classifying writing systems are revealing about the complex process of lexical retrieval and the organization of representations in the mental lexicon.

After reviewing some of the important tasks and effects that have influenced visual word recognition research, Chapter 4 discusses the two central issues for the mental lexicon. The first issue of lexical retrieval, or lexical access, is primarily concerned with the journey from the printed word to the retrieval of a word's meaning. The second issue of lexical representation is primarily interested in how words, especially polymorphemic words, might be arranged and linked in the mental lexicon. As we shall see later when we look at how some of the most influential models deal with the two issues, lexical retrieval and lexical representation are, however, intricately related. This is because the way in which lexical information is represented and organized in the mental lexicon will greatly determine what kind of mechanisms can be assumed as underlying lexical retrieval, just as mechanisms of lexical retrieval limit and constrain the representation and arrangement of lexical information. From a brief review of a number of models that are concerned primarily with lexical representation, we shall also see how most allow for some degree of morphological involvement in the lexical retrieval of polymorphemic words, highlighting the language-universal importance of morphology for the organization of the mental lexicon.

1.4 The Japanese mental lexicon

The central question of how kanji function as a writing system is, as we have already mentioned, a serious matter of direct relevance for our understanding of writing and language. Reflecting this and the importance of morphology for lexical retrieval and representation, the answer proposed here that kanji function as a morphographic writing system clearly has fundamental implications for our understanding of the Japanese mental lexicon.

Chapters 5 and 6 examine these implications for the organization of the Japanese mental lexicon by investigating the lexical retrieval and representation of two-kanji compound words from a morphological perspective. Specifically, this thesis advocates a lemma unit model version of the multilevel interactive-activation framework as a model for the Japanese mental lexicon

(Joyce, 1999, in press).

After a brief review of psycholinguistic studies of kana and kanji processing, Chapter 5 discusses a number of studies from both the Chinese and Japanese literature that have focused on two-kanji compound words. In particular, two proposals are considered as models of the Japanese mental lexicon that differ in terms of lexical retrieval and representation. The first proposal is Hirose's (1992, 1994, 1996) hypotheses which evoke search mechanisms and claim that the mental lexicon for two-kanji compound words is structured so that words sharing the same first kanji are linked in clusters with the first kanji serving as a retrieval cue, but that words sharing the same kanji as a second constituent are not. In contrast, activation mechanisms are assumed in the second proposal made by Joyce (1999, in press) for a Japanese lemma-unit version of the multilevel interactive-activation framework adapted from the recently modified version for Chinese (Taft, Liu, & Zhu, 1999).

As any model of the Japanese mental lexicon must in some way capture the morphological relations that exist between polymorphemic words, we shall investigate how well these two proposals cope with the diversity in the morphology of two-kanji compound words. Because these proposals make different predictions concerning constituent-morpheme priming, the patterns of facilitation will be examined in two experiments varying stimulus onset synchronicity (SOA) with five word-formation principles as experimental conditions. The results from these experiments, of similar levels of priming for both constituents in all but one word-formation condition, are more consistent with the predictions from the Japanese lemma-unit model.

The major feature of the lemma unit model is the incorporation of lemma units that mediate the links between access representations and semantic units. As such, the model provides a very promising approach to modeling the complex relations between meaning, orthography and phonology that exist within the Japanese mental lexicon. In Chapter 6, the lemma unit model is investigated further through a series of seven experiments that seek to examine various aspects of this model of the Japanese mental lexicon. While these experiments continue to focus on the lexical retrieval and representation of two-kanji compound words from a morphological perspective, they also investigate the extent to which the results in the first two experiments might be task specific, the possibility that the lemma unit model offers to integrate both kana and kanji processing within a single model, and the potential to model the dual reading system of on-readings and kun-readings. In addition, by employing recently compiled constituent-morpheme frequency data (Joyce & Ohta, in press), the issue of positional sensitivity is examined for verbal constituent morphemes of two-kanji compounds words. Although the results from this series of experiments raise further questions, they are generally consistent with the lemma unit model, which is undoubtedly an extremely appealing way of thinking about the relations that exist between semantic representations, and access representations for orthography and phonology, as well as accounting for the morphological information that underlies two-kanji compound words in the Japanese mental lexicon.

Thus, the answer to our central question that emerges from the exploration of the four broad areas of this thesis is that *kanji function as a morphographic writing system*; a fact which is reflected in the organization of the Japanese mental lexicon. In investigating the lexical retrieval and representation of two-kanji compound words within the Japanese mental lexicon from a morphological perspective, this research provides important insights not only for our understanding of kanji, but also into the nature of writing systems and of language itself.

Chapter 2

The Japanese writing system: Classification and terminology issues

In expressing their language in writing, the Japanese people use a mixture of different types of scripts as elements of a largely complementary system of writing (Kabashima, 1977; Kaiser, 1993a; Kess & Miyamoto, 1999; Smith, 1996; Taylor & Taylor, 1995). The elements of the Japanese writing system are 漢字 *kanji* (literally, 'Chinese characters'), 平仮名 *hiragana*, 片仮名 *katakana*, and, increasingly,ローマ字 *rōmaji* (the alphabet), as well as Arabic numerals. As Kess and Miyamoto (1999) observe, with this mixture of scripts, the "Japanese may have the unique distinction of employing all three extant means" (p. 9) of expressing language in writing.

After a brief description of the Japanese writing system in terms of its multi-script nature and the dual-reading system, this chapter addresses the central question of this thesis of how do kanji function as a writing system by examining linguistic classifications of the Japanese writing system and, in particular, the issues underlying the terminology debates relating to kanji.

2.1 Descriptions of the Japanese writing system: A collection of superlatives

The Japanese writing system is often portrayed as being extremely complex.

Indeed, even from its first contact with the West, the Japanese writing system has had a bad press:

> The complex Japanese language and its writing system are inventions of the devil, designed to prevent the spread of Gospel. (attributed to Francis Xavier (1506-1552), cited in Taylor & Taylor, 1995, p. 279)

While modern scholars no longer see the hand of the devil at work, the notion that the Japanese writing system is extremely complex persists, and is commonly expressed in works on writing systems. For instance, commenting on the adaptation of Chinese characters to the Japanese language, Coulmas (1989) writes that,

> under the hands of the Japanese, Chinese characters were transformed to become what is often said to be the most intricate and complicated writing system ever used by a sizeable population. (p. 122)

Smith (1996) suggests that there are two reasons for this portrayal. The first is the multi-script nature of the Japanese writing system, employing kanji, kana syllabaries, as well as rōmaji. Claiming that Japanese can be written entirely in kana, Unger (1987) believes that the system is unnecessarily complex, and accordingly describes kanji as "just a burdensome collection of visual abbreviations" (p. 35).[4] A similar position is taken by DeFrancis (1989), who remarking on the development of kana syllabaries from Chinese characters, comments,

> it is an ironic fact, however, that while the Japanese developed a system of sound representation that was almost perfectly suited to their language, they ended up with one of the worst overall systems of writing ever created. (p. 138)

The other reason suggested by Smith (1996) for this portrayal is the complexity arising from the dual system of on-readings and kun-readings for kanji. The remarks of Sansom (1928), writer of an early grammar of Japanese, have often been quoted in this respect, which refer to the common custom in his day of indicating readings for kanji in newspapers with 振り仮名 *furigana* glosses.

> One hesitates for an epithet to describe a system of writing that is so complex that it needs the aid of another system to explain it. There is no doubt that it provides for some a fascinating field of study, but as a practical instrument it is surely without inferiors. (p. 44)

Certainly, it cannot be denied that these two aspects of the Japanese writing system, which are discussed further below, add to the complexity; on the other hand, the fact that the Japanese writing system is capable of meeting the functional needs of a modern, literate society (Smith, 1996)[5] seems to suggest that there has been a tendency to overstate the situation.[6] I believe that Sproat (2000) actually comes closer to identifying the major reason for the portrayal of the Japanese writing system as being extremely complex when he observes that,

> Japanese is surely the most complex modern writing system, and the hardest to force into any taxonomic mold. (p. 132)

The main objective of this chapter is, therefore, to discuss the issues in appropriately classifying the Japanese writing system, and, in particular kanji, within a linguistic taxonomy of writing systems, which are taken up in Section 2.4.

2.2 A multi-script writing system

Implicit in the criticisms of the multi-script nature of the Japanese writing system is the assertion that kanji are simply too numerous and too complex, and on the claim that Japanese can be written entirely in kana, completely unnecessary. This section describes the development of this multi-script system.

2.2.1 Introduction of writing to Japan

Perhaps one of the most remarkable things about writing is the fact that it has only been independently invented a few times in the history of mankind. All modern writing systems can be traced back in some way to one of three ancient inventions of writing; the Sumerian cuneiform writing system, the Mayan script in Mesoamerica, and Chinese characters (Coulmas, 1989, 1996; Daniels, 1996, 2001; DeFrancis, 1989; Gaur, 1992; Gelb, 1952; Nakanishi, 1980; Sampson, 1985; Senner 1989). As the process in which one culture learns of writing from a neighboring culture and borrows their writing system is one that has been repeated many, many times in history, it should be nothing of a surprise to discover that the Japanese people number among the vast majority of peoples who did not invent writing for themselves.[7]

For Japan, the neighboring culture was Chinese, the writing system used Chinese characters, and the borrowing probably started as early as the first century CE.[8] Miller (1967) remarks of this writing system that "in many ways it suited the Chinese language admirably" (p. 92). However, he also notes later that “Japanese and Chinese are historically two quite different languages of totally different genetic origin and entirely different structures” (1986, p. 30), which might lead us to speculate about how suitable this system was for the Japanese language. Certainly the Japanese have had to adapt it to fit the needs of their language, but perhaps we would do well to remember that for the Japanese, writing was not something which was introduced to them with a range of options. Rather, at that time, writing must have appeared inextricably linked to the Chinese language and its writing system.[9] Viewed in this way, the fact that the Japanese adopted Chinese characters for their writing system is clearly a matter of historical coincidence rather than choice.

2.2.2 Development of kana

During the ninth century, the Japanese developed two syllabaries, 片仮名 katakana and 平仮名 hiragana. Although these were developed separately, both are based on the use of kanji in phonetic transcription. The practice of using a character only for its phonetic value, with no reference to its semantic value, was similar to how the Chinese themselves dealt with foreign names. This method was employed in the 万葉集 *Man'yōshū* (759 CE), an anthology of Japanese verse, and consequently, characters used in this way are often referred to as 万葉仮名 *man'yōgana* (Miller, 1967; Shibatani, 1990).

Katakana was developed by Buddhist priests as part of a system for annotating their texts written in Chinese. In the reading convention known as 訓読 *kundoku*, texts would be marked to indicate Japanese word order and man'yōgana were used to write some particles, inflections, and kun-readings between the Chinese characters. However, due to limited space between the lines and the need to write quickly while taking notes, the priests would write the man'yōgana in a simplified form, usually by isolating a distinctive feature of the character (Habein, 1984).

Although hiragana also developed from man'yôgana, the process was somewhat different. From the early Heian period (794-1185), man'yôgana were written in a cursive hand called 草仮名 *sōgana* 'grass style,' and hiragana represent a simplified and more cursive form of sōgana, bearing little resemblance to the man'yōgana from which they were ultimately derived (Habein, 1984).

The development of these kana scripts made it possible to write the sounds of the Japanese language without the use of kanji, as indeed women did during the Heian period. However, the development of kana scripts did not replace the use of kanji completely,[10] but led to a mixed writing system in which kana came to represent elements of Japanese grammar, and kanji continued to be used to represent nouns and verbal stems. This is essentially the system which is still used and is now known as 漢字かな混じり文 *kanji-kana-majiribun* 'mixed kanji and kana writing.'

2.2.3 Kanji limitations and the present Jōyō kanji list

As the Chinese language developed over time, so the number of Chinese characters increased. From the earliest forms found on bones, which numbered about 2,500, the number of characters in use had increased to almost 10,000 by the Han dynasty. This trend continued, so that by the twelfth century the number was 23,000 and in the eighteenth century the figure was almost 49,000 (Coulmas, 1989). In theory, any character known in Chinese was available for use in Japanese (Miller, 1986). However, these counts are a little misleading, for they include various graphic versions of the same kanji, as well as very obscure kanji used only a few times in a rare text, but still the number of characters in use in writing Japanese was more than 10,000 at the end of the Tokugawa period (1603-1868) (Twine, 1991).

国字問題 *Kokuji mondai* 'problems of the national script' refers to the Japanese debate over the difficulties involved in the Japanese writing system. This debate can be traced back to the Meiji period, when the written language was seen as an obstacle to Japan's modernization. It was in this context that intellectuals turned their attention to the Japanese writing system, with a number of proposals being put forward (Koizumi, 1989; Seeley, 1991; Twine; 1991). These fell into three groups, calling for: (a) the use of kana only, (b) romanization, and (c) a continuation of the existing writing system, but with restrictions on the number of kanji employed (Seeley, 1984).

Although there was some enthusiasm for the first two kinds of proposals, the greatest support was for a continuation of the existing writing system with restrictions on the numbers of kanji used. Although the implementation of script reforms was delayed until after the Second World War, the general tendency has indeed been to reduce the number of kanji in daily use, along with some simplification of character forms. The 常用漢字表 *Jōyō kanjihyō* 'List of Characters for General Use,' which the Japanese Cabinet introduced in October 1981, includes 1945 kanji.

To put this figure in clearer perspective, however, it must be noted that this is not an upper limit for the number of kanji in daily use, for the present list is actually only a guideline for kanji usage representing a less restrictive attitude than that behind the previous official list of the 当用漢字表 *Tōyō kanjihyō*. A stronger candidate for the upper limit figure would be the 6,355 kanji of the Japanese Industrial Standard (JIS) code (JIS X-0208-1990 includes 2,965 level 1 kanji and 3,390 level 2 kanji) (Lunde, 1993), which define character sets for computers. On the other hand, newspapers and official documents do generally follow the Jōyō kanji guidelines, and it has been estimated that the list covers over 98% of all newspaper kanji, with 90% being covered by the 1,006 教育漢字 *kyôiku kanji* 'educational kanji,' that are a subset of the Jōyō kanji taught during the 6 years of elementary school (Kaiser, 1993a). The small percentage not covered is due to kanji outside the list in family and place names, which keep the total number of kanji used in newspapers at over 3,200 (Seeley, 1984).[11]

2.2.4 *Orthographic features of the Japanese writing system*

The present Japanese writing system is, then, multi-script in nature, incorporating morphographic kanji, and two kana syllabaries, hiragana and katakana, as well as allowing for the use of the alphabet and Arabic numerals. These orthographic elements are employed in largely separate and complementary ways to represent the Japanese language in writing (Kabashima, 1977; Kaiser, 1993a; Kess & Miyamoto, 1999; Smith, 1996; Taylor & Taylor, 1995).

Kanji are used to write nouns, either singularly or in compounds with other kanji, and the stems of verbs and adjectives. Hiragana is used for inflectional elements, and other grammatical elements. The other kana script, katakana, is used to write foreign names and 外来語 *gairaigo* words of foreign origin (apart from those written in kanji), species names for animals and plants, and onomatopoeic expressions for emphasis, as a form of italics. The use of the Roman alphabet is common, especially in advertising and the mass media.

2.3 The dual-reading system

In this section, we will look at the two ways in which the Japanese came to use Chinese characters for writing their own vastly different language, for these are essentially the same ways that kanji are employed today. As Martin (1972) observes, they were to have far-reaching consequences, for they are also the methods that created the dual on-reading and kun-reading system.

2.3.1 Adapting Chinese characters for the Japanese language

At first, “writing” in Japan meant writing in Chinese, following Chinese syntax.[12] However, this pattern did not last long, for the Japanese set up conventions for reading classical Chinese according to Japanese syntax, mentioned above, to indicate reading order and pronunciations (Habein, 1984; Miller, 1967). Of course, once this happened, the language was no longer Chinese but Japanese, and the Chinese words and morphemes represented by the characters entered the Japanese language as loan words and morphemes, with

Japanese imitations of the Chinese readings, known as on-readings (Coulmas, 1989). The first way in which the Japanese came to use Chinese characters was, therefore, to write these new loan words and morphemes.

Although many Chinese words and morphemes for things entered the Japanese language as on-readings for kanji, the Japanese people already had their own words for many of these things. The second method of using kanji was, therefore, to associate the kanji with these native Japanese words referring to the same things, as kun- readings.

Thus, the character 人 'person' has in modern Japanese the on-reading *jin* from the Old Chinese word *jen*, and the kun-reading *hito*, based on the association of the kanji with the Old Japanese word *fitö* (Coulmas, 1989). Based on the shared semantic association to the morpheme represented by a given kanji, Morioka (1968) has referred to on-readings and kun-readings of a kanji as allomorphs.[13]

2.3.2 Multiple on-readings

Unfortunately, this is not the end of the story, for some kanji actually have more than one on-reading and one kun-reading. This is because Chinese characters were borrowed repeatedly at different periods and from different regions of China, bringing new readings and meanings (Kaiser, 1993a; Miller, 1967; Shibatani, 1990).

The earliest readings for kanji reflect the phonology of southern Chinese from the fourth century to the end of the sixth century. These readings are referred to as 呉音 *go'on*. The second wave of Chinese readings came around the beginning of the eighth century, as many Japanese returned from studying in Chang'an, the capital of the T'ang period (618-907), and these on-readings are referred to as 漢音 *kan'on*. This was followed during the fourteenth century by pronunciations from later varieties of Chinese, including T'ang and Sung (960-1279), and these are referred to as either 唐音 *tō'on* or 唐宗音 *tō'sō'on*. Examples of these on-readings are given in Table 2.1.

Table 2.1

Examples of On-Readings

	Reading		
Kanji character	*Go'on* 呉音	*Kan'on* 漢音	*Tō'on* 唐音
行 go; conduct	修行 *shugyō* training	孝行 *kōkō* filial piety	行灯 *andon* lantern
請 invite; request	起請 *kishō* vow, pledge	懇請 *konsei* entreaty	普請 *fushin* (temple) construction
頭 head	頭巾 *zukin* hood, cowl	頭髪 *tōhatsu* hair, head of hair	饅頭 *manjū* bean-jam cakes
経 longitude; sutra	経文 *kyōmon* Buddhist scripts	経験 *keiken* experience	看経 *kankin* (silent) reading of Buddhist scripts

Note. Adapted from Fujii (1996).

The different readings were associated with different spheres of learning: go'on with Buddhism, kan'on with Confucianism and other secular learning and tō'on with Zen Buddhism (Shibatani, 1990). This in part explains the distribution of the different readings, for although only a few kanji have tō'on readings, most have go'on and kan'on readings, and it was the association of go'on readings with Buddhist terms that prevented them from being replaced by the later kan'on readings. As an example of a kanji with multiple readings, Coulmas (1989) cites 頭 (general meanings of 'head,' 'chief,' 'top,' or 'beginning') as having four on-readings; the go'on reading *zu*, the kan'on reading *tō*/*do*, and the tō'on reading *ju*, as well as having six kun-readings; *saki*, *atama*, *kashira*, *kōbe*, *kaburi* and *tsumuri*. (p. 126)[14]

2.3.3 Multiple readings for Jōyō kanji

In addition to limiting the number of kanji in daily use, postwar script reforms have also reduced the number of recognized on-readings and kun-readings associated with kanji. Table 2.2 overleaf shows the distribution of on-readings and kun- readings for the Jōyō kanji list.

A problem in Japanese, which is particularly associated with on-readings, is the high incidence of homophones. As already noted, on-readings are Japanese imitations of pronunciations for Chinese words that were borrowed into Japanese with kanji. However, because of the simpler syllable structure of Japanese, distinctions between many different pronunciations in Chinese in terms of sounds and tones were lost. As an example, Martin (1972) remarks that the on-reading *kō* corresponds to such diverse syllables of classical Chinese as, *ko*, *kau*, *kou*, *kang*, *kwang*, and *kong*, each with aspirated vs. unaspirated initial and each with three separate tones. (pp. 98-99). The consequence of this is that many kanji share the same on-reading. Inspection of the Jōyō kanji list reveals, for instance, that 64 kanji have the on-reading *kō*, and 63 kanji have the on-reading *shō*.

Table 2.2

Distribution of On-readings and Kun-readings Among the 1945 Jōyō Kanji

	On-readings per character				
Kun-readings per character	0	1	2	3	Total
0	-	694	43	-	737
1	32	691	61	-	784
2	7	247	24	-	278
3	1	82	8	-	91
4	-	33	6	1	40
5	-	7	-	-	7
6	-	1	-	-	1
7	-	4	-	-	4
8	-	-	-	-	-
9	-	-	1	-	1
10	-	-	2	-	2
Total	40	1,759	145	1	1,945

Note. Excluding special readings. Based on Kaiho and Nomura (1983)

2.4 *Issues in classifying the Japanese writing system*

Before discussing the issues in classifying the Japanese writing system, it will be useful to look briefly at the classification of kanji according to their principles of formation, for much of the classification and terminology debates relate to how these principles have been perceived.

2.4.1 Creation and formation of kanji

From archeological evidence, it is generally believed that Chinese characters were first used no later than 1200 BCE. The earliest forms of Chinese characters known are the forms inscribed on ox scapulas and turtle plastrons during the Shang dynasty in north-central China as records of royal divinations, and often referred to as 甲骨文 *kōkotsubun* ‘oracle-bone inscriptions’ (Boltz, 1996; Habein & Mathias, 1991; Keightley, 1989).

Kanji forms are often classified according to the principles of their formation. Traditionally, six groups have been recognized, but, because two of these groups are actually based on principles of usage rather than formation, this is a little misleading.[15] The tendency recently has therefore been to focus on the four formation principles, often distinguishing between these in terms of two simple kanji groups and two complex kanji groups (Habein & Mathias, 1991; Halpern, 1990; Kaiser, 1993a). Examples of these formation principles are given in Table 2.3.[16]

The first group of 象形文字 *shōkei moji* pictographs are based on simple pictures of the physical objects they represent. Today, these kanji are highly stylized in form, due in part to changes in writing implements, so it is not always clear what is being depicted. Although many pictographs have relatively few strokes, some are more complex. The second group of 指示文字 *shiji moji* ideographs represent simple concepts. For example, for small numbers, this was done with tally-like strokes, for marking directions, by adding marks to a baseline, and other simple concepts, by highlighting a part of a pictograph. Kanji formed by these two principles are often referred to as simple kanji, and according to Habein and Mathias (1991) account for 152 Jōyō kanji (7.8%).

Table 2.3

Examples of Formation Principles for Kanji

Formation principle	Examples and explanations			
Simplex characters				
Pictographs	木	tree	人	person
象形文字	日	sun; day	山	mountain
	象	elephant; image	鳥	bird
Ideographs	一	one	二	two
指示文字	上	up	下	down
	本	roots; origin	末	end
Complex characters				
Semantic compounds	林	woods; grove	木 tree + 木 tree	
会意文字	休	rest	人 person + 木 tree	
	信	trust; believe	人 person + 言 word	
Phonetic compounds	侍	*ji* serve	人 person + 寺 *ji* temple	
形声文字	持	*ji* have	手 hand + 寺 *ji* temple	
	時	*ji* hour; time	日 sun; day + 寺 *ji* temple	

The other two formation principles involve the combination of kanji from the first two groups. The third group of 会意文字 *kaii moji* semantic compounds are based on combining characters from the first two groups to represent a meaning, which is a function of the meanings of the elements. These account for 483 Jōyō kanji (24.8%) (Habein & Mathias, 1991). The last formation principle is that of 形声文字 *keisei moji* phonetic compounds. This is by far the most important group, accounting for 1,310 Jōyō kanji (67.4%) (Habein & Mathias, 1991). These characters consist of an 意符 *ifu* semantic marker, or 部首 *bushu* radical, indicating the semantic field of the kanji, and an 音符 *onpu* phonetic marker indicating the reading. Kaiser (1993a) points out that for 58% of phonetic compounds, the phonetic marker gives a perfect indication, and that for nearly 33% the indication is partly reliable. Thus, phonetic markers provide no clue to the on-reading in less than 10% of these kanji. Historically, the number of radicals used in phonetic compound characters has been over 500, but this number was limited to 214 in the eighteenth century, and this is the number currently in common use.

2.4.2 Classifications of writing systems

Although Anderson (1992) is referring primarily to typologies of language, his observations about the nature of typologies are equally applicable to the classification of writing systems.

> We can conclude that the parameters of a typology ought to be ones from which something follows: that is, they ought to identify groups of properties that co-vary with one another, so that knowing how one things works entails knowing about others as well, as a direct consequence of whatever it is that motivates the typological labels. (p. 322)

Although the fact that there are no 'pure' writing systems (Gelb, 1952) means that classifications of writing systems can never be totally objective (Coulmas, 1996),[17] still the distinctions we apply in differentiating different systems should be as far as possible informative about how different writing systems function in representing language.

This section presents short reviews of a number of important classifications of writing systems. The aim of this section is to highlight some of the assumptions about writing that have shaped these typologies of writing systems, and to examine the treatment of kanji within these typologies, by noting some of the typological labels that have been applied to kanji as used in both the Chinese writing system and the Japanese writing system.

2.4.2.1 Taylor (1883) classification

One of the earliest classifications of writing systems seems to be that proposed by Taylor (1883) (as cited in Daniels, 1996, 2001; DeFrancis, 1989; Diringer, 1962; Hill, 1967; Kaiser, 1995).

DeFrancis (1989) describes Taylor's classification as being "an evolutionary model consisting of a succession of five stages" (p. 59),[18] from (1) pictures, (2) pictorial symbols, (3) verbal signs, (4) syllabic signs, to (5) alphabetic signs, with the first three stages being referred to as ideograms and the last two stages as phonograms. DeFrancis also notes that Taylor used the terms pictographs, ideograms and verbal phonograms with respect to Chinese characters. (p. 59)

2.4.2.2 Gelb's (1952) classification

No Writing: *Pictures*

Forerunners of Writing: *Semasiography*

1. Descriptive-Representational Device
2. Identifying-Mnemonic Device

Full Writing: *Phonography*

1. *Word-Syllabic:*	Sumerian (Akkadian)	Egyptian	Hittite (Aegean)	Chinese
2. *Syllabic:*	Elamite Hurrian etc.	West Semitic (Phoenician) (Hebrew) (Aramaic)	Cypriote Phaistos? Byblos?	Japanese
3. *Alphabetic:*		Greek Aramaic (vocalized) Hebrew (vocalized) Latin Indic etc.		

Figure 2.1. Gelb's (1952) stages of the development of writing.

Gelb's (1952) seminal work attempted to lay the foundation for the scientific study of writing, for which he coined the term *grammatology*.[19] The stages in the development of writing that Gelb presented are depicted in Figure 2.1.

While Gelb makes an important distinction between pictorial representation and mnemonic devices, which he referred to as the forerunners of writing, and full writing, his classification is undoubtedly flawed by his zeal to present a teleological account of writing, placing the alphabet as the final stage in a evolution via logography and syllabary (see Coulmas (1996), Daniels (2001) and Sproat (2000) for discussion of problems with Gelb's classification).

Reflecting his belief that full writing systems are phonographic in nature, Gelb (1952) refers to Chinese characters by the term word-syllabic.

2.4.2.3 Diringer's (1962) classification

As Hill (1967) observes, Diringer's (1962) classification is very similar to that proposed by Taylor (1883). Like Gelb's (1952) classification, Diringer's classification is also greatly influenced by the view that the alphabet represents the "most flexible and useful method of writing even invented" (p. 24). Although Diringer makes a distinction between what he calls embryo-writing and writing proper, his notion of *writing proper* is more inclusive than that underlying Gelb's conception of full writing. Thus, under writing proper, Diringer includes the five categories of (1) pictography, (2) ideography, (3) analytic transitional scripts, (4) phonetic scripts, and (5) alphabetic writing.[20]

The third category of analytic transitional scripts refers to writing systems that Diringer (1962) regarded as being in transition from pure ideographic to pure phonetic writing. The term analytic is intended to imply that the basic units of representation in these writing systems are words. Accordingly, Diringer classifies Chinese writing under this category of analytic transitional.

2.4.2.4 Hill's (1967) classification

Hill (1967) makes three criticisms of Diringer's (1962) classification. The first point is that phonetic scripts should include both syllabaries and alphabets. The second point is that alphabetic scripts can be used in various ways. And, the third criticism relates to the term ideographic. Hill claims that his classification deals with these points, and places "every system of writing in relation to that which all systems represent, language" (p. 92).

Hill's (1967) classification consists of three divisions—discourse systems, morphemic systems, and phonetic systems. After noting that all discourse systems are only partial systems, Hill remarks,

> it is quite striking, though not by any means always recognized, that there are no systems based on words. The reason seems to be that unless words are identified by their phonetic content, they must be identified in terms of units of meaning. The result is that a non-phonetic system forces analysis of the meaning of words, and results in symbols which are nearly equated with the linguists' concept of the morpheme, which has often enough been defined as the smallest linguistic unit possessed of meaning. (p. 93)

Chinese characters are sorted under morphemic systems in this classification. However, after suggesting that the term logographic should be reserved to refer only to the practice of supplementing basic morphemic signs with phonetic indicators, Hill states that Japan "has most fully developed this logographic use of a basically morphemic system by the addition of phonetic indicators" (p. 96).

2.4.2.5 Haas' (1976) classification

Haas' (1976) classification of writing systems is based on a set of three binary choices. The first choice is *derived-original*; pictographs are regarded as original because they do not correspond to speech in a regular way. The second choice is *empty-informed*; whether or not a graphic unit directly determines a meaning. The third choice is *motivated-arbitrary*; whether or not the relation between graphic unit and referent is pictorial.

These choices are logically independent, but not all of the combinatory possibilities are real. This is because an empty script, for example, cannot also be motivated, so this scheme actually only recognizes five kinds of script. Moreover, while the contrasts in this scheme are useful in differentiating types of pictorial representation, there is essentially only one contrast between other scripts; the extremely important contrast of empty-informed. Haas (1976, 1983) refers to scripts distinguished by this contrast as being either cenemic or pleremic. From the Greek word for 'empty,' in a cenemic writing systems, the graphic units only represent sounds and are, therefore, empty of semantic reference, such as alphabets and syllabaries. In contrast, from the Greek for 'full,' pleremic refers to writing systems where the graphic units are semantically informed denoting both sounds and meanings, such as kanji.

2.4.2.6 Halliday's (1985) classification

Defining 'writing' to mean "a system of visual representation that is language" (p. 14), Halliday's (1985) classification of kinds of writing systems is based on the level of linguistic unit represented, in which three kinds of script are distinguished; charactery, syllabary, and alphabet, as shown in Table 2.4.

Table 2.4

Halliday's (1985) Classification of Writing Systems

Level of language represented:	Lexico-grammatical (wording)	Phonological (sound)	
Linguistic unit represented:	word/morpheme	syllable	phoneme
Type of symbol:	character ('logogram')	syllabic sign	letter
Type of script:	charactery	syllabary	alphabet

Presumably as an analogy of syllabary, Halliday's (1985) appears to coin the term charactery (a system in which the symbols are characters) for the Chinese writing system to emphasize his view that these represent the words (or more accurately, the morphemes) of the language, for he states that the technical term for a character is a logogram.

2.4.2.7 Sampson's (1985) classification

In Sampson's (1985) classification of writing systems, which is shown in Figure 2.2, the first distinction made is between semasiographic and glottographic writing systems (see DeFrancis (1989), DeFrancis & Unger (1994) Sampson (1994) and Unger & DeFrancis (1995) for discussion of this).[21]

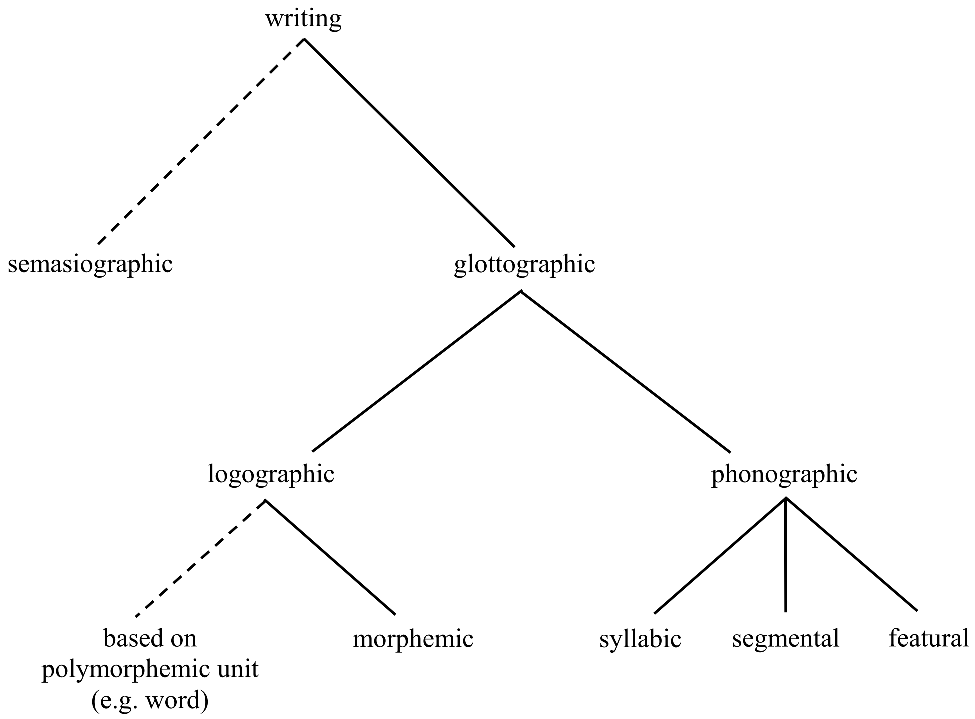


Figure 2.2. Sampson's (1985) classification of writing systems.

In the next level of this classification, glottographic is divided into logographic and phonographic. Under logographic, Sampson (1985) provides for what he considers to be a logical possibility, the polymorphemic unit, but uses a dotted line to indicate the fact that no such systems actually exist. Although the non-existence of systems based on polymorphemic units would seem to render the term logographic redundant in this classification, Sampson uses the term logographic, rather than morphemic, for Chinese characters.

Interestingly, Sampson (1985) also tackles the problem of the Japanese writing system of not fitting well within traditional classification schemes.

Sampson has elected to describe Japanese as a mixed writing system, and although Kaiser (1995) comments that this is not particularly informative, reminding us of Gelb's (1952) observation that there are no pure writing systems, in defense of his term, Sampson (1985) remarks,

> in some cases the different types of writing are mixed in such proportions within a given script that one cannot say which predominates. (p. 42)

This definitely applies to the multi-script nature of Japanese, and thus, the term 'mixed system' would seem to be well motivated. The aim of the present review of writing-system classifications and of the following section on the terminology issues is to determine exactly what types of script are being mixed in the case of the Japanese writing system.

2.4.2.8 DeFrancis' (1989) classification

DeFrancis' (1989) treatment of writing systems deserves our attention for at least two important reasons. The first is that DeFrancis is a Chinese specialist who has written at length about Chinese characters and the issues surrounding their classification (DeFrancis, 1984, 1989; DeFrancis & Unger, 1994; Unger & DeFrancis, 1995). The second reason is that DeFrancis (1989) provides perhaps the most extensive and extreme articulation of what I shall refer to as the 'language is speech' view. These points are taken up in more detail in Section 2.4.3 on the terminology issues, and for the moment, we note only the basic divisions in DeFrancis' classification, which is presented in Figure 2.3, and DeFrancis' term for Chinese characters.

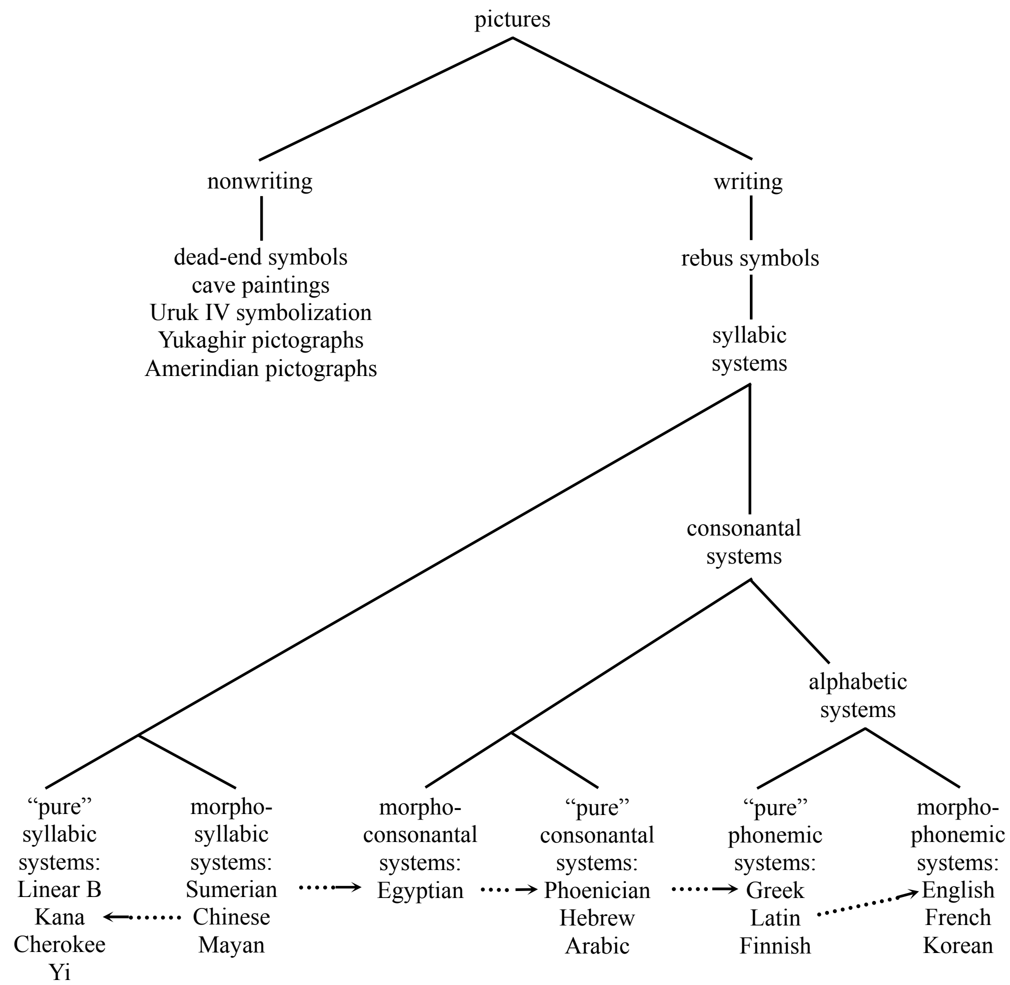


Figure 2.3. DeFrancis' (1989) classification of writing systems.

At the heart of DeFrancis' classification of writing systems is the dichotomy between what he refers to as partial and full writing systems,[22] and directly linked to this, DeFrancis' conviction in the phonetic basis of all full writing systems.

As noted above we shall return to discuss the issue in more detail later, although, in line with his belief that writing is simply the visual representation of speech, DeFrancis (1989) emphasizes the phonological aspect of Chinese characters, within the classification scheme DeFrancis opts to use the term morpho-syllabic.

2.4.2.9 Daniels' (1996, 2001) classification

Table 2.5

Daniels' (1996, 2001) Classification of Writing Systems[23]

Writing system		Explanation
1	Logosyllabary (morphosyllabary)	Each character stands for a word (or a morpheme)
2	Syllabary	Each character stands for a syllable
3	Abjad (Semitic-type script)	Each character stands for a consonant
4	Alphabet (Greek-type script)	Each character stands for a consonant or vowel
5	Abugida (Sanskit-type script)	Each character stands for a consonant accompanied by a particular vowel, with other vowels indicated by additions to the consonant symbol
6	Featural	Shapes of the characters correlate with phonetic features of designated segments

Note. Based on Daniels (1996, 2001).

As presented in Table 2.5, Daniels (1996, 2001) suggests that there are six fundamentally different kinds of writing systems.[24] While Daniel (2001) uses the term logosyllabary as the principal label for this category in his typology, he observes that the term morphosyllabic is more precise, commenting that,

> each character stands for a morpheme, and the characters can be used for the sound of the morpheme as well as for its meaning (p. 43).

2.4.2.10 Sproat's (2000) classification

Breaking with the conventional tree-format used in classifying writing systems, Sproat (2000) arranges writing systems according to two-dimensions, namely, the type of phonography and amount of logography involved in a system, as shown in Figure 2.4. Sproat's definition of logography is perhaps more inclusive than traditional definitions, for he regards,

> any component of a writing system as having a logographic function if it formally encodes a portion of nonphonological linguistic structure, whether it be a whole morpheme or merely some semantic portion of that morpheme (p. 134).

While Sproat refers to Chinese as a partly logographic writing system, based on this notion of logography, Sproat remarks that,

> Japanese apparently exhibits the most extensive use of logography of any modern writing system (p. 154).

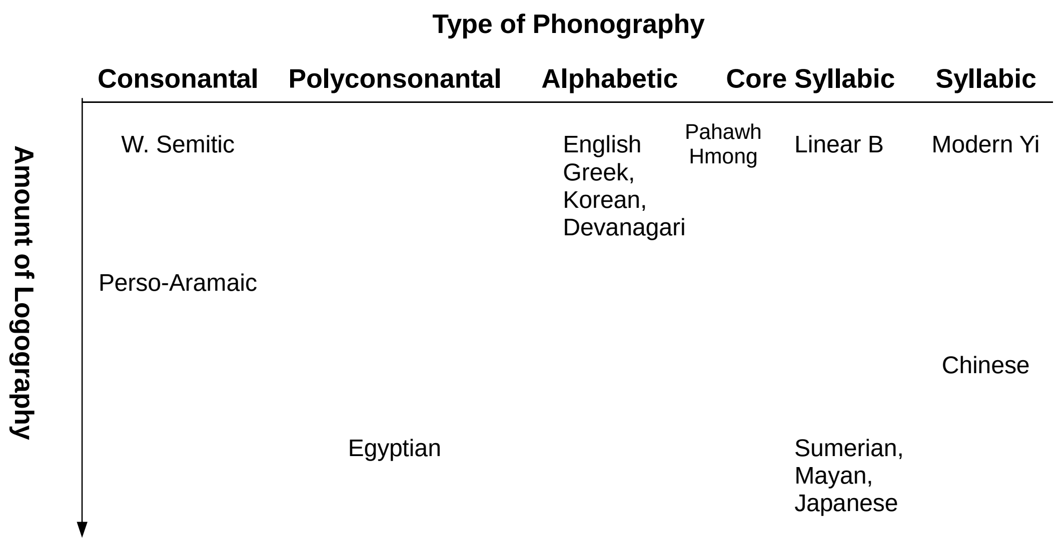


Figure 2.4. Sproat's (2000) non-arboreal classification of writing systems.

2.4.3 Terminology issues

Based on the short reviews of various classifications in the previous section, in this section we focus on the terminology that have been used in these various typologies, as shown in Table 2.6, and examine the precepts underlying these different terms. These terms, particularly, the loose usage by some of the terms 'pictographic,' 'ideographic,' and 'logographic,' have generated much heated debate (see, for example, DeFrancis (1984, 1989) for discussion of this in respect of Chinese, and Matsunaga (1996), Miller (1967, 1986) and Unger (1987) in respect of Japanese).

Table 2.6

Terms used for Kanji in Various Classifications of Writing Systems

Classification	Term
Taylor (1883)	Pictographs, ideograms and verbal phonograms
Gelb (1952)	Word-syllabic
Diringer (1962)	Ideographic (analytic transitional)
Hill (1967)	Morphemic
Halliday (1985)	Charactery (logogram)
Sampson (1985)	Logographic
DeFrancis (1989)	Morpho-syllabic
Daniels (1996, 2001)	Logosyllabic
Sproat (2000)	Chinese = Syllabic + medium level of logography Japanese = Core syllabic + high level of logography

While Hansen (1993) may feel the debate over terminology to be a "matter of truly mind-numbing triviality" (p. 376),[25] such comments clearly fail to fully appreciate the differences in the theoretical stances that the debate reflects.

2.4.3.1 Language and mediums of expression

How we approach the problem of classifying the world's writing systems will depend not only on how we perceive the creation and historical development of writing systems, but also on our notions about how speech and writing are related, and even on what we understand language to be (Henderson, 1982).

In this section, we examine some of the important issues involved in the classification of writing systems. The first is what to regard as writing. Although Gelb (1952), Diringer (1962), Haas (1973) and, speculatively, Sampson (1985) distinguish forms of semasiographic writing from full/proper/glottographic writing in their classifications, DeFrancis (1989) is undoubtedly correct to stress the importance of distinguishing between partial writing and full writing, which are defined as follows.

> Partial writing is a system of graphic symbols that can be used to convey only some thought.
> Full writing is a system of graphic symbols that can be used to convey any and all thought. (p. 3)

The second issue that a classification of writing system must consider, then, is how can graphic symbols convey "any and all thought," that is, how does writing relate to language. This question is, I would suggest, the single most important issue for understanding writing and how different types of writing systems function.

On the relation of writing to language, it is possible to discern two schools of thought. The first holds that language should be conceived of in terms of speech, and that writing is merely a means of transcribing speech, with the graphic unit being defined as primarily representing units of speech (Bloomfield, 1933; Daniels, 2001; DeFrancis, 1989; Miller, 1967, 1986; Sproat, 2000; Unger, 1987). As mentioned earlier, DeFrancis (1989) makes perhaps the strongest claims for this 'language is speech' position. At the risk of greatly oversimplifying DeFrancis' (1989) arguments, there are two main elements to his argument. The first element is the often-cited arguments for the primacy of speech over writing that state that speech exists in all human communities although writing does not, and linked to this, that while speech is naturally acquired, writing requires explicit instruction. In the 'language is speech' view, these facts are mistakenly interpreted to mean that sound is a defining attribute of language. The second element of DeFrancis' arguments is what he refers to as,

> the epoch-making invention that marked the birth of true writing. That invention was the *rebus principle*, whereby a pictographic symbol was used not for its original meaning value but specifically to represent the sound evoked by the name of the symbol. (p. 50)

To borrow DeFrancis' example, the rebus principle refers to using a picture of a *bee* and a picture of a *leaf* together to express the sounds of the word *belief*. While this is unquestionably the key to how the ancient inventions of writing, Sumerian, Mayan and Chinese, developed from partial writing systems to full writing systems, the application of this principle in the creation of phonetic compounds did not, however, contrary to what DeFrancis would have us believe, render Chinese characters as being a cenemic writing system.

In the alternative, and preferable, approach to language, sound is not regarded as being a defining feature of language, and while speech is undeniably the primary medium of expression, other mediums of expression—writing and sign—are not secondary to speech in terms of their relation to language (Coulmas, 1989; Crystal, 1987; Garman, 1990; Halliday, 1985; Lyons, 1981; Morioka, 1968; Sandler & Lillo-Martin, 2001; Steinberg, Nagata & Aline, 2001). Language should be seen as an abstract entity, where speech and writing, as well as sign, are different mediums for expressing language, which are linked not in a hierarchical relation but rather in terms of relations of inter-translatability. These two views are represented schematically in Figure 2.5.

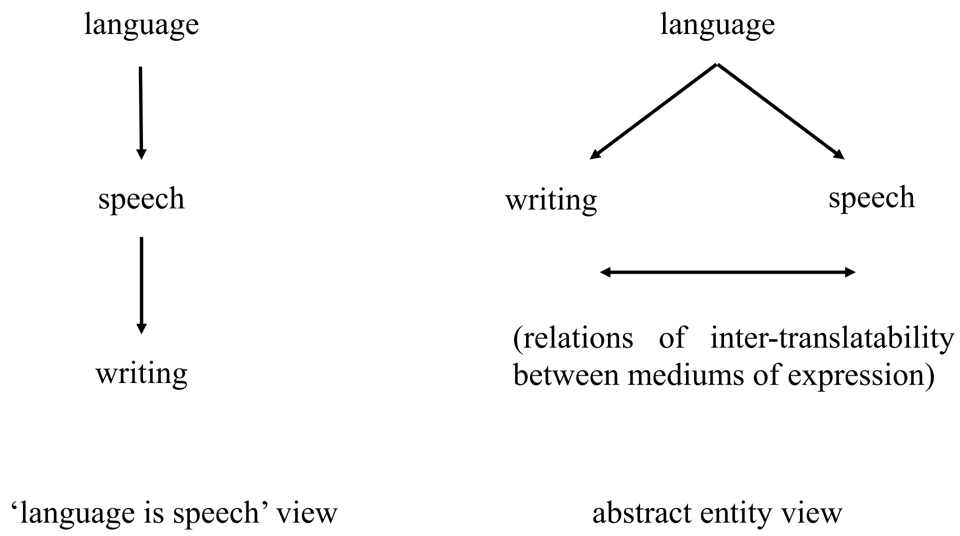


Figure 2.5. Schematic representation of the 'language is speech' view and the abstract entity view.

To clarify the differences in these views of language a little further, it is useful to briefly return to the two elements of DeFrancis' (1989) arguments. As mentioned, the conception of language solely in terms of speech stems from the facts that speech exists in all human communities while writing does not and that writing requires explicit instruction while speech is naturally acquired. The alternative position does not deny that speech is the primary medium of expression for hearing persons; it merely correctly recognizes that speech is not the only modality in which language can be given expression, and so does not privilege speech over writing or sign as mediums of expression.[26] Turning to the second point, the significance of the rebus principle for writing systems, DeFrancis is correct in describing this as the key to the development of partial writing systems into full writing systems. However, while the application of the rebus principle—using a graphic unit for its sound reference only—has in almost all instances eventually led to the creation of cenemic writing systems at either the syllabic level (such as kana) or the phonemic level (the alphabet), the way the principle was employed in the creation of the phonetic compound Chinese characters, as discussed in Section 2.4.1, did not lead to a cenemic writing system. This actually presents a real dilemma for the 'language is speech' position, for if language is speech and if writing is merely representing speech, then, what exactly is the non-phonological or semantic element of Chinese characters? Although DeFrancis claims not to deny the existence of what he refers to as the "secondary but nonetheless important nonphonetic, that is semantic or morphemic, aspect" (p. 58) of morphosyllabic systems (which includes Chinese in his classification), his response to this dilemma for the 'language is speech' position is to suggest that "Sampson, Haas, and others have seriously undervalued the phonetic element in Chinese writing" (p. 52) and to apparently go to the extreme in the opposite direction in overvaluing the phonetic element when he claims that Chinese writing is a 100 percent syllabic script.[27] The notion of a pleremic writing system is not, however, a problem for the abstract entity view of language. Because this approach recognizes that both speech and writing are mediums of expressing language and that these mediums are linked in relations of inter-translatability, it is able to offer a more flexible understanding of how both semantic and phonological elements can combine in Chinese characters.[28]

The final issue relating to the classification of writing systems that we need to consider is the level of linguistic units represented by pleremic writing systems. Actually, Hill (1967) has already provided us with the answer to this issue, for as he astutely observes there are no writing systems based on words. Because of the sheer number of symbols that would be required for a pure word writing system, we find that the only level that a writing system can function at above the syllabic level is the morphemic level. As Hill points out, in contrast to the phonological analysis of words in cenemic writing systems, a pleremic writing system entails analysis of word meaning, where it will settle on the morpheme, the smallest element of linguistic meaning.

From this brief discussion of the central issues in the classification of writing systems, we can see that while we must be careful to remember that there are no pure writing systems (Gelb, 1952), still, the terms that we apply in differentiating different systems should as far as possible seek to capture the complex relations between abstract linguistic elements, meanings, and the mediums of expression, namely, the orthographic and phonological units for different classes of systems. Once we recognize that speech and writing are functionally equal in terms of expressing language, the issue of how to classify a writing system becomes a matter of identifying the level of linguistic unit that the graphic units of the writing system principally represent. Thus, in response to the basic question addressed in this thesis of *how do kanji function as a writing system*, I suggest that the most appropriate answer is that *kanji are a morphographic writing system*—a writing system where the orthographic units represent morphemes, the minimal units of meaning in the language.

It is extremely regrettable that misconceptions about writing systems and kanji, reflected in terms like pictograph, ideograph, and logograph, which we consider next, continue to hinder a general appreciation of the morphographic nature of kanji.

2.4.3.2 Pictographs and the ideographic myth

As noted in our short review of important classifications of writing systems, a variety of typological labels that have been applied to kanji. The first of these is the term pictograph. Suffice to say, while the term may be used in reference to the limited numbers of simplex kanji that are derived from simple pictures of the physical objects they represent, these alone only constitute a partial writing system.

The second term is ideograph, which has probably created the most misunderstandings and the fiercest debate.[29] The arguments against the term ideograph being used for kanji is that it is simply misleading in suggesting that,

> unlike other forms of writing they [kanji] express an idea directly, and only by extension the sound associated with that idea. (Kaiser 1993a: p.59)

This is the line that Miller (1986) takes, for although he states that his objection is not with the terms, 'ideograph' or 'ideogram' themselves, but with their conventional use in referring to Chinese characters which he sees as "potentially extremely misleading" (p. 17), he is clearly focusing on a literal interpretation of the terms when he repeatedly insists that the characters are not 'graphs' for writing 'ideas' (p. 19).

While the use of the term 'ideographic' is generally accepted in a narrow sense to refer to the second group of simplex kanji, that represent simple concepts, it is extremely unfortunate that this term has been used in the past to refer to all kanji. Once again, the reason why the term is totally inappropriate to refer to all kanji is that the ideographic principle in the narrow sense alone could, simply, never underpin a full writing system (DeFrancis, 1989; Miller, 1986). As we have noted, to be a full writing system, a system must be capable of expressing the entirety of a language, and to be do this the graphic units of the system must be representing linguistic units, at either the phonemic level or the syllabic level in the cases of cenemic writing systems, or at the morphemic level in pleremic writing systems.

2.4.3.3 Logographic versus morphographic

Based on the discussion of the issues surrounding the classification of writing systems, I have proposed that the answer to the basic question of *how do kanji function as a writing system* is that *kanji are a morphographic writing system*.

The preference for the term morphographic over the more common term logographic may seem rather pedantic, especially given the wider inclusive definition which seems to be accorded to logographic recently,[30] and might be dismissed as merely a matter of emphasis. However, I believe, that would be a mistake for the shift in focus that this terminological revision requires is of fundamental significance for our understanding of kanji in the Japanese writing system, writing systems in general, the mental lexicon, and even language itself, as we shall see in Part 2 of this thesis which investigates the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon.

Some may claim that we need to retain the term logographic because many single kanji represent words, but the term morphographic is perfectly adequate to cover this fact, since by their very definition *free morphemes* are simple words. Seen in this light, compound labels such as logo/morphographic (Smith, 1996) can only appear repetitious, as Figure 2.6 shows.

logo "word"	+	*graph* "writing"	=	implies only words
morpho "morpheme"	+	*graph* "writing"	=	covers both free and bound morphemes

Figure 2.6. Logographic versus morphographic.

Others may object that the term morphographic is not completely correct, citing the existence of *jukujikun* and *ateji* words (this objection, of course, could also be levied at the term logographic). Although written with two or more kanji, these words must be considered as being monomorphemic words. However, rather than refuting the basic principle underlying the relationship between units of language and the majority of kanji, these minority groups of exceptions merely support the axiom that there are no pure writing systems (Gelb, 1952). As we have noted, writing systems may be classified as being semantically empty or semantically informed, but in reality, all writing systems include, to some extent, a mixture of the two orthographic principles. The task is to identify the dominant principle and to classify the writing system accordingly.

When kanji are used in writing Japanese words, in the vast majority of cases, they do not appear alone, but rather in combination with other graphic units, either with other kanji or hiragana, which are also representing morphemic elements of the Japanese words. It is because the term morphographic more accurately reflects and emphasizes this fact of kanji usage that it is preferable over logographic. As we shall note shortly in our discussion of Japanese word formation in the next chapter, the principles underlying the combination of kanji are not the provinces of orthography or phonological representation, but are rather the domain of morphology. Compounding is the combining of morphemes into larger word units, with the graphemes representing the union visually.

2.5 The Japanese writing system: Summary

The Japanese writing system is often portrayed as being extremely complex. Two factors that undoubtedly contribute to this are its multi-script nature and its dual-reading system. The present Japanese writing system is multi-script in nature, incorporating morphographic kanji, and two kana syllabaries, hiragana and katakana, as well as allowing for the use of the alphabet and Arabic numerals, which are employed in largely separate and complementary ways to represent the Japanese language in writing (Kabashima, 1977; Kaiser, 1993a, Kess & Miyamoto, 1999; Smith, 1996; Taylor & Taylor, 1995). In common with the vast majority of peoples, the Japanese did not invent writing for themselves, but rather borrowed writing from the Chinese. The two ways in which the Japanese came to use Chinese characters for writing their own vastly different language were to have far-reaching consequences leading to the dual on-reading and kun-reading system.

This chapter has argued that another factor for the portrayal of the Japanese writing system as extremely complex is the controversy surrounding the classification of kanji as a writing system, and the general failure to fully appreciate the morphographic nature of kanji. As Part 2 of this thesis, which investigate the lexical retrieval and representation of two-kanji compound words within the Japanese mental lexicon, argues the representation of morphology is fundamental to the organization of the mental lexicon.

Chapter 3

Japanese word formation: Two-kanji compound word morphology

This chapter provides an overview of Japanese word formation,[31] focusing in particular on the morphology of two-kanji compound words. Our interest in Japanese morphology follows directly from the central claim made in this thesis that kanji function as a morphographic writing system.

As mentioned in Chapter 1, the emphasis in this thesis on the two-kanji compound word rather than on the single kanji character follows from a belief that in classifying how a writing system actually functions it is essential to identity the principles that govern the combination of graphic units. The discussion of Japanese word formation and of the morphology of two-kanji compounds in this chapter is precisely about identifying the principles that underlie the combination of kanji in two-kanji compound words. The important insight that will emerge through this discussion is that in the relations between abstract linguistic elements, morphemes, and elements of expression—orthographic and phonological units, which in this context are primarily kanji and their readings—priority is at the abstract linguistic level, and the combination of elements is primarily morphologically motivated, with the representation of the resulting polymorphemic units being based on the representations of component elements. It is crucial that we do not confuse this relation between external form and internal representation. As noted at the outset of our discussions, the term two-kanji compound word is being used as a shorthand convention from an orthographic perspective for what are more accurately polymorphemic words formed by the combination of two morphemes, which are conventionally represented orthographically by the kanji associated with those morphemes. The claim that kanji function as a morphographic writing system is being made here precisely because the term morphographic most accurately reflects and emphasizes this basic reality—the fact that the principles underlying the combination of graphic units in two-kanji compound words are primarily morphological.

The discussion of two-kanji compound word morphology also serves to provide background to the selection of representative word formation principles for two-kanji compound words used as experimental conditions in the priming experiments that will be presented in Part 2 of the thesis, which were conducted to investigate the lexical retrieval and representation of two-kanji compounds in the Japanese mental lexicon. While the main emphasis of the thesis is on two-kanji compound words, many of the same structures discussed in this chapter are seen in other polymorphemic words.

3.1 The Japanese lexicon

Japanese word formation processes involve lexical items from all lexical stratums of the Japanese language. Awareness of lexical origin can often clarify the complex interaction between word structure and readings for kanji compound words. We shall, therefore, start our discussions with a few words concerning the various lexical stratums. In this section, we shall also refer to a structural classification of the Japanese lexicon, distinguishing compound words from other complex words, before taking up the discussion of compound words in some detail in the next section.

3.1.1 Lexical stratum

One common way of classifying the Japanese lexicon is according to the origin of the words (Shibatani, 1990; Kaiser, 1993a; Fujii, 1996; Tsujimura, 1996). Words are divided into three basic groups: 和語 *wago*, or 大和言葉 *yamatokotoba*, native Japanese words, 漢語 *kango* Sino-Japanese words, and 外来語 *gairaigo* foreign loanwords, which have entered Japanese from foreign languages. A further group of 混種語 *konshugo* hybrid compounds, represents combinations of elements from the other groups. A subdivision of native Japanese words can also be made for 擬態語・擬声語 *gitaigo/giseigo* phonaestetic words, which include onomatopoeia expressions (Kaiser, 1993a). Technically, Sino-Japanese words are also foreign loanwords. However, they are generally distinguished from loanwords from Western languages, reflecting the fact that they are so ingrained into the Japanese lexicon and are written in kanji (Morioka, 1968).

Many of the most frequently used words in the Japanese language are native Japanese words. For instance, all grammatical words, such as particles and auxiliary verbs, are native Japanese words. These are written in kana. When native Japanese words are represented by kanji, they are read according to the kun-reading, e.g., 山 *yama* ‘mountain.’ One problem with this classification is that, unfortunately, some words generally thought to be native Japanese words are actually not. For instance, 馬 *uma* 'horse,' 梅 *ume* 'plum,' were borrowed from Chinese (Shibatani, 1990), and 寺 *tera* 'temple' and 靴 *kutsu* 'shoes' were borrowed from Korean (Kaiser, 1993a).

The status of Sino-Japanese words is also problematic, for this stratum includes not only items borrowed in ancient times, but also more recent borrowings, such as 哲学 *tetsugaku* 'philosophy' and 科学 *kagaku* 'science.' The category also includes items of Japanese coinage, such as words now normally read according to on-readings for native Japanese words transcribed in kanji, such as 大根 *daikon* 'large radish' based on *ône* and 返事 *henji* 'reply' based on *kaerigoto*.

Written in katakana, foreign loan words are words that have been borrowed into Japanese, undergoing modifications to conform to Japanese phonological structure (Kaiser, 1993b). As Fujii (1996) observes, because this process involves syllabification of the foreign words, the loan words often become quite long, as in パーソナルコンピューター *paasonaru konpyuutaa* 'personal computer' and エアーコンディショナー *eaa kondishonaa* 'air conditioner,' which, then, are often abbreviated to four mora length, as in パソコン *pasokon* and エアコン *eakon*, respectively (p. 93).

Hybrid words involving the combination of native Japanese elements with Sino-Japanese ones give rise to combinations of on-readings and kun-readings. An example of an on-reading plus a kun-reading is 書棚 *shodana* 'bookshelf.' This combination of readings is known as 重箱読み *jūbako yomi*, after another example meaning 'nested boxes.' In contrast, an example of kun-reading plus on-reading is 手製 *tesei* 'handmade.' Combinations of this sort are, again after an example of a word meaning 'bathtub,' called 湯桶読み

yutō yomi (Fujii, 1996).

Because of lexical borrowing for Sino-Japanese and other foreign loan words, words of similar meanings exist at different lexical stratums in Japanese, as the examples in Table 3.1 indicate. The impressions from native Japanese words are more common, from Sino-Japanese words more formal, and from foreign loans more modern.

Table 3.1

Examples of Similar Words at Different Lexical Stratums

Lexical Stratum		
Native Japanese 和語	Sino-Japanese 漢語	Foreign Loans 外来語
宿屋 *yadoya* inn	旅館 *ryokan* Japanese-style hotel	ホテル *hoteru* hotel
取り消し *torikeshi* cancellation (general)	解約 *kaiyaku* cancellation (contracts)	キャンセル *kyanseru* cancellation (bookings)
開く *hiraku* open (a shop)	開店する *kaiten suru* open a shop	オープンする *ôpun suru* open a shop

3.1.2 Simplex and complex words

Another way of classifying the Japanese lexicon is according to word structure. Tamamura (1985) provides such a classification,[32] a slightly modified version of which is given in Figure 3.1.

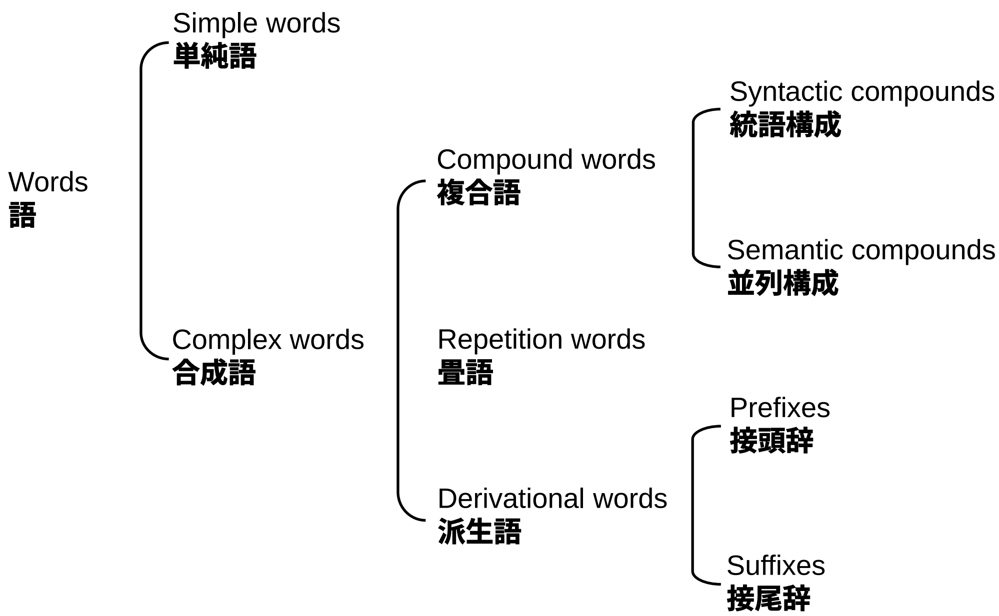


Figure 3.1. Structural classification of Japanese words (based on Tamamura, 1985, p. 9).

In this classification, the first division is between 単純語 *tanjungo* simple words and 合成語 *gōseigo* complex words, which are formed by combinations of free and bound morphemes. Complex words are further subdivided into three groups; 複合語 *fukugōgo* compound words, 畳語 *jōgo* repetition words, and 派生語 *haseigo* derivative words.

Compound words, which will be discussed in some detail in the next section, are in turn divided into 統語構成 *tōgo kôsei* syntactic compounds and 並列構成 *heiretsu kōsei* semantic compounds. Tamamura (1985) points out, however, that the terms complex words and compound words are often used interchangeably, and, indeed, Nomura's (1988a. 1988b) classification of the combination patterns underlying two-kanji compound words, which we shall consider shortly, includes repetition words and derivatives.

Repetition words express plurality and emphasis. For instance, repeating 人 *hito* 'person' gives 人人 *hitobito* 'people' normally written as 人々 with a special character to represent repetition, and repeating 黙 *moku* 'be silent' gives 黙々 *mokumoku* 'silence' (Kuratani, Kobayashi, & Okunishi, 1982; Habein & Mathias 1991). It is also possible to have adjective repetition words with inflectional endings, such as 騒々しい *sōzōshii* 'noisy.'

Derivation compounds are also further subdivided into prefixes and suffixes. Both Shibatani (1990) and Kageyama (1982) discuss affixation in some detail, observing that affixes can be distinguished in terms of lexical stratum and productivity, with Sino-Japanese affixes being generally are more productive than native Japanese ones. Common prefixes include 不 *fu* negation form 'un-,' and 本 *hon* 'this,' and the honorific form お *o*. Common suffixes include 的 *teki* adjectival form '-ic,' 化 *ka* verbalizing form '-ize,' and さ *sa* the adjective nominalizer.

3.2 Japanese compound words

As already indicated, compounding is an extremely productive process of word formation in Japanese, involving elements from all word classes, and all lexical stratums (Kageyama, 1999; Shibatani, 1990). In this section, we shall look first at syntactic compound formations, and then at semantic compound formations.

3.2.1 Syntactic compounds

Tamamura (1985) organizes his discussion of compounding primarily according to the word classes of the component elements and the word class of the compound, which is determined by the class of the final element. We shall, however, break with this to a certain extent, in order to emphasize the syntactic relationships between the elements, distinguishing the three relationships of modifier + modified, verb + complement, and complement + verb.

3.2.1.1 Modifier + modified

Conforming to the syntax of modification, with modifiers preceding modified nouns, these compounds have a right-hand head structure (Kageyama, 1982).

Table 3.2.

Examples of Adjective + Noun and Verb + Noun Modifier + Modified Compounds

Relationship	Compound words		
Adjective + Noun			
	古本	*furuhon*	old book
	高値	*takane*	high price
Verb + Noun			
	食べ物	*tabemono*	food
	泣き声	*nakigoe*	crying voice

Table 3.3.

Examples of Noun + Noun Modifier + Modified Compounds

Relationship	Compound words		
Instance	母親	*hahaoya*	mother
Subject	交通渋滞	*kōtsūjūtai*	traffic congestion
	地盤沈下	*jibanchinka*	subsidence
Object	歯ブラシ	*haburashi*	toothbrush
	ガラス工場	*garasukōjō*	glass factory
Means	トラック輸送	*torakkuyusô*	truck transportation
Material	ビニール袋	*biniirubukuro*	plastic bag
	石橋	*ishibashi*	stone bridge
Cause	事故死	*jikoshi*	accidental death
Location	山桜	*yamazakura*	mountain cherry
	裏門	*uramon*	rear gate
Time	朝ご飯	*asagohan*	breakfast
	秋空	*akizora*	autumn sky
Quality	親心	*oyagokoro*	parental love
	勇気	*yūki*	courage

Lexicalized examples of adjectival and verbal modification, as shown in Table 3.2, where inflectional endings are dropped, such as 高値 *takane* 'high cost' from 高い値, are extremely common (Okutsu, 1975; Tamamura, 1985). However, the majority of these compounds are noun + noun compounds, which express a variety of relationships, such as object, means, material, location, and time, as shown in Table 3.3.

3.2.1.2 Verb + complement

The syntactic structure of verb + complement is Chinese in origin rather than being native to Japanese, and accordingly only Sino-Japanese elements are involved. These compounds have left-hand heads, and are treated as verbal nouns, forming verbs with the dummy verb する *suru* 'do' (Kageyama, 1982; Shibatani, 1990). Examples are given in Table 3.4.

Table 3.4

Examples of Verb + Complement Compounds

Relationship	Compound words		
Intrans. Subj.	出火	*shukka*	outbreak of fire
	停電	*teiden*	power failure
Direct Object	洗顔	*sengan*	washing one's face
	駐車	*chūsha*	parking a car
Goal	帰国	*kikoku*	returning to one's country
	着陸	*chakuriku*	landing
Location	在宅	*zaitaku*	staying home
	滞京	*taikyō*	staying in capital
Source	離日	*rinichi*	leaving Japan
	落馬	*rakuba*	falling from horseback

There appear to be syntactic restrictions with this pattern, for, as Kageyama (1982) observes, there is an absence of indirect objects and transitive subjects as complements, with the majority of these Sino-Japanese compounds being combinations of verb + direct object. As we shall see in the next section, Sino-Japanese words with predicative subject, indirect object, and adverbial complements have the syntactic structure of complement + verb.

3.2.1.3 Complement + verb

In contrast to the verb + complement structure, Sino-Japanese words with a complement + verb structure involve examples of subject + predicate, indirect object complements, and adverb complements (Ozaki, Todome, Nishioka, Yamada, & Yamada, 1992), as shown in Table 3.5.

Table 3.5

Examples of Sino-Japanese Complement + Verb Compounds

Relationship	Compound words		
Subject + predicate	地震	*jishin*	earthquake
	日没	*nichibotsu*	sunset
Indirect object	毒殺	*dokusatsu*	kill by poisoning
	郵送	*yūsō*	send by mail
Adverb	再会	*saikai*	meet again, reunion
	独奏	*dokusō*	play a solo

This order of elements is also a native Japanese syntactic structure, and accordingly, there are many native Japanese compound words of this form. According to the form of the final element, it is possible to distinguish between two kinds of compounds possessing this structure of complement + verb structure. The first are compound verbs, such as 気付く *kizuku* 'notice, realize,' although these will not be discussed further here (see Kageyama, 1982, 1999; Tamamura, 1985; also Tagashira & Hoff, 1986). The second are compound nouns, such as 山歩き *yamaaruki* 'mountain walking,' lexicalized from the sentence 山を歩く *yama o aruku* 'to walk in the mountains.' Examples of Native Japanese compounds with this structure are provided in Table 3.6, with semantic relationships indicated.

Although these compounds can be associated with syntactic structures relatively easily, Kageyama (1982) argues that given the rather unpredictable and wide range of the referents of these compounds, they should be interpreted as being formed in the lexicon. Tamamura (1985) provides an example that illustrates this problem. The compound 酒飲み *sakenomi* '(alcohol) drinker' can be linked to the syntactic structure 酒を飲む *sake o nomu* 'to drink alcohol,' here referring to a person who drinks alcohol. However, although a similar structure of 湯を飲む *yu o nomu* 'to drink hot water (liquids)' can be said to underlie 湯飲み *yunomi* 'teacup, cup,' here the referent is a teacup with which to drink. It is therefore possible to classify these compounds according to the referent, examples of which are in Table 3.7.

Table 3.6

Examples of Native Japanese Complement + Verb Compounds

Relationship	Compound words		
Subject	値上がり	*neagari*	price increases
	雨降り	*amefuri*	rain-fall
Object	子守り	*komori*	baby-sitter
	人殺し	*hitogoroshi*	murder
Movement	家出	*iede*	leave home
	谷渡り	*taniwatari*	crossing valley to valley
Instrumental	砂遊び	*sunaasobi*	playing with sand
	鉄板焼き	*teppanyaki*	meat cooked on hot plate
Destination	里帰り	*satogaeri*	home-coming
	外国行き	*gaikokuyuki*	foreign-bound
Location	沖釣り	*okizuri*	offshore fishing
Source	天下り	*amakudari*	appointed from above
Collaborative	人付き合い	*hitozukiai*	mixing with people
Standard	親勝り	*oyamasari*	surpassing one’s parents
Cause	霜枯れ	*shimogare*	withered by frost
	日焼け	*hiyake*	sunburn, suntan
Direction	南向き	*minamimuki*	south facing
Qualification	客扱い	*kyakuatsukai*	receiving as guest

Table 3.7

Examples of Referents of Native Japanese Complement + Verb Compounds

Referent	Compound words		
Activity	墓参り	*hakamairi*	visiting grave
	草取り	*kusatori*	weeding
Result	野菜炒め	*yasaiitame*	fried vegetables
	水溜まり	*mizutamari*	puddle
Person	嘘吐き	*usotsuki*	liar
	髪結い	*kamiyui*	hairdresser
Tool	缶切り	*kankiri*	can-opener
	筆入れ	*fudeire*	brush (pen) case
Place	手洗い	*tearai*	toilet
	車寄せ	*kurumayose*	car porch
State	二日酔い	*futsukayoi*	hangover
Time	夜明け	*yoake*	dawn
	日暮れ	*higure*	dusk

The existence of both verb + complement compounds and complement + verb compounds occasionally leads to compounds of similar meanings. For instance, 殺人 *satsujin* and 人殺し *hitogoroshi* both mean 'murder;' 止血 *shiketsu* and 血止め *chidome* both mean 'stop bleeding;' and 投球 *tōkyū* and ボール投げ *boorunage* both mean 'throwing a ball.'

3.2.2 Semantic compounds

A characteristic of semantic compounds is that the elements are of the same word class, and these can be divided into associative and synonymous pairs depending on the relationship between the elements.

3.2.2.1 Associative pairs

Kageyama (1982), Shibatani (1990) and Tsujimura (1996) referring to these compounds as dvandva compounds, emphasize that each element is an independent reference. For example, 親子 *oyako* may be taken as denoting any combination of parent(s) and child(ren), with both elements having equal status.

Table 3.8

Examples of Associative Pairs

Lexical stratum	Compound words		
Native Japanese	あちこち	*achikochi*	here and there
	売り買い	*urikai*	buying and selling
	好き嫌い	*sukikirai*	like or dislike
Sino-Japanese	出欠	*shukketsu*	presence or absence
	苦楽	*kuraku*	pain and pleasure
	老若	*rōnyaku*	old and young
Foreign	プラスマイナス	*purasu mainasu*	plus or minus

The conjunction of these elements may also be inclusive or exclusive, i.e., linked by 'and' or by 'or,' as the examples in Table 3.8 illustrate. Kageyama (1982) also observes that the order of the elements in these compounds is fixed, and in many cases, there is a priority for positive elements to be first, such as 正否 *seihi* 'right or wrong,' 有無 *umu* 'existence-nonexistence.' However, there are exceptions to this, as in 貧富 *hinpu* 'poor and rich, and 難易 *nan'i* 'difficult and easy.'

3.2.2.2 Synonymous pairs

Table 3.9

Examples of Synonymous Pairs

Word class	Compound words		
Noun + Noun	道路	*dōro*	road
	山岳	*sangaku*	mountains
	河川	*kasen*	rivers
Adj. + Adj.	詳細	*shōsai*	detailed
	善良	*zenryō*	good
	美麗	*birei*	beautiful
Verb + Verb	変化	*henka*	change
	増加	*zōka*	increase
	死亡	*shibō*	die

In contrast to associative pairs, synonymous pairs involve elements with similar meanings. Tamamura (1988) suggests that the two kinds of pairings are similar for they are both combining elements from a semantic category, the difference being that associative pairs are combinations of examples from opposing extremes. Table 3.9 gives examples of synonymous pairs.

3.3 The morphology of two-kanji compound words

Following on from the brief discussion of Japanese word formation in the last section, here the focus will be on the structure of two-kanji compound words. Specifically, this section will introduce a classification of the combination patterns underlying two-kanji compound words proposed by Nomura (1988a. 1988b). Nomura (1988a. 1988b) takes a more inclusive interpretation of the term compound words, for the classification also covers other complex words, such as repetition, derivation, and abbreviation words.

Nomura (1988a) discusses some of the difficulties of analyzing two-kanji compounds, suggesting that perhaps the biggest problem is declining awareness of Sino-Japanese structure, which gives rise to tautological expressions such as *犯罪を犯す *hanzai o okasu* ‘to commit committing a crime.’ An additional problem, as mentioned earlier, relates to the origin of many Sino-Japanese words. One source of confusion here is the existence of 和製漢語 *wasei kango* Japanese-made Sino-Japanese words, such as 工業 *kōgyō* ‘manufacturing industry’ and 野球 *yakyū* ‘baseball.’

Table 3.10

Examples of Orthographic Modifications

Modification Principle			Compound words		
Same etymology					
	廻転	→	回転	*kaiten*	rotation, revolve
	吃水	→	喫水	*kissui*	draft of a ship
Common forms					
	火焔	→	火炎	*kaen*	flames
	史蹟	→	史跡	*shiseki*	historic ruins
Similar meaning					
	衣裳	→	衣装	*ishō*	clothes
	蒐集	→	収集	*shūshū*	collection
New formed words					
	根柢	→	根底	*kontei*	root, basis
Phonetic based					
	一挺	→	一丁	*itchō*	one block
	庖丁	→	包丁	*hōchō*	kitchen knife

Another problem is the existence of 音訳語 *onyakugo* phonetic loan words. Jukujukun words are actually examples of phonetic loans into Chinese and from there into Japanese, such as 刹那 *setsuna* 'moment, instant' and 檀那 *danna* 'master' from Sanskrit. More recent transliterations from European languages also exist, such as 護謨 *gomu* 'rubber' and 瓦斯 *gasu* 'gas.' The situation is further complicated by ateji formations, which Nomura (1988a) suggests has been accelerated by orthographic modifications as part of the simplification process within the script reforms. Examples of such changes are given in Table 3.10.

Table 3.11

Nomura's (1988a) Classification of Word Bases and Affixes

Constituent Type	Examples
Word Bases	
Uninflecting word bases (i.e., nouns (N)) 体言類語基 *taigenrui goki*	客 *kyaku* 'guest' 駅 *eki* 'station'
Inflecting word bases (i.e., adj. (A)) 相言類語基 *sōgenrui goki*	急 *kyû* 'quick' 新 *shin* 'new'
Declinable word bases (i.e., verbs (V) 用言類語基 *yōgorui goki*	帰 *ki* 'return' 集 *shû* 'collect; gather'
Modifiers (M) 副言類語基 *fukugenrui goki*	最 *sai* 'most' 再 *sai* 'again'
Affixes	
Prefixes 接頭辞 *settōji*	不 *fu* 'un-' 可 *ka* 'possible'
Suffixes 接尾辞 *setsubiji*	的 *teki* '-ic' 然 *zen* 'state of –'

In view of such problems, and after discussing some previous attempts to classify two-kanji compound words, Nomura (1988a) proposes nine combination patterns for their formation. Defining the constituent elements as 字音形態素 *jion keitaiso* Sino-Japanese morphemes,[33] Nomura also distinguishes between four kinds of 結合専用語基 *ketsugō senyō goki* bound word bases[34] and two kinds of affixes, which are detailed in Table 3.11.

Nomura's (1988a) classification of two-kanji compound words is based on the patterns in which these elements are combined, as shown in Table 3.12 shown over the next two pages.

An important feature of Nomura's (1988a) concise classification is his separation of problematic items like jukujikun and ateji words as phonetic borrowings. Another is the fact that this classification eliminates the need to specify the underlying syntactic structure in terms of stipulating case markers for complements, such as subject marker が *ga* or object を *o*.

However, although the modification (1) pattern of verbal modification appears well motivated in contrast to modification (2) of nominal modification, the division of verbal patterns into those of modification and those with complements would seem to be at the expense of differentiating between the two syntactic patterns of verb + complement and complement + verb, with both included under the complement pattern. This point will be taken up again when discussing the selection of representative word formation principles for two-kanji compound words as experimental conditions in the experiment presented in this thesis, which were conducted to investigate the lexical retrieval and representation of two-kanji compounds in the Japanese mental lexicon.

Table 3.12

Nomura's (1988a) Classification of the Combination Patterns for Two-Kanji Compound Words

Combination Pattern	Compound words		
1 Complements 補足 *hosoku*			
(1) [N] + [A]	胃弱	*ijyaku*	weak stomach, indigestion
(2) [A] + [N]	多才	*tasai*	talented
(3) [N] + [V]	地震	*jishin*	earthquake
	肉食	*nikushoku*	eat meat
(4) [V] + [N]	降雨	*kōu*	rainfall
(5) [V] + [N]	登山	*tozan*	mountain climbing
2 Modification (1) 修飾 *shūshoku* (1)			
(1) [A] + [V]	細分	*saibun*	divide finely
(2) [V] + [V]	代弁	*daiben*	act/ speak by proxy
(3) [M] + [V]	再発	*saihatsu*	re-issue
(4) [M] + [A]	最高	*saikō*	highest
3 Modification (2) *shūshoku* (2)			
(1) [A] + [N]	幼児	*yōji*	young child, infant
(2) [V] + [N]	祝日	*shukujitsu*	celebration day, holiday
(3) [N] + [N]	牛乳	*gyūnyū*	cow's milk

table continues

Combination Pattern	Compound words		
4 Synonymous pairs 並列 *heiretsu*			
(1) [N] + [N]	身体	*shintai*	body
(2) [A] + [A]	永久	*eikyū*	long time
(3) [V] + [V]	破壊	*hakai*	destroy
5 Opposites 対立 *tairitsu*			
(1) [N] + [N]	左右	*sayū*	left + right
(2) [A] + [A]	強弱	*kyōjyaku*	strong + weak
(3) [V] + [V]	売買	*baibai*	buying + selling
6 Repetitions 重複 *jūfuku*			
[ ] + [ ]	段々	*dandan*	gradually
7 Auxiliary 補助 *hojo*			
(1) [X] + [ ]	不明	*fumei*	unclear
(2) [X] + [ ]	可能	*kanō*	possible
(3) [ ] + [X]	史的	*shiteki*	historic
8 Abbreviations 省略 *shōryaku*			
医学大学	医大	*idai*	university of medicine
9 Phonetic borrowings 音借 *onsha*			
(1) [ ] + [ ]	葡萄	*budō*	grapes
(2) [ ] + [ ]	面倒	*mendō*	care

3.3 Japanese word formation: Summary

Our interest in Japanese morphology has followed directly from the central claim being made in this thesis that kanji function as a morphographic writing system.

This chapter's discussions of Japanese word formation processes, and, in particular, of the morphology of two-kanji compound words have sought to demonstrate the morphographic nature of kanji. As Table 3.13 shows, the principles that underlie the formation of two-kanji compound words are primarily morphological.

Table 3.13

Word-Formation Principles Underlying Two-Kanji Compound Words[35]

	Principle	Examples			Morphological
1	Modifier + modified	山桜	国道	高値	Yes
2	Verb + complement	登山	殺人	投球	Yes
3	Complement + verb	外食	毒殺	夜勤	Yes
4	Associative pairs	親子	生死	左右	Yes
5	Synonymous pairs	山岳	苦痛	変化	Yes
6	Repetitions	段々	個々	黙々	Yes
7	Derivation	不明	以上	史的	Yes
8	Abbreviations	農協	教祖	春闘	Yes
9	Phonetic borrowing	葡萄	阿片	面倒	No

As only one of the word-formation principles for two-kanji compound words is not morphologically based,[36] clearly, in the vast majority of cases, compounding should be seen as the combining of morphemes to form larger word units, with the representation of such words being a product of combining the constituent representations.

While the arguments being made here may appear to be somewhat superfluous, the basic point is of fundamental importance to our understanding of kanji, writing systems and even of writing and language. The recognition that kanji function as a morphographic writing system is also going to have profound implications for our understanding of the Japanese mental lexicon. The objective of Part 2 of this thesis is to examine these implications by examining the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon from a morphological perspective.

Part 2

The Japanese mental lexicon

Chapter 4

Visual word recognition: Tasks, effects, issues and models

The shift in perspective from the linguistic description of external form in Part 1 to the psycholinguistic investigation of internal representation in Part 2 of this thesis closely parallels the transition from printed word to internal representation in the process of lexical retrieval, which is the central concern of visual word recognition and mental lexicon research.

In transferring our attention onto the internal representation, however, we must be careful to guard against taking the "short and innocent journey" from linguistic description to asserting a psychological doctrine that Henderson (1982, p. 7) warns of, such as insisting on simplistic and misleading dichotomies derived from the notions embodied in terms like logographic and phonographic.[37]

We must also be vigil in watching for those who would make the 'return journey' of misconstruing findings from the visual word recognition literature as providing indisputable evidence for distinctions within classifications of writing systems. DeFrancis and Unger (1994), Sproat (2000), Unger and DeFrancis (1995), for example, appear to be making this journey when they cite the findings from a study by Horodeck (1987) in support of their claims for the phonological basis of kanji (see also Matsunaga (1996), and Hayes (1988) for similar study in Chinese). Horodeck's (1987) reports of both higher writing errors and higher errors in detection involving homophonic kanji are described by Unger and DeFrancis (1995) as,

> impressive statistical evidence that kanji are primarily associated with phonological strings and only secondarily with categories of shape or "meaning." (p. 54)

While these findings are very interesting, there are limits as to what they prove.[38] Such findings only indicate that phonology is co-activated within the process of lexical retrieval (just as we would expect given the relations of inter-translatability between orthography and phonology and the primacy of speech as a medium of expression for hearing persons); they certainly do not show that lexical retrieval is necessarily mediated via phonological activation. The questions of whether lexical retrieval is mediated by indirect, phonological activation and whether direct, semantic activation is possible based on visual or orthographic information are some of the most hotly debated issues within visual word recognition research (Baron, 1973; Coltheart, Daveleer, Jonasson, & Besner, 1977; Seidenberg, 1990; Shen & Forster, 1999; Van Orden, 1987).[39] Van Orden (1987), for instance, has interpreted higher errors for English homophonic foils in a semantic categorization task as evidence for phonological mediation. On the other hand, because Shen and Forster (1999) found orthographic priming effects in both masked-priming lexical decision and naming tasks, but only phonological priming in the naming task which explicitly emphasizes phonological processing, they argue that lexical retrieval of Chinese characters does not depend on prior activation of phonological information.[40]

As background to our investigation and discussion of the Japanese mental lexicon in Chapters 5 and 6, this chapter briefly reviews psycholinguistic research into visual word recognition and the nature of the mental lexicon.

Firstly, the chapter introduces the two most commonly used tasks in visual word recognition, the lexical decision task and the naming task, as these are also the tasks that are employed in the experiments to be reported later. Secondly, the chapter introduces some of the main effects that have been identified within visual word recognition research. These effects are extremely important not only for their implications in the design of visual word recognition experiments, but because they provide important insights into the nature of the mental lexicon. As such, all models of the mental lexicon must take these effects into consideration. And finally, the chapter discusses the central issues of lexical retrieval and lexical representation by looking at how some of the most influential models of the mental lexicon account for these issues.

4.1 Tasks

While a number of tasks are commonly used in visual word recognition research, such as reading tasks for eye-movement tracking, tachistoscopic identification tasks, semantic judgment or categorization tasks, and proofreading tasks to mention just a few (Harley 2001), the most commonly employed tasks are the lexical decision task and the naming task (Balota, 1994; Neely, 1991; Shen & Forster, 1999; Taft, 1991). As these are also the tasks that were used in the experiments to investigate the lexical retrieval and representation of two-kanji compound words which will be presented later, these tasks and some of the issues associated with them are briefly outlined in the following two sub-sections.

4.1.1 Lexical decision

Essentially, the lexical decision task asks participants to judge whether a presented letter string is a word or not, and to respond by pressing a ‘yes’ button for words and a ‘no’ button for non-words. The logic behind this task is that as long as the non-words conform to orthographic rules, participants must consult their mental lexicons in order to judge whether a letter string is a word or not. This is why the task is often preferred over the naming task, where lexical access is arguably not necessary (Coltheart, Daveleer, Jonasson, & Besner, 1977; Taft, 1991). Differences in reaction times and error rates in making the lexical decision responses are taken as evidence of differences in the ease of making the decisions. As Harley (2001) points out, however, experimenters should be aware of the speed-accuracy problem associated with this task. It is necessary to emphasize to participants the importance of both speed and accuracy, because when encouraged only to respond quickly, participants will tend to make more errors in responding, whereas if accuracy is stressed, then responses will be tend to be slower.

4.1.2 Naming

In the naming task, participants are typically presented with a visual word that they are required to name—to read aloud. The latency to initiate the pronunciation of the word is taken as a measure of how long it takes to access the word and retrieve the pronunciation (Harley, 2001; Seidenberg, 1990; Taft, 1991). There are, however, two issues relating to the naming task that need to be considered. The first problem for the naming task is the claim that because it is possible to pronounce legal non-words, which are obviously not in the mental lexicon, it could be possible to pronounce regular words without recourse to lexical access. For instance, by simply converting letters to sounds it would be possible to pronounce the word C-A-T, rather than having to rely on lexical information (Coltheart, Daveleer, Jonasson, & Besner, 1977; Seidenberg, 1990; Taft, 1991). A second issue for the naming task is the technical issue of how to measure the reaction times. Although naming tasks usually measure naming latencies with a voice key that triggers a timer when the sound amplitude exceeds a preset level, there is evidence to suggest that voice keys are not always reliable (Kello and Kawamoto, 1998; Sakuma, Fushimi, & Tatsumi, 1997)—an issue that we shall discuss further in Chapter 6 when presenting the two naming experiments conducted for the present research.

Although these two tasks have yielded important findings concerning the effects influencing visual word recognition, which we discuss in the next section, as Balota (1994) comments, task analysis studies suggest that researchers should exercise caution in attributing all word-frequency effects to access processes in the tasks used to measure word recognition. This is because there is evidence to suggest that effects may be exaggerated by the decision component in the case of the lexical decision task, and by post-access components related to the generation and output of phonological code in the case of the naming task. Certainly, the findings from the lexical decision task and the naming task have not always been consistent, which in itself requires some explanation (Henderson, 1982; Neely, 1991; Taft, 1991).

4.2 Effects

In this section, we briefly introduce some of the more important findings obtained in visual word recognition research that have shaped the models that attempt to explain the mental lexicon (Neely, 1991; Taft, 1991).

4.2.1 Word frequency and word familiarity effects

Word frequency and word familiarity effects—the findings that high frequency or familiar words are recognized quicker and more accurately than low frequency or unfamiliar words—are among the most robust effects within visual word recognition research. Rubenstein, Garfield, and Millikan (1970) provided an early demonstration of the word frequency effect. In their lexical decision task experiment, English words were divided according to word frequency statistics, and according to whether or not they were homographic words (same orthographic word representing more than one meaning). The results showed that reaction times were faster for words than non-words, and that they were faster for words of high frequency than for words of low frequency. The results also showed that reaction times for homographic words were faster than for non-homographic words.

Gernsbacher (1984) has claimed that word familiarity is actually a more reliable indicator of word recognition latencies than printed word frequency counts, particularly for low printed frequency words. For instance, although all the low frequency words *mumble*, *giggle*, *drowsy* and *rend*, *cant*, *pithy* have the same low printed word frequency of 1 per million, reaction times have been found to be slower for the second set (Gernsbacher, 1984, p. 261). This is partially because, by their very nature, printed frequency counts are based on samples of written language usage (p. 260), whereas word familiarity is more a measure of personal frequency, and is likely to influenced by the number of times a word is spoken, heard, and written, rather than read.

4.2.2 Non-word legality

This effect refers to the finding that responses to non-words will depend on the legality of the non-word (Novik, 1974; Stanners & Forbach, 1973; Stanners, Forbach, & Headley, 1971). Garman (1990) suggests that legality may reflect purely visual constraints, but pronounceability through grapheme-phoneme conversion rules may also be involved.

Comparing reaction times in the lexical decision task for three letter strings forming either consonant-vowel-consonant (CVC) words (e.g., *sat*), consonant-vowel-consonant (CVC) nonwords (e.g., *sut*), and all consonant (CCC) non-words (e.g., *svt*), Stanners, Forbach, and Headley (1971) found that reactions times were fastest in rejecting CCC nonwords, followed by reaction times to CVC words, with reaction times in rejecting CVC nonwords being the slowest. This pattern of results was duplicated by Stanners and Forbach (1973) using five letter strings, where reaction times were fastest in rejecting CCCCC nonwords (e.g., *crnss*), followed by reaction times to CCVCC words (e.g., *cross*), with reaction times in rejecting CCVCC nonwords (e.g., *cruss*) being the slowest. Although these studies were interpreted as indicating a two-stage process involving an initial evaluation of phonological lawfulness, where phonologically illegal letter strings could be rejected, and a later stage of a search process, which would be required to reject phonologically legal nonwords, such an account has problems with the findings from Novik (1974) where meaningful CCC nonword strings, such as JFK and LSD, were slower to reject than meaningless CCC nonwords.

4.2.3 Neighborhood effects

In addition to word frequency and word familiarity effects described above, there is also evidence that suggests that the recognition of a word is influenced by orthographically similar words, referred to as ‘neighbors.’

Much of the research into neighborhood effects has adopted Coltheart, Davelaar, Jonasson and Besner’s (1977) straightforward definition of an orthographic neighbor—any word that can be generated by changing just one letter of a given word while preserving letter positions (e.g., *mice* and *race* are both neighbors of *rice*), with the neighborhood being the set of such neighbors. However, while this definition is simple enough, there has been much controversy surrounding neighborhood effects, especially over whether these are inhibitory or facilitatory in nature.

For instance, although Coltheart et al. (1977) only reported an effect for nonwords, where nonwords with large neighborhoods were slower to reject, Andrews (1992) found facilitated lexical decision and standard naming latencies for low-frequency words. The situation is further complicated by the claim by Grainger and his colleagues (Grainger, 1990; Grainger, O’Regan, Jacobs, & Segui, 1989) that it is not neighborhood size per se but rather neighborhood frequency that is important. In Grainger, et al. (1989) having at least one neighbor of higher frequency produced inhibition in both lexical decision latencies and gaze durations, although this might be task-specific, for while Grainger (1990) also found inhibition in lexical decision, having at least one neighbor of higher frequency produced facilitation in naming. Examining the apparent contradiction in the findings from Andrews and Grainger, et al., Sears, Hino and Lupker (1995) point out that neighborhood size and neighborhood frequency are likely to covary. Their results, which are more consistent with Andrews, showed that a large neighborhood facilitated processing for low-frequency words, and the existence of higher frequency neighbors facilitated processing rather than inhibiting it. More recently, Pollatsek, Perea and Binder (1999) found that when the frequency of the highest frequency neighbor was equated, a larger neighborhood facilitated lexical decision judgments, although this had an inhibitory effect on reading sentences containing the target words.

While the results from the research into neighborhood effects have not always been consistent, much of the interest in these effects lies in the fact that they potentially offer a means of differentiating contrasting models of lexical access. For example, Andrews (1992) has argued that although her results are incompatible with lexical-access models that involve serial comparison, activation models assuming either localized or distributed representations are capable of accounting for neighborhood effects.

4.2.4 Morphological family effects

As well as the purely visual overlap or similarity of orthographic neighbors, word recognition is also influenced by morphologically related words. For instance, Taft (1979) provided an early demonstration of this in his experiments that manipulated both surface frequency, the frequency for a given form, and base frequency, the summed frequencies of inflected forms. When surface frequency was kept constant, responses were faster for words with higher base frequencies.

More recently, investigating the components of word frequency effects with monomorphemic, or simplex, nouns in Dutch, Schreuder and Baayen (1997) reported an effect of morphological family size. In addition to distinguishing between the surface frequencies for singular and plural forms (i.e., *table*, *tables*), which would give a stem frequency for inflectional forms when combined, they used the term morphological family to refer to the set of words sharing a stem formed either by derivation (i.e., *tablet*, *tabular*) or compounding (i.e., *tablespoon*, *timetable*). There are two counts related to the morphological family—family size is a type count of the number of different words in the set, while cumulative family frequency is the summed token frequencies of those words. However, although Schreuder and Baayen found that response times in lexical decision and subjective ratings of frequency were influenced by morphological family size, there was no effect for cumulative family frequency. This family size effect has also been observed for English simplex nouns (Baayen, Lieber, & Schreuder, 1997), for inflected and derived Dutch nouns (Bertram, Baayen, & Schreuder, 2000), and for Dutch verbs (De Jong, Schreuder, & Baayen, 2000). Evidence concerning semantic selection restrictions for affixes, semantically transparent family members (Bertram, et al.), verb morphology, and mediation via abstract central morphological representations (De Jong, et al.) suggests the family effect is a semantic-morphological effect. Krott, Baayen and Schreuder (2001) introduce the term 'constituent family' to refer to a set of compound words that share the same first or second constituent. Although the selection of linking morphemes in novel noun-noun compounds in Dutch is very unpredictable by rules, Krott, et al. argue, based on the results of production experiments and analogy simulations, that this can be predicted reasonably well by analogy based on the constituent families of both the left and right constituents.

Supporting Bertram, et al.'s (2000) claim that much of the interconnectivity in the human mental lexicon is based on networks of morphologically related words, these morphological family effects must also have important implications for models of the mental lexicon. These papers reporting morphological family effects have primarily interpreted their results within the parallel dual-route model of morphological processing proposed by Schreuder and Baayen (1997) and Baayen, Dijkstra, and Schreuder (1997), which is discussed later in Section 4.3.2.5.[41]

4.2.5 Semantic priming effects

Balota (1994) observes that the semantic (associative) priming paradigm is undoubtedly the most researched area within the priming literature, suggesting that the interest is, at least in part, because semantic priming appears to be well suited to the investigation of the structure of meaning-level representations and the processes of retrieval that operate on these representations (p. 337).[42]

The seminal study of semantic priming is Meyer and Schvaneveldt (1971). Using pairs of associatively related words, such as *doctor–nurse* and *bread–butter*, participants were asked to respond ‘yes’ if a pair of simultaneously presented letter strings were both English words (Exp. 1), and to respond ‘same’ if the letter strings were either both words or both non-words (Exp. 2). In both experiments, responses were faster for pairs of associatively related words than for unrelated pairs of words. That is, participants responded ‘yes’ quicker to *doctor–nurse* than to *doctor–butter*.

While this extremely simple yet elegant finding has been immensely influential within visual word recognition research, in his review of semantic priming effects, Neely (1991) argues that the single-word semantic paradigm has two important advantages over the procedure that Meyer and Schvaneveldt (1971) used (p. 265). The first is that the task for the participants is simpler because they are judging the lexicality of just one letter string rather than two. The second advantage is that it is possible to vary the SOA between the prime and target. If a response is required to both the prime and the target, then the SOA must be sufficient to allow for the response to the prime, but if no response is required to the prime, then it is possible to use very short SOAs to investigate semantic priming.

As the length of the SOA used in Experiment 1 to be reported in Chapter 5 required further investigation, it is worth briefly discussing Neely’s (1977) influential study where the manipulation of the SOA was an important aspect. This study provided experimental support for the theory of Posner and Snyder (1975) that there are two separate components of attention—automatic processing and attentional or expectancy-based processing. The automatic spreading–activation process is fast, parallel, and independent of intention or conscious awareness, whereas the attentional process is slow, serial, and intention or conscious dependent.

In his experiment using a lexical decision task, Neely (1977) investigated these components of attention by varying the relationship between a category word prime and target word. Neely manipulated the relationship between prime and target in three ways, based on the instructions that were given to the participants. As they were told to expect after *bird* an exemplar of that category, but after *body* a word for a part of a building, and after *building* a word for a part of the body, the first way was whether there was a shift (*body–door*) or no shift (*bird–robin*) in attention. Having thus created expectations in the participants, the second way in which the prime-target relationship varied was according to whether it was expected (*bird–robin*) or unexpected (*body–heart*). The third way was whether the prime was semantically related (*bird–robin*) or unrelated (*body–door*). In addition, then, to a neutral prime condition (XXX) as a baseline, there were five prime-target relation conditions: (1) no shift-expected-related (*bird–robin*), (2) shift-expected-unrelated (*body–door*), (3) no shift-unexpected-unrelated (*bird–arm*), (4) shift-unexpected-unrelated (*body–sparrow*), and (5) shift-unexpected-related (*body–heart*). The SOA was varied both between and within participants over four intervals of 250 ms, 400 ms, 700 ms, and 2,000 ms.

The important finding from this experiment was the way in which the patterns of facilitation and inhibition relative to the baseline condition varied across the range of SOAs. In condition (1) of no shift-expected-related (*bird–robin*), facilitation was constant over all SOAs. Although there was no difference from baseline at 250 ms in condition (2) of shift-expected-unrelated (*body–door*), there was facilitation at 2,000 ms. In condition (3) of no shift-unexpected-unrelated (*bird–arm*), there was neither facilitation nor inhibition at 250 ms, but there was inhibition at 2,000 ms. This pattern was also found in condition (4) of shift-unexpected-unrelated (*body–sparrow*). The changing pattern for condition (5) of shift-unexpected-related (*body–heart*) was the most important, for although at 2,000 ms there was inhibition for this condition, at 250 ms there was facilitation. This indicates that at a short SOA of 250 ms, priming is automatic and not subject to intention or conscious awareness, which only becomes influential at longer SOAs.

4.2.6 Constituent morpheme priming effects

The constituent-morpheme priming paradigm provides an important technique with which to examine morphological involvement within lexical retrieval (For a brief review of different forms of morphological priming, see Drews, 1996).[43] Here, the effect on the lexical decision responses to a compound word of a prior presentation of one of the constituents of that word is compared to responses in an unrelated control condition. As this is the basic paradigm adopted in the experiments to be presented in Chapters 5 and 6 of this thesis to investigate the lexical retrieval and representation of two-kanji compound words from a morphological perspective, as background, we briefly introduce some studies conducted in English, Dutch, Greek and Polish with alphabetic writing systems.

Monsell (1985) conducted a constituent-priming experiment with English compound words, comparing semantically transparent compounds (e.g., *tightrope*) with opaque compounds (e.g., *butterfly*), as well as with a monomorphemic 'pseudo-compound condition' (e.g., *bone* in *trombone*). While he found facilitation in both the first and second element prime conditions for both the semantically transparent and opaque compound conditions, there was also an effect in the pseudo-compound condition, which suggests that the effects might be orthographic.

Sandra (1990) has conducted a series of three similar experiments with Dutch compounds. In the first, primes were associatively related to a constituent element of either opaque compounds or pseudo-compounds. Thus, for example, *bliksem* 'lightning,' a word related to *donder* 'thunder,' was a prime to *donderdag* 'Thursday,' and *peer* 'pear,' a word related to *appel* 'apple,' was a prime to *aardappel* 'potato.' The results of this experiment showed no priming effects for either word type. However, the results from a second experiment, which used the same primes with semantically transparent compounds, such as *bliksem* to *donderslag* 'thunder-stroke' and *peer* to *winterappel* 'winter-apple,' did show priming for both constituent conditions. In a third replication experiment, which focused on the word-final constituent, the finding that associatively related primes to a word-final constituent do not prime opaque compounds (*rente* 'interest' to *zandbank* 'sand-bank'), but do prime semantically transparent compounds (*brand* 'fire' to *kampvuur* 'camp-fire') was confirmed.[44]

Zwisterlood's (1994) study also compared different types of Dutch compound words, claiming that priming for transparent compounds can always be interpreted as arising at a semantic level, but if priming could also be obtained for opaque compounds, it would provide strong evidence for a separate contribution of morphology. In a reversal of the prime and target pairing seen in Sandra (1990), however, Zwisterlood (1994) presented a compound word as a prime to constituent target. In her first experiment, Zwisterlood (1994) used pairs of compounds sharing a constituent, with one compound always being semantically transparent, e.g. *kerkorgel* 'church organ.' The other compound was either partially or fully opaque, such as *drankorgel* 'drunkard.' These compound words were presented as primes to both constituents as targets, e.g., *kerkorgel–orgel* and *drankorgel–orgel*, as well as *kerkorgel–kerk* and *drankorgel–drank*. The stimulus items also included words with orthographic overlap, such as words like *kers* 'cherry' in the beginning of *kerstfeest* 'Christmas' and *aap* 'ape' in the ending of *koolraap* 'turnip,' and pseudowords like **troos* in the beginning of *troostprijs* 'consolation prize' and **rempel* in the end of *pijndrempel* 'pain threshold.' The results from this experiment showed priming for all the morphological conditions. However, importantly, no facilitation from form overlap was found in the orthographic word conditions. This suggests that the priming effect in the morphological conditions was indeed due to the morphological relation between the primes and targets. Zwisterlood stresses the importance of the fact that priming was obtained for both of the constituent elements for truly opaque compounds, which cannot be explained away by semantic relatedness. In a second experiment that attempted to separate morphological and semantic contributions, the same types of compounds were used as primes to targets that were semantic associates of the constituents of the primes. Thus, for instance, the fully transparent compound *kerkorgel* appeared as a prime to *priester* ‘priest’ in a first component condition, to *muziek* ‘music’ in a second component condition, and as a prime to *bier* ‘beer’ in a third condition based on the paired compound *drankorgel*. No priming was found for the truly opaque compounds or for the pseudo-compounds, which suggests that these words are not linked in any way to component words. However, for transparent and partially opaque compounds a semantic priming effect was obtained, for both prime types and for both constituents.[45]

Kehayia, Jarema, Tsapkini, Perlak, Ralli, & Kadzielawa (1999) carried out constituent-morpheme priming experiments using transparent noun-noun and adjective-noun compound words in Greek and Polish, which are highly inflected languages. Kehayia, et al. reported priming for both constituent conditions compared to an unrelated prime condition in both languages. However, they also reported an advantage for the first constituent with faster reaction times than for the second constituent, despite the fact that in these modifier + modified compounds the second constituent is the head. They interpreted this result as evidence that the first constituent plays an important role in compound word recognition.

4.3 Issues and models

Having briefly introduced some important findings from the visual word recognition literature, we turn now to consider two central issues for the mental lexicon and to see how some influential models tackle them.[46]

The two central issues within visual word recognition are undoubtedly lexical retrieval and lexical representation. Lexical retrieval, or lexical access, is primarily concerned with the journey from the printed word to the retrieval of a word's meaning. As noted already, much of the debate here has focused on the relative contributions of orthographic activation and phonological activation, but it has also been concerned with the nature of the retrieval mechanism. In contrast, lexical representation is primarily interested in how words might be arranged and linked in the mental lexicon. As we shall see this issue is particularly important for polymorphemic words. These two issues are, however, mutually defining aspects of the mental lexicon. Just as certain lexical retrieval mechanisms will require certain representations and arrangements of lexical information, equally, the way in which lexical information is represented will greatly determine the type of mechanism underlying lexical retrieval.

4.3.1 Lexical retrieval: From print to meaning

Models that attempt to explain the journey from printed word to the retrieval of a word's meaning may be distinguished by the mechanism assumed to underlie the retrieval process, such as search mechanisms and activation mechanisms. Here, we shall briefly introduce Forster's (1976, 1989) serial search model and two related models that assumed activation mechanisms, the interactive-activation and competition (IAC) model proposed by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) and the distributed activation model proposed by Seidenberg and McClelland (1989).[47]

4.3.1.1 Serial search model: Forster (1976, 1989)

While a few other models employ forms of ordered serial searching, such as the activation-verification model (Paap, Newsome, McDonald, and Schvaneveldt, 1982), the checking model (Norris, 1986) and the verification model (Becker, 1976), we shall only discuss the serial search model proposed by Forster (1976, 1989), which is probably the most well known, in order to illustrate search mechanisms, depicted in Figure 4.1.

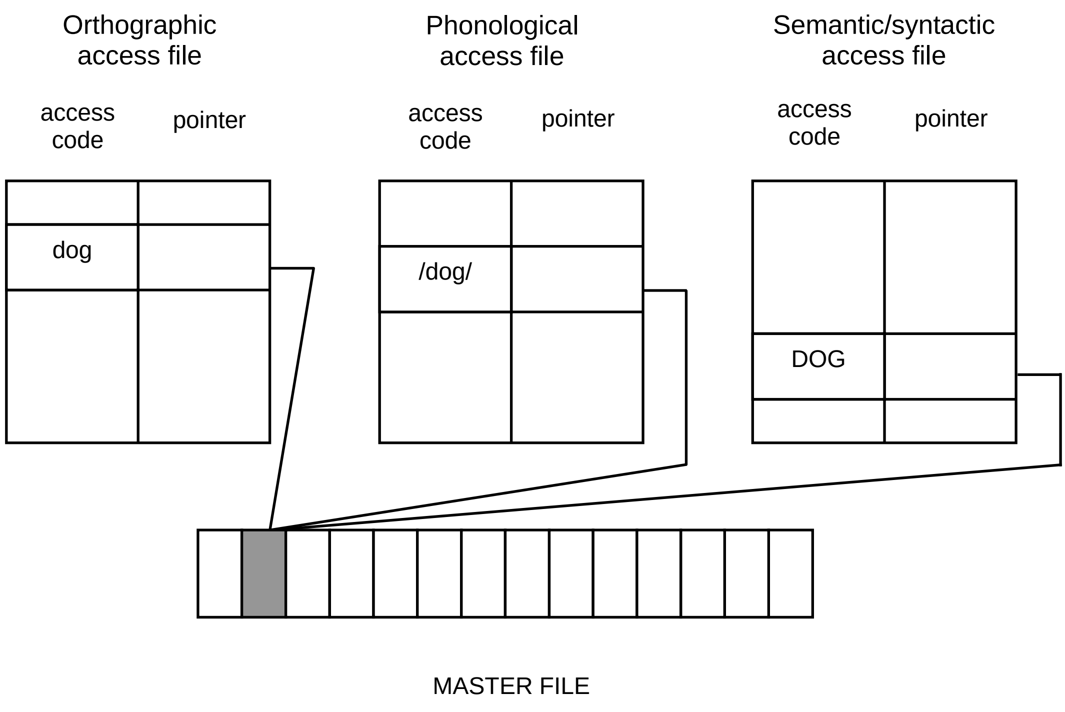


Figure 4.1. Forster's (1976) serial search model, showing the organization of peripheral access files and the master file (based on Forster, 1976, p. 268).

As the name implies, this model claims that when a letter string is presented, a serial search is made of the lexicon on the basis of sensory information from the letter string until a match is made with a lexical entry. Forster (1976) actually advocates a two-tiered architecture for this model on the assumption that the mental lexicon is accessed during the different activities of listening, speaking, and reading, but that it is impossible to organize entries according to three kinds of properties–phonological, semantic/syntactic, and orthographic–in one place. Thus, the model consists of three access files linked to a modality-free master file, which contains all the information about a word, such as meaning, spelling, word class, etc.

Rather than claiming that the complete lexicon is scanned every time a word is processed, however, the model assumes that entries in the access files are grouped in sets, called 'bins,' based on shared properties. Sensory information about the visual form of a presented letter string will be used to direct the search to an appropriate bin. This is then serially searched until a reasonable match is made with an entry in the access file, which leads to the entry in the master file. Once full information from the master file is available, the match between spelling and sensory input can be checked, and if they match, the letter string is recognized as the word. If, however, the sensory input does not match, the search in the orthographic access file is resumed until a better match is found. This two-stage model has been likened to searching for a book in a library, where a search would first be made of a card catalogue to find the location of the book in the library (Harley, 2001; Singleton, 2000). In this analogy, the card catalogue corresponds to the peripheral access files, and the shelves with the books to the entries in the master file.

In the serial search model, the semantic priming effects are explained in terms of cross-references between entries in the master file, linking, for instance, *doctor–nurse*. The model suggests that when an entry is accessed, this cross-referencing network will activate related words. When a related word is subsequently presented, it may be found quicker by means of a search of the activated words, which is carried out simultaneously with the search of the orthographic access file, and thus lead to faster reaction times. Frequency effects are explained by having the entries in the bins arranged in terms of frequency, so that entries for high frequency words will appear towards the top of the lists and thus will be found sooner.

4.3.1.2 Interactive-activation and competition (IAC) model: McClelland & Rumelhart (1981), Rumelhart & McClelland (1982)

An extremely important model of lexical access is the interactive-activation and competition (IAC) model. McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) presented this early connectionist model to explain word context effects in letter identification, the finding that letters are easier to recognize in words than when seen as isolated letters.

The model proposes three levels of processing units, which are represented in Figure 4.2. At the lowest level of input into the system, there are visual feature units. Next, there are letter units, and at the level of output from the system, there are word units. A network of connections links the various units.

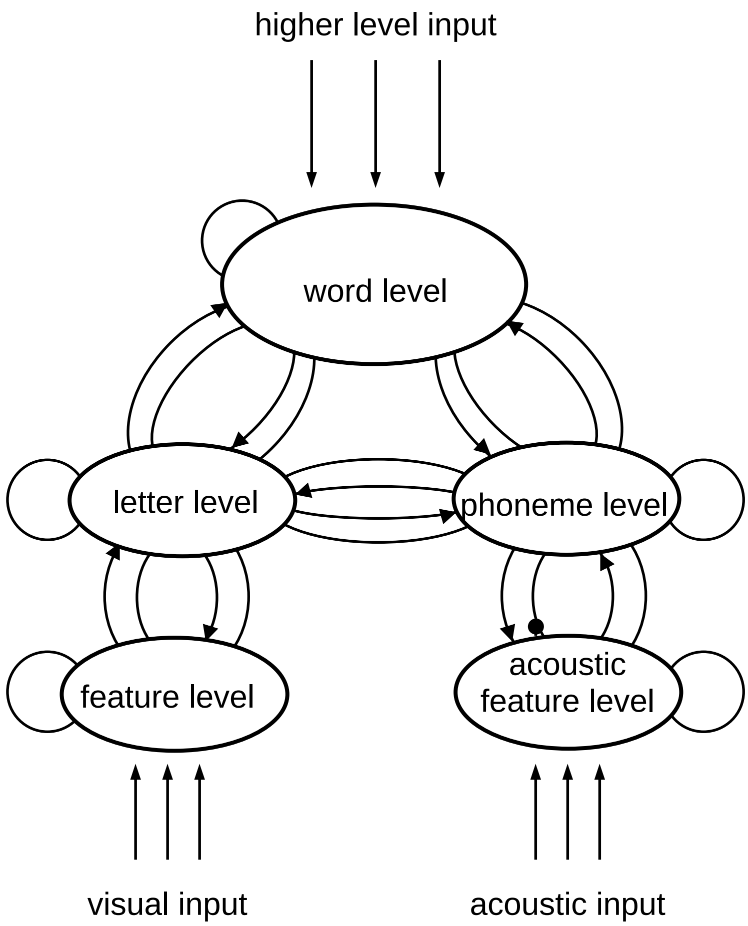


Figure 4.2. The IAC model, showing an outline of some of the processing levels involved in visual and auditory word perception, with interconnections (based on McClelland and Rumelhart, 1981, p. 378).

The connections between different levels are either facilitatory or inhibitory in nature (represented in the figure by arrows and circles, respectively). For instance, the letter unit for ‘T’ would be linked by facilitatory connections to the word units for ‘TRIP,’ ‘TRAP,’ and ‘CART,’ but by an inhibitory connection to the word unit for ‘ABLE.’ However, connections between units at the same level, for instance, such as between the letter units for ‘T’ and ‘C,’ are all inhibitory. When a word is presented, relevant visual feature units are activated, and this activation is passed along to connected units, either facilitating or inhibiting their activation. In this way, appropriate letter units are activated, which in turn activate word units. The model is interactive, because activation from higher units can also feed back to lower units, influencing their activation levels. For instance, when the word ‘TRIP’ is presented, activation from feature units will start to activate the letter units ‘T,’ ‘R,’ ‘I,’ and ‘P.’ As well as activating the word unit for ‘TRIP,’ some activation will also pass to the word unit for ‘TRAP.’ However, as the activation for ‘TRIP’ builds up, inhibitory activation is sent to the word unit ‘TRAP’ and inhibitory feedback is also sent down to the letter unit ‘A.’ Eventually, through this process, the activation for the word ‘TRIP’ exceeds that of other competing units, and it is recognized.

4.3.1.3 Distributed activation model: Seidenberg & McClelland (1989)

Closely related to the interactive-activation and competition (IAC) model proposed by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) is the distributed activation model proposed by Seidenberg and McClelland (1989).

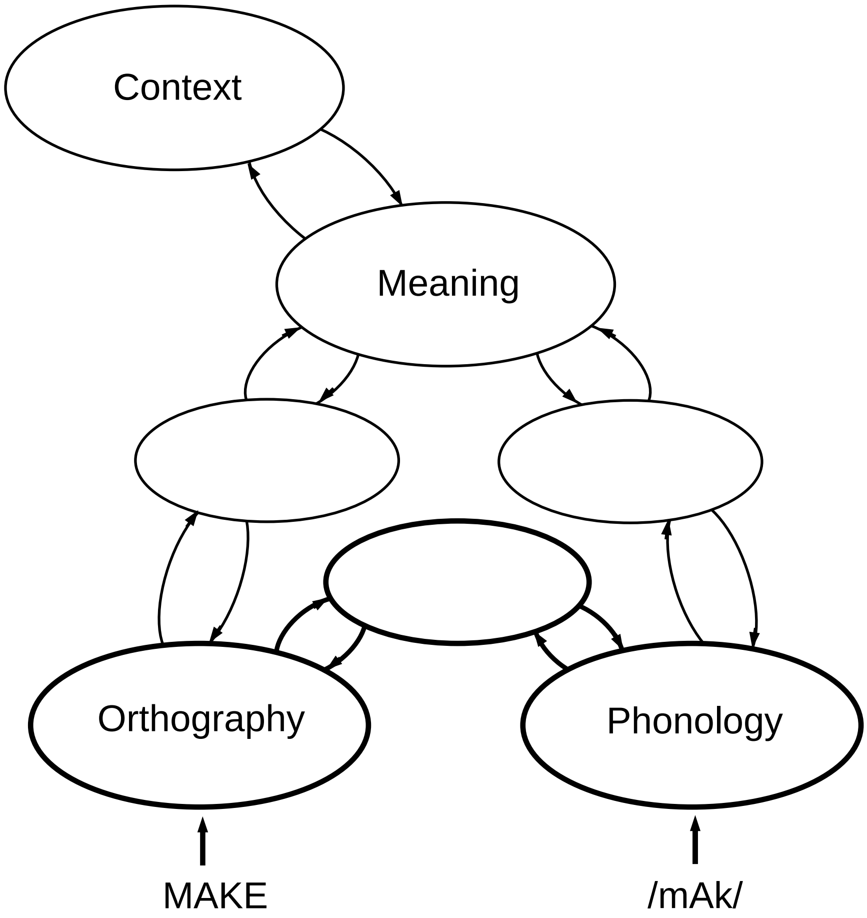


Figure 4.3. General framework for lexical processing in the Seidenberg and McClelland (1989) distributed activation model (the implemented model is in bold lines, based on figure on p. 526).

As the name of the model suggests, in contrast to the localist structure of the interactive-activation model, this model assumes distributed representations (Rumelhart, McClelland, & the PDP Research Group, 1986). Although seen as part of a larger framework, as shown in Figure 4.3, Seidenberg and McClelland (1989) only implemented a part of this in their simulation model (as indicated in bold in the figure). This larger framework assumes that reading words involves the computation of three types of codes: orthographic, phonological, and semantic. Seidenberg and McClelland note that other codes are probably also computed, such as those concerning the syntactic and thematic functions of words, but they did not attempt to include them in the model because they probably are more relevant to comprehension processes than to the recognition and pronunciation of mono-syllabic words, which was the focus of the simulation model of the naming task.[48] Each of these codes is assumed to be a distributed representation; that is, to be a pattern of activation distributed over a number of primitive representational units. Each processing unit has an activation value that in their model ranges from 0 to 1. The representations of different entities are encoded as different patterns of activity over these units.

Processing in the model is assumed to be interactive. That is, Seidenberg and McClelland assume that the process of building a representation at each of the three levels both influences, and is influenced by, the construction of representations at each of the other levels. In line with this inherently interactive view, they also assume that word processing can be influenced by contextual factors arising from syntactic, semantic, and pragmatic constraints, although these were not incorporated within the simulation model.

As in other connectionist models, processing is mediated by connections among the units. However, commenting that there are limits on the processing capabilities inherent in networks in which there are only direct connections between units at different representational levels, the simulation model includes a set of hidden units, mediating between the pools of representational units.

As Seidenberg (1990) discusses in detail, this model differs from other accounts of the mental lexicon in a fundamental way. This is the manner in which lexical knowledge is represented and processed in the model. The standard view, common to models such as Forster's (1976, 1989) serial search model discussed earlier and dual-route models (Coltheart, 1978; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), is that lexical memory consists of entries corresponding to the different codes of words in a mental lexicon. The Seidenberg and McClelland (1989) model departs from these models in not regarding the mental lexicon as consisting of entries for individual words. Rather, knowledge of words is seen as being embedded in a set of weights on the connections between processing units encoding orthographic, phonological, and semantic properties of words, and the correlations between these properties. The spellings, pronunciations, and meanings of words are not listed in separate stores; hence, lexical processing does not involve accessing these stored codes. In contrast, lexical information is computed on the basis of the input string in conjunction with the knowledge stored in the network structure, resulting in the activation of distributed representations. Accordingly, the notion of lexical access does not play a central role in Seidenberg and McClelland's model, because it is incongruent with the representational and processing assumptions made by the model.[49]

4.3.2 Lexical representation: Representing polymorphemic words

Having looked at models of lexical retrieval that contrast in their underlying mechanisms, with either search mechanisms or activation mechanisms, we now turn to consider the related issue of lexical representation, and in particular, to consider how polymorphemic words might be represented in the mental lexicon.

Psycholinguistic interest in morphology stems from the fact that information concerning the structure of polymorphemic words is an important part of our linguistic knowledge (Feldman, 1995; Jarema, Kehayia, & Libben, 1999; Sandra & Taft, 1994; Taft, 1991).[50] This is clearly evidenced by the sheer numbers of polymorphemic words that already exist in the lexicons of all languages, and in the relative ease with which language users are able to produce and understand new polymorphemic words. Neither of these phenomena can be adequately accounted for unless we grant that language users possess some awareness of the morphological structure of polymorphemic words (Sandra, 1994). However, although we may readily acknowledge the existence of some form of morphological awareness, it is far from clear how this might best be characterized. The challenge is to discover the nature and extent of this awareness.

Cast very much in terms of logical alternatives, interest has largely been concerned with deciding between competing models of lexical representation, and the assumptions regarding lexical organization and processing that these models make. At one extreme, there are models with only whole-word representations. At the other extreme, there are models of fully decomposed storage. In the middle are models that propose the existence of both whole-word and morpheme representations, but with differing approaches to access.

Full listing models hold that whole words are the only representations in the lexicon (Rubin, Becker, & Freeman, 1979; Butterworth, 1983). In such accounts, morphological information can only become available after the whole-word representation has been accessed. The obvious conclusion from this is that morphological information, which is only available post-lexically, can have no role in lexical access. The opposite position is that polymorphemic words are stored in decomposed form, and that morphological parsing is obligatory in lexical access. This claim was made in the influential papers by Taft and Forster (1975, 1976) that proposed a 'prefix-stripping' model of visual word recognition for polymorphemic words.

4.3.2.1 Prefix-stripping model: Taft & Forster (1975, 1976)

According to this model, prefixed words are morphologically decomposed (i.e., the prefix is stripped off) and accessed via a search for the stem morpheme in the mental lexicon. This hypothesis was based on a series of experiments using bound stems. Non-words in one experiment were either 'real stem' non-words formed by removing a prefix from prefixed words, e.g., *juvenate* from *rejuvenate*, or 'pseudo stem' non-words formed by removing letters from pseudo-prefixed words, e.g., *pertoire* from *repertoire*. The result of this experiment was that 'real stem' non-words were slower to be judged non-words than the 'pseudo stem' non-words. The result of a second experiment was that words that are both free and bound morphemes, such as *vent*, were slower to classify when the bound form is more frequent than the free form. In a third experiment, non-words were formed by adding a prefix to 'real stem' non-words and 'pseudo stem' non-words, e.g., *dejuvenate* and *depertoire*, respectively. The result of this experiment was that non-words containing a 'real stem' were slower to be judged non-words than non-words containing a ‘pseudo stem.’ All of these results were taken as indicating the existence of lexical representations for bound stems.

Discussing the question of what purpose such morphological decomposition could serve, Taft and Forster (1975, p. 645) suggested three possible explanations. The first is the storage economy from only having to list word stems once. The second is that with decomposed storage, semantically related words would be stored close together. Being a search model of lexical access, the third possible explanation offered was that decomposed storage could provide retrieval that is more efficient for polymorphemic words. However, Sandra (1994), in a discussion of the morphological issues of lexical representation and processing, has pointed out that there are problems with these notions of economy and processing efficiency.

Other problems include the finding by Andrews (1986) that stem frequency effects seem to depend on experimental context. This is a problem for the prefix-stripping model, because stem frequency is seen as a characteristic of the stem representation, and should be independent of experimental context in a model that claims obligatory decomposition. Another problem is the implicit requirement that there is a ‘prefix store’ or list in the mental lexicon that makes prelexical prefix stripping possible. This implies that readers have an awareness of what is and is not a prefix, even though the status of some letter strings are questionable, such as *mini*, *bio*, and *pseudo* (Taft, 1994, p.278).

4.3.2.2 Augmented addressed morphology model: Caramazza, Laudanna, & Romani (1988)

As already mentioned, models of lexical representation covering the range of logical possibilities have been proposed, from models claiming that there are only whole-word representations at one extreme, to models requiring fully decomposed representation at the other extreme. Occupying the middle ground between the two extremes are models that hold that both whole-word and morpheme representations exist.

One such model is the augmented addressed morphology model proposed by Caramazza, Laudanna and Romani, 1988 (see also Chialant & Caramazza, 1995; Laudanna & Burani, 1995). Although this model has been proposed mainly to account for inflectional morphology, rather than compounding which is the central interest of this thesis, the model combines whole-word access with parsing in a form of dual-route model, with a direct route employing whole-word access representations and an indirect parsing route. According to earlier versions of this model, 'known' words are all handled by the direct route, and the parsing route is a backup option for rare or novel morphologically regular and orthographically transparent complex words. The second route comes into play only after completion of the first. In later versions, a slightly less restrictive position is taken, where words with a low surface frequency but with high-frequency constituents might be effectively processed via the parsing route. Thus, when a letter string is presented, for known words whole-word access is usually faster, but the parsing route can be faster in the case of novel words.

4.3.2.3 Multilevel interactive-activation framework: Taft (1991, 1994)

Partly because of the problems with the prefix-stripping model, mentioned above, Taft (1991, 1994) has more recently adopted a multilevel interactive-activation framework to account for the morphological processing of polymorphemic words.

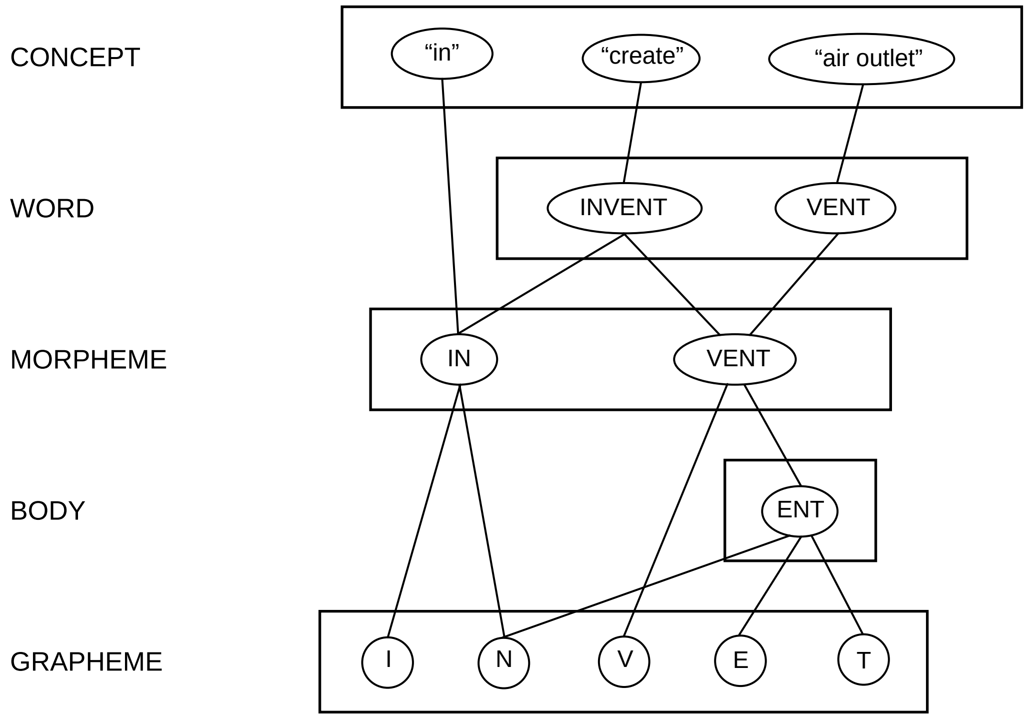


Figure 4.4. The multilevel interactive-activation model incorporating a morpheme level, depicting the representation for INVENT (based on Taft, 1994, p. 280).

In expanding on the basic units of the interactive-activation model discussed earlier, Taft has included word bodies, such as *int* of *pint*, to capture orthographic-phonological relationships above the single grapheme-phoneme level (1991), as well as morphemes (1994). By incorporating a morpheme level into the model, Taft (1994) eliminates the need for a 'prefix store,' but by virtue of the fact that these will also constitute independent activation units at the morpheme level, the model accounts for stem frequency effects. Figure 4.4 depicts the orthographic units for this model.

Units at the concept level are included here to represent the links or interface between meaning and orthographic form and are not intended to imply that lexical meanings are represented as single nodes. The basic mechanisms underpinning this model are the same as those outlined earlier. When a word is presented, visual feature units are activated, and this activation is passed on to the grapheme units. From there, the activation moves up the levels, until the activation passed to a particular word exceeds some activation threshold and the word is recognized. The frequency with which a pathway is used will influence the strength of activation passed along, so that high-frequency words will be activated more strongly than low-frequency words.

This model can explain priming effects by assuming that when a unit is used again shortly after activation, lingering activation in the unit will give it an advantage over other units that have not been activated recently. Although Taft (1994) offers this as a possible interpretation, he does, however, stress the need for caution in explaining results of priming studies, which might be open to non-lexical strategic factors (p. 291).

4.3.2.4 Parallel dual-route model of morphological processing: Schreuder & Baayen (1995), Baayen, Dijkstra and Schreuder (1997), and Baayen & Schreuder (1999)

Another model that occupies the middle ground between the two extreme positions on the lexical representation of polymorphemic words is the parallel dual-route model of morphological processing proposed by Schreuder and Baayen (1995), Baayen, Dijkstra and Schreuder (1997), and Baayen and Schreuder (1999).[51]

Schreuder and Baayen (1995) suggest that lexical representations consist of a concept node, which is connected to syntactic and semantic representations, as depicted in Figure 4.5. In this model, the mapping of access representations onto the relevant syntactic and semantic representations is seen as being mediated by these concepts nodes (p. 135). In giving computational formulation to this basic structure, Baayen, Dijkstra, and Schreuder (1997) describe this model as a race model with fully parallel routes, based on a spreading activation network with three representational layers: a layer of form-based modality-specific access representations (lexemes) and a layer of integration nodes (lemmas) that, in turn, are linked to a third layer of semantic and syntactic representations. The direct route maps a full-form access representation onto its associated lemma node, which in its turn activates its semantic and syntactic representations. In addition, the model contains a parsing route that operates in parallel with the direct route. Three stages in the parsing process are distinguished.

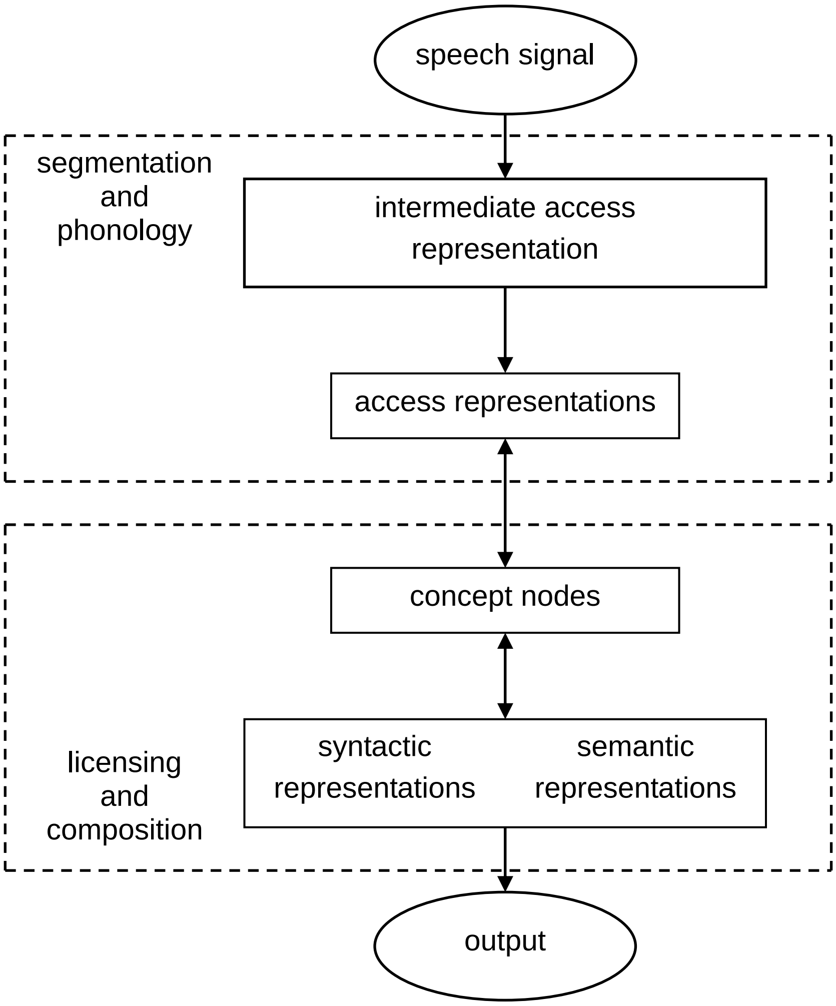


Figure 4.5. General outline of the Schreuder and Baayen (1995) model of morphological processing (p. 134).

In the first stage, access representations of affixes and stems become active along with full-form representations, leading to the activation of lemma nodes of the stems and the affixes involved. This stage is referred to as the segmentation process. More recently, Baayen and Schreuder (1999) have presented a computational tool that implements this segmentation stage of the morphological processing. Baayen and Schreuder comment that they regard this segmentation to be a fully bottom-up, autonomous process of pattern recognition. Thus, it is blind to the linguistic properties of access representations, such as whether they represent morphemes, simplex words, or complex words, properties that only contribute to the processing at later stages (p. 31). Following segmentation, two additional processes take place: licensing and composition. In the licensing process, the compatibility of subcategorization features of the activated constituents is checked. The subcategorization properties of a morpheme specify the properties that another morpheme should have if the two are to be combined into a single word. For instance, the English suffix -*ness* is subcategorized for attaching to adjectives and hence is excluded from attaching to, e.g., verbs: *kindness*, **thinkness*. In what is referred to as the composition process, the meaning of a complex word is computed from the meanings of its constituents. Finally, Schreuder and Baayen's (1995) model contains a mechanism of activation feedback from the syntactic and semantic layer via the lemma nodes to the access representations of constituents that are fully present in the signal. Activation feedback is hypothesized to allow cumulative frequency effects for transparent complex words only. Over time, activation feedback tunes the system such that an advantage for the parsing route results for transparent words, but a disadvantage results for semantically opaque words.

The most important difference between this model and the augmented addressed morphology model is that the parsing route and the direct route are engaged in the recognition process from the very beginning of the recognition process.

4.4 Visual word recognition research: Summary

This chapter has provided a very brief review of visual word recognition research, at the boundary of our transition from external form to internal representation. After introducing the lexical decision task and the naming task, which are employed in the experiments presented in this thesis, the chapter described some of the main effects within visual word recognition research. Finally, the chapter discussed the central issues for the mental lexicon of lexical retrieval and lexical representation, by looking at how some of the most influential models. As we have seen, these two issues are mutually defining. For instance, in the serial search model proposed by Forster (1976), the search mechanism assumed for lexical retrieval require an ordered arrangement of lexical representations based on shared features. In contrast, when activation is assumed as the mechanism of lexical retrieval, for example in the multilevel interactive-activation framework proposed by Taft (1991, 1994), the relationships between related polymorphemic words are modeled by the links or connections between lexical representations.

Looking at the models focusing primarily on lexical representation, it is striking that they all allow for some form of morphological involvement in the lexical retrieval of polymorphemic words, highlighting the language-universal importance of morphological information for any model of the mental lexicon. In the next two chapters, we focus on the organization of the Japanese mental lexicon by investigating the lexical retrieval and representation of two-kanji compound words from a morphological perspective.

Chapter 5

The Japanese mental lexicon: The lemma unit model

Based on a linguistic description of external form, Part 1 of this thesis has argued that kanji function as a morphographic writing system. Here, we continue our psycholinguistic investigation of internal representation by examining the implications of this for the organization of the Japanese mental lexicon by investigating the lexical retrieval and representation of two-kanji compound words from a morphological perspective.

Specifically, this chapter focuses on two proposals for a model of the Japanese mental lexicon that differ in their explanations of lexical retrieval and representation. The first proposal is Hirose's (1992, 1994, 1996) hypotheses, which evoke search mechanisms. The second proposal is Joyce's (1999, in press) suggestion for a Japanese lemma-unit version of the multilevel interactive-activation framework, which assumes activation mechanisms. As any model of the Japanese mental lexicon must in some way capture the morphological relations that exist between polymorphemic words, we shall consider how well these two proposals cope with the diversity in the morphological structure of two-kanji compound words, which was discussed in Chapter 3.

However, before turning directly to discuss these two proposals for the Japanese mental lexicon, we make a highly selective review of psychological studies of kana and kanji processing, looking in particular at a few studies from both the Chinese and Japanese literature that have focused on two-kanji compound words.

5.1 Psycholinguistic studies of kana and kanji processing: A brief review

As Kess and Miyamoto (1999) observe, one of the most active areas of research within Japanese psycholinguistics has been the study of kanji and kana processing (see also, Chen, 1997; Chen & Tzeng, 1992; Chen & Zhou, 1999; Flores d'Arcais, 1992; Hatta, & Saito, 1999, 2000; Kaiho, & Nomura, 1983; Leong & Tamaoka, 1998; Morton & Sasanuma, 1984; Paradis, Hagiwara, & Hildebrandt, 1985; Saito, 1997; Sugishita, Otomo, Kabe, & Yunoki, 1992; Tamaoka, 1991, 1994). The highly selective review below will not even begin to scratch the surface of this literature, and only attempts to single out a few studies which illustrate the important lesson that Kess and Miyamoto draw from this rich body of research, which is that simple "early dichotomies" between semantic routes and phonological routes are inadequate to explain the complex interaction involved in reading (pp. 196-197).[52]

5.1.1 Kana studies

As observed in Chapter 2, the present Japanese writing system comprises a mixture of scripts, including morphographic kanji, two kana syllabaries, the alphabet and Arabic numerals, which are employed in largely separate and complementary ways (Kabashima, 1977; Kaiser, 1993a, Kess & Miyamoto, 1999; Smith, 1996, Taylor & Taylor, 1995). This mixture of scripts presents a variety of extremely interesting opportunities to further our understanding of the processes involved in visual word recognition, allowing for both comparisons of similar scripts (hiragana and katakana) and of different scripts (kana and kanji) (Besner & Hildebrandt, 1987; Feldman & Turvey, 1980; Shimamura, 1987; Shimomura, & Yokosawa, 1991; Steinberg, & Yamada, 1978-9; Yamada, 1992b; Yamada, Mitarai, & Yoshida, 1991).

While there is a very real danger that apparent contrasts may be confounded by factors like orthographic familiarity because the scripts are used in largely separate and complementary ways, some studies have tried to exploit such factors. Besner and Hildebrandt (1987), for example, manipulated orthographic familiarity in their naming study for katakana words. Comparing orthographically familiar words (normally written in katakana), orthographically unfamiliar (normally written in kanji) as well as nonwords, they found that orthographically familiar words were named faster than both unfamiliar words and non-words. Besner and Hildebrandt interpreted this result as evidence that lexical retrieval for katakana words "can be achieved without recourse to preliminary phonological recoding" (p. 340).

In an early and often cited comparative study of kana and kanji, Feldman and Turvey (1980) also contrasted orthographic familiarity, although arguably less successfully. Presenting six color words both in kanji, which they claim would be orthographically more familiar, and in kana, Feldman and Turvey found that naming latencies were faster for the kana words. They suggest that this result is because kana "conform to a phonographic principle" (p. 141) whereas kanji do not. However, as Kess and Miyamoto (1999) point out, clearly the color words used in this study vary greatly in terms of frequency, which certainly casts doubts over their claim that the kanji representation is the more orthographically familiar form for these words.[53]

A more enlightening comparative study of kana and kanji processing is the study by Shimamura (1987) using the Stroop test and an interesting variation of this paradigm. Presenting the conflicting color words in either kanji or kana, Shimamura reported greater Stroop interference for words in kanji than for words written in kana, although the words written in kana were named faster. In variations of the Stroop test, participants were asked to indicate the spatial location of stimuli, which could be a conflicting arrow or a conflicting word (e.g., a *left*-pointing arrow or the word *left* in the right position). Running this variation with English participants, Shimamura found that conflicting arrows produced more interference than conflicting words, while words were named faster. In a Japanese version of this variation, the greatest interference was found for conflicting kanji words, where again words written in kana were named faster than words written in kanji. These results dissociate word comprehension and word naming, for although kana were named faster than kanji, the greater Stroop interference for kanji words indicates that these were identified faster than kana.[54]

5.1.2 Kanji studies: Single characters

The main interest of studies of single kanji has been to determine the balance between semantic activation and phonological activation and their relative time courses in lexical retrieval (Flores d'Arcais, 1992; Flores d'Arcais, Saito, & Kawakami, 1995; Kadota & Ishikawa, 1999; Kess & Miyamoto, 1999; Mizuno, 1997; Saito, Masuda, & Kawakami, 1998; Sakuma, Sasanuma, Tatsumi, & Masaki, 1998; Tan & Perfetti, 1997, 1998; Perfetti & Tan, 1998; Perfetti & Zhang, 1995; Wydell, 1991; Yamada, 1998; Zhou & Marslen-Wilson, 1999a, 1999b). Many of these studies have focused on phonetic compound kanji, which, as discussed in Chapter 2, consist of a semantic marker and a phonetic marker, in order to investigate the complex interactions between orthographical, phonological, and semantic activation in the lexical retrieval of single kanji.

Flores d'Arcais (1992), for example, investigated these activation processes in a series of experiments, from which he concluded that in the recognition of complex characters, the meanings of the component radicals can be retrieved, even when they are not semantically related to the meaning of the word character. Flores d'Arcais comments that when the pronunciation of a kanji is emphasized in a task such as naming, then phonological information appears to become available before semantic information is fully available (p. 62).[55] Flores d'Arcais and Saito (1993) have also found evidence for the semantic activation of the component radicals of complex kanji. In a speeded semantic categorization task, they found interference in all the critical experimental conditions, which were graphically similar kanji pairs (e.g., 伸 'extend' and 仲 'friend'), part-whole related kanji pairs (e.g., 石 'stone' and 口 'mouth') and opaque-component related kanji pairs (e.g., 石 'stone' and 目 'eye').

There is also evidence that phonological activation contributes to the lexical retrieval of kanji words (Saito, Masuda, & Kawakami, 1998; Sakuma, Sasanuma, Tatsumi, & Masaki, 1998; Shimomura & Yokosawa, 1995; Wydell, Patterson,

& Humphreys, 1993). For instance, in addition to a significant effect of visual similarity, Wydell, Patterson, and Humphreys (1993) found a significant homophone effect, in a semantic categorization task, where reaction times were longer and errors greater when responding to homophone foils to the correct exemplars.[56] Sakuma, Sasanuma, Tatsumi, and Masaki (1998) have also been reported similar results for this task, although they also found that the homophone effect was reduced under pattern-masking conditions, whereas orthographic similarity effects remained strong under these conditions.

Saito, Masuda, and Kawakami (1998) provide further evidence for both orthographic and phonological activation using a delayed matching task. In this study, after brief presentation of two ‘source’ kanji (e.g., 略 and 伴), participants were asked to decide whether a ‘probe’ kanji (e.g., 畔) was one of the two source kanji. Saito, Masuda, and Kawakami only found a homophone effect when the probe was orthographically similar to the source kanji, and interpreted this result as evidence that phonological information for both the whole kanji and the radicals is automatically activated, even when not explicitly demanded by a task.

Taken together, these single kanji studies clearly indicate that lexical retrieval involves complex interactions between orthographic, phonological, and semantic activation.

5.1.3 Kanji studies: Two-kanji compound words

We now turn to look at a number of studies that have focused on two-kanji compound words, including some relevant studies from Chinese.[57] Our main interest here will be to look at how a number of models attempt to account for the lexical retrieval and lexical representation of these polymorphemic words.

5.1.3.1 Clusters and serial search: Hirose (1992)

Hirose (1992) has evoked search mechanisms[58] in his hypotheses concerning the organization of two-compound words based on the results from a constituent-morpheme priming study using the lexical decision task.

Hirose's (1992) Experiment 1 asked whether the processing of compound words is serial, using either the first constituent or second constituent as a retrieval cue, or whether it is holistic in nature. To test this, Hirose (1992) used a priming paradigm to compare the facilitation from five prime-target relationship conditions on the identification of a compound word in a lexical decision task, which are shown in Table 5.1.

As the basic procedure for presenting the stimulus items during a trial of the experiment is be adopted in Experiment 1 presented in this thesis, this will be described in detail later, but, in passing, we should note that the participants in this experiment were asked to name the prime when it was presented (presumably by its on-reading, although this is not clearly stated).

Table 5.1

Prime-Target Relation Conditions in Hirose (1992) Experiment 1

Prime-target relationship conditions	Prime	Target
First constituent condition	会	会社
Second constituent condition	計	設計
First same on-reading condition	犯	判定
Second same on-reading condition	練	関連
Unrelated condition	歴	責任

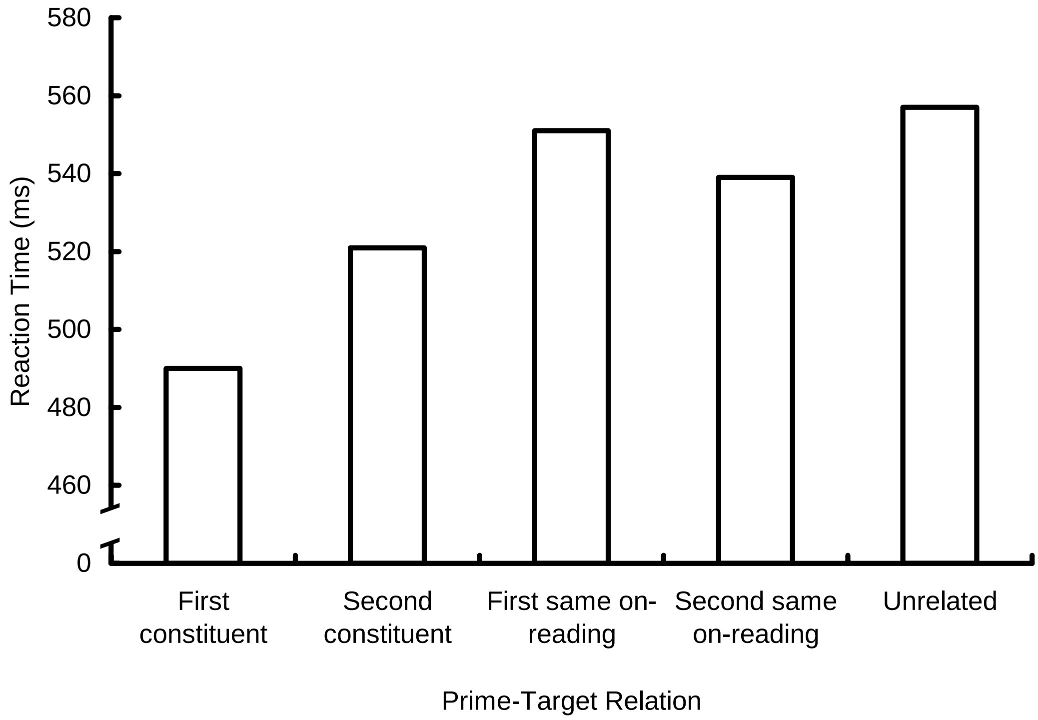


Figure 5.1. Mean reaction times (in milliseconds) as a function of prime-target relation in Hirose's (1992) Experiment 1.

Figure 5.1 presents the mean reaction times for the five conditions in Hirose's (1992) Experiment 1. Noting that there are no significant differences between the two same on-reading conditions and the unrelated condition, Hirose (1992) focuses his discussion of these results on the pattern of facilitation between the two same kanji conditions and the unrelated condition.[59] Both conditions when the prime kanji is a constituent of the compound word show significant priming compared to the unrelated prime condition. However, the reaction time is significantly faster for the first constituent condition compared to the second constituent condition. Hirose (1992) interprets this as evidence of serial processing, from left to right, of the compound words.

Based on the results of his first experiment, Hirose (1992) suggested that retrieval of compound words can be primed by presentation of a constituent, and that the retrieval process is serial in nature, proceeding from left to right. However, he pointed out that it was unclear whether this facilitation is due to the activation of phonological or semantic information for the prime. To examine this question in Experiment 2, Hirose (1992) added extra target conditions to contrast compound words which share a kanji but with different phonological values, such as 横顔 *yokogao* 'profile' and 横断 *ōdan* 'crossing.' The assumption was that if semantic information is facilitating retrieval of the compound, then there should be similar levels of facilitation for both conditions when the kanji are the same, irrespective of whether the kanji reading is the same or different. The six experimental conditions are given in Table 5.2.

Table 5.2

Prime-Target Relation Conditions in Hirose (1992) Experiment 2

Prime-target relationship conditions	Prime	Target
First constituent of prime compound is same as first constituent of target compound and with same reading	確率	確認
First constituent of prime compound is same as second constituent of target compound and with same reading	全国	安全
First constituent of prime compound is same as first constituent of target compound but with different reading	横顔	横断
First constituent of prime compound is same as second constituent of target compound but with different reading	楽園	音楽
Non-word prime to target compound	績晩	新鮮
Neutral condition	# #	現実

The presentation procedure for Experiment 2 was similar to that of Experiment 1, except that one group of participants responded to the prime by naming it, and another group made a lexical decision for the prime.

There were some differences in the pattern of priming between the two groups in the results of Experiment 2, but these were not significant. Overall, the pattern of priming was similar to that found in Experiment 1 of significantly faster reaction times when the prime was the first constituent of the target compound compared to the second constituent condition, irrespective of whether the reading of the kanji was the same or different. Thus, Hirose (1992) claimed that it is semantic rather than phonological information from the activation of the prime that facilitates the recognition of the compound word.

Table 5.3

Prime-Target Relation Conditions in Hirose (1992) Experiment 3

Prime-target relationship conditions	Prime	Target
Prime is the first constituent of few compound words	節	節約
Prime is the first constituent of many compound words	学	学歴
Neutral condition	#	天候

The notion of semantic links was investigated further in the third experiment that asked whether there would be differences in facilitation when the prime kanji is the first constituent of many compound words compared to instances where it is the first constituent of few compounds. Accordingly, Experiment 3 consisted of a comparison of three prime-target relationships, which are presented in Table 5.3. The result of this experiment was that both related conditions showed facilitation compared to the neutral condition, but the facilitation was greater when the prime was the first constituent of few compound words.

It is on the basis of these experimental results that Hirose (1992) has claimed that the mental lexicon for compound words is structured so that words sharing the same first kanji are linked in clusters, with the first kanji serving as a retrieval cue, but words sharing the same kanji as a second constituent are not.

5.1.3.2 Distributed activation model for Japanese: Ijuin, Fushimi, Patterson, and Tatsumi (1999)

Although it does not address morphological aspects relating to the representation of two-kanji compound words, Ijuin, Fushimi, Patterson, and Tatsumi (1999) have implemented a connectionist model of word naming for two-kanji compound words, as shown in Figure 5.2, which has similarities with the Seidenberg and McClelland model (1989) discussed in Section 4.3.1.3.[60]

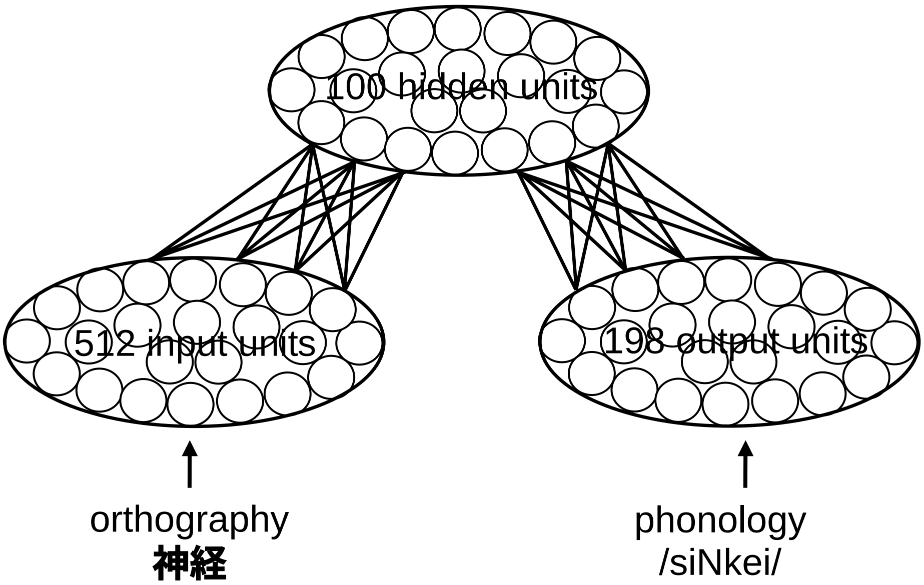


Figure 5.2. The network architecture of Ijuin, et al.'s (1999) connectionist model of word naming for two-kanji compound words. (Based on Ijuin, et al. 1999, p. 272).

This network consists of three layers of units. On the input layer, which has 512 units, each kanji of a compound word was represented as a 16 x 16 grid pattern, and on the output layer, which has 198 units, there were phonological codes for the two-kanji compound words. Ijuin, et al. (1999) report that after training this network for 900 epochs on a corpus of 4,136 two-kanji compound words, it showed frequency and consistency effects, as well as interaction between these, that were generally comparable to those observed by Fushimi, Ijuin, Patterson, and Tatsumi (1999) for Japanese skilled readers, although performance for nonwords (novel strings) was poorer.

While this model does not consider the representation of two-kanji compound words in terms of their morphological structure, it is an important demonstration of the application of the connectionist approach to the Japanese writing system, and as Ijuin, et al. (1999) comment, highlights the fact that,

> the precise method of implementation (in terms of factors such as the size of the training corpus, the coding system, and so on) requires careful consideration depending on characteristics of a particular writing system. (p. 278)

5.1.3.3 Decomposed storage: Zhang & Peng (1992)

Zhang and Peng (1992) investigated the issue of morphologically decomposed representation for Chinese words, with lexical decision task experiments that manipulated word frequency, character frequency, and morphological structure for the compound word stimuli.[61]

In their short discussion of morphological structure in Chinese, Zhang and Peng (1992) claim that there are two basic kinds of relationship between two morphemes. The first kind, which they name coordinative words, consists of two constituents of equal importance. Judging from the stimuli listed in their appendix, these would appear to mainly correspond to the synonymous pairs discussed in Chapter 3. The second kind is modifier words, for which they claim the second morpheme is more important.[62]

In their first experiment, conducted in two parts, Zhang and Peng (1992) investigated whether coordinative words are morphologically decomposed in lexical access. Maintaining the word frequency constant in all conditions, in the first part of the experiment Zhang and Peng manipulated the character frequency, either high or low, of the first constituent of the compound words, and in the second part they manipulated the character frequency of the second constituent. The results for both parts of this experiment was that reaction times were faster in high-frequency conditions compared to the low-frequency conditions. Reasoning that with word frequency held constant, the differences between the high-frequency and the low-frequency conditions must be due to differences in the rates of activation for the constituent characters, Zhang and Peng offer these results as support for the position that compound words are morphologically decomposed and accessed via the constituents. They also suggest that same pattern over the high- and low-character frequency conditions for both the first and second constituents indicates that both constituents are of equal importance for coordinative words.

Zhang and Peng's (1992) Experiment 2 repeated the experimental manipulations of Experiment 1 for modifier words, although this time, no significant differences in reaction times were found. However, Zhang and Peng claim that there was a significant difference in errors between the two parts of the experiments, with these being very low for the second constituent (Exp.2, part 2), and that this supports the decomposition position only for the second character in the case of modifier words.

Based on these results, Zhang and Peng (1992) claim that two-character words are stored in decomposed form in the mental lexicon.[63] Furthermore, they suggest that two-character compound words are organized according to information relating to morphological structure, which includes word type, word class of the characters, and semantic distance and strength between component characters in the lexicon.

However, Taft, Huang, and Zhu (1994) observe that "the difference in response pattern between coordinative and modifier words was not as marked as Zhang and Peng's conclusion warrants" (p. 63). Taft et al (1994) sought to expand on Zhang and Peng's (1992) findings. They duplicated the result that responses to high-high character frequency compound words were faster than to high-low character frequency compound words.

5.1.3.3 Dual-route model of morphological processing for Chinese: Liu & Peng (1997)

Liu and Peng (1997) suggest that Zhang and Peng (1992) were wrong to stress the existence of morpheme units to the exclusion of whole word entries, since they also found word frequency effects. Liu and Peng (1997), therefore, examined the time course for lexical access for Chinese two-character compound words, by contrasting semantically transparent and opaque compounds over three different short SOAs of 143ms, 86ms, and 43 ms.

The aim of their first experiment was to investigate whether the meanings of constituent morphemes in opaque compounds would be accessed in processing the compound word, and if so, whether this would be before or after accessing the whole-word. There were six stimulus conditions for this experiment. In the first three conditions, the opaque compound prime was either semantically related to the target compound word at the whole-word level, the first constituent of the opaque compound prime was related to the target compound, or the second constituent of the opaque compound prime was related to the target compound. These three experiment conditions were matched with three control conditions, in terms of word frequency, component character frequency and number of strokes. The SOA in the first experiment was 143ms. The finding of this experiment was priming in all three experimental conditions. That is, not only did the related whole-word prime the target word, but the first constituent and the second constituent also facilitated the recognition of the target word, suggesting that even with opaque Chinese compound words morphemes have an effect in lexical access.

In a second experiment with an SOA of 43ms, however, priming was only obtained in the whole-word related condition, and not for the morpheme conditions. In a third experiment, semantically transparent and opaque compounds were compared as primes to target words related to one constituent (the distinction between first and second constituent conditions was collapsed, with the whole-word related condition omitted) using an SOA of 86ms. Although a priming effect was obtained for transparent compounds, no priming was found for opaque compounds. Based on the results of the first and third experiment, Liu and Peng (1997) suggest that the activation of morpheme meaning in opaque compounds occurs somewhere between 86ms to 143 ms.

In their discussion of the results, Liu and Peng (1997) also propose a morpheme level of mental lexical representation. However, this morpheme level is not seen as containing information concerning meaning, orthography and phonology, but is treated rather as a link connecting elements of meaning and form.

5.1.3.5 Multilevel interactive-activation model for Chinese: Taft & Zhu (1997)

As mentioned in Chapter 4, Taft (1991, 1994) has proposed the multilevel interaction-activation framework as a model for thinking about and accounting for experimental findings concerned with the morphological processing involved in recognizing polymorphemic words in English. Taft and others (Taft, Liu, & Zhu, 1998; Taft & Zhu, 1994, 1997a, 1997b) have subsequently expanded on this model to account for visual word recognition for Chinese. A version of this interactive-activation model for Chinese is shown in Figure 5.3.

The lowest level representing features corresponds to the strokes of Chinese characters. Above the feature level comes the sub-lexical level, which is where radicals, both semantic and phonetic markers, are incorporated into the model. The next level is the morpheme level, at which single kanji characters are represented. At the highest level come two-character Chinese polymorphemic words. We shall briefly look at the experiments carried out by Taft and Zhu (1995, 1997a, 1997b) to support the existence of the radical level and the lexical levels of morpheme, single kanji characters, and polymorphemic words, two-character compound words.

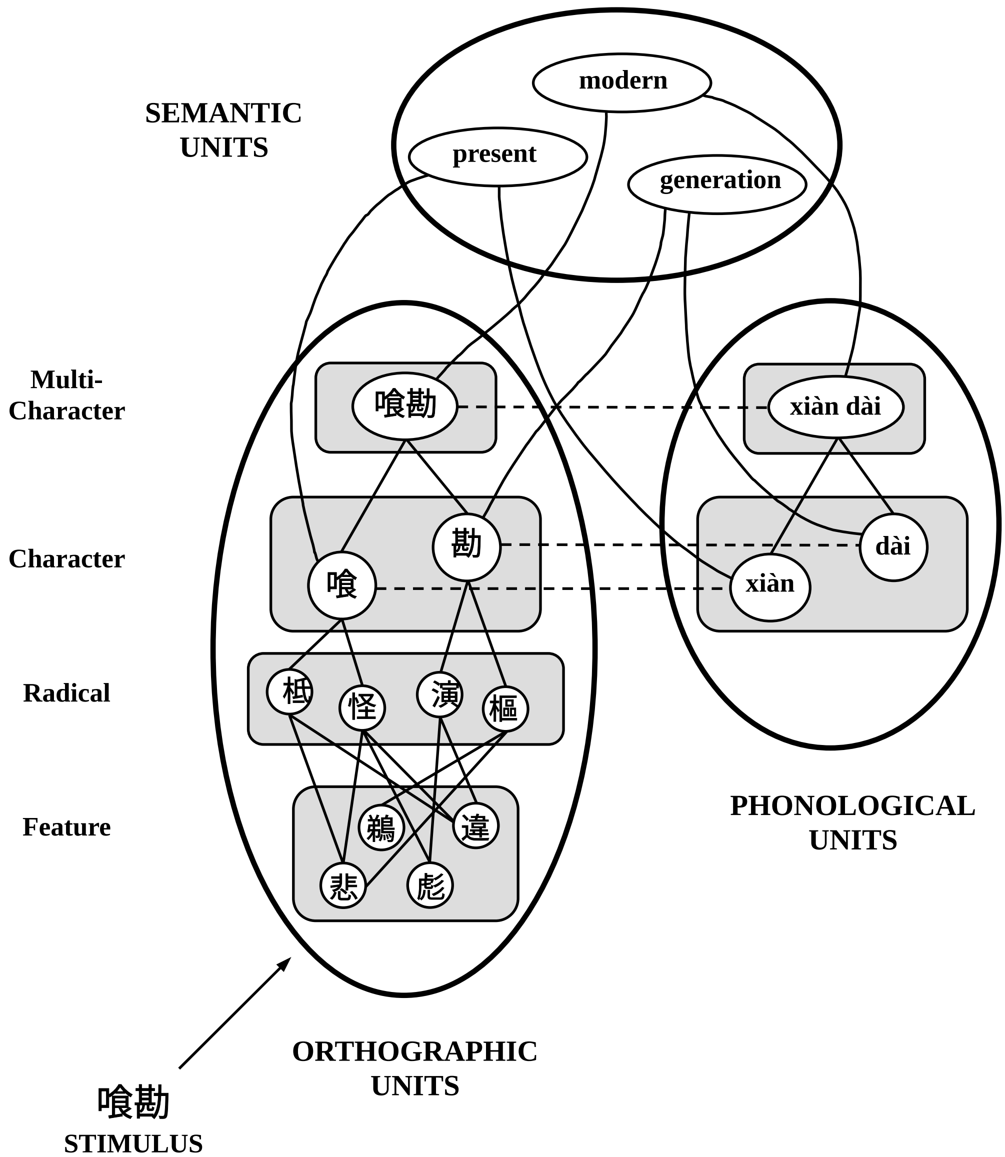


Figure 5.3. Taft and Zhu's (1997) multilevel interactive-activation framework for conceptualizing the lexical processing of Chinese words. (Based on Taft, Zhu & Peng, 1999).

Taft and Zhu (1997a) argue for the existence of a sub-morphemic or radical level within the framework. The basic assumption for their series of experiments was that if there is a level where representational units for radicals are activated in the recognition of the kanji character, then we should expect the frequency with which a radical is used in the language to have an impact on the recognition times for the kanji character.

In Experiment 1, then, there were four conditions for a character decision task for horizontal characters, that is, characters consisting of a 偏 *hen* 'left-side radical' plus 旁 *tsukuri* 'right-side radical' structure, based on the frequency of the radicals, giving high-high, high-low, low-high, and low-low radical frequency conditions. The experiment found an effect for the frequency of the right-side radical, with characters with high-frequency right radicals being responded to faster than characters with low-frequency right radicals, but no effect for the left radical frequencies. In explanation of this, Taft and Zhu (1997a) suggest that although both radicals are involved in the activation stage, there is a priority in processing from left to right, which obscures the effect of the left radical frequency.

In a second experiment, Taft and Zhu (1997a) considered the effect on character recognition of radical frequencies in terms of positional sensitivity, that is, the differences in frequencies for a radical appearing on the left-side of a character and appearing on the right-side, by comparing position-sensitive frequency with total frequency. The result of this experiment was that there was only an effect for the position-sensitive frequency condition, leading Taft and Zhu (1997a) to comment that not only do radical level representations exist, but "it seems that the radical level is somehow sensitive to the position in which the radical occurs" (p. 769).

Although Taft and Zhu (1997a) claimed a priority in the processing of radicals from left to right, a study by Masuda and Saito (December, 1997) questions this. Masuda and Saito (December, 1997) classified the radicals that can appear in horizontal characters into 'left-anchored' radicals appearing only on the left, 'right-anchored' radicals appearing only on the right, and 'free-floating' radicals that can appear in either position. The radicals are further classified in terms of the number of different radicals that occupy the other position, which they refer to as the number of companions. For instance, they treat the water radical in 海 *umi* 'sea,' 池 *ike* 'lake,' 河 *kawa* 'river,' etc, as a left-anchored radical with many companions, but 客 *kyaku* 'guest' in 額 *gaku* 'amount; tablet' as a left-anchored radical with few companions. According to the interactive-activation framework, radicals with more companions should receive more feedback activation from units at the character level in the recognition process.

Masuda and Saito (December, 1997) sought, therefore, to investigate the activation of both left and right radicals by manipulating the number of companions for radicals in an identification and report task for briefly presented kanji characters.[64] The premise for this experiment was that if a radical with many companions receives more feedback activation from character units at the next higher level, then report for such radicals should be more accurate than for radicals with few companions. Also, if both left and right radicals are activated in the recognition of briefly presented kanji, then accuracy of report will depend on the number of companions, but if processing is serial from left to right, then accuracy should be higher for left radicals, because the right radical is unlikely to be activated under such brief presentation conditions.

The results of this experiment showed that high frequency characters were more accurately reported than low frequency characters, which is in line with other research. The results also showed that low frequency characters with many companions were reported more accurately than those with few companions, confirming the hypothesis of greater feedback activation from the character level. This effect of number of companions was found for both left and right radicals, indicating that both are activated in the recognition process, even at very short SOAs of 40 ms, and questioning the notion that radicals are only processed from left to right.

The relationship between the morpheme and the polymorphemic word levels is considered in three studies by Taft and others (Taft and Zhu, 1995, 1997b; Taft, Liu, and Zhu, 1998). Much of this discussion has focused on the status of two-character words that Taft and Zhu (1995) refer to as 'binding words.' A binding word is defined as a word transcribed with two Chinese characters that are only used in writing that particular word, and would correspond to jukujikun words in Japanese.[65] Taft and Zhu (1995) differentiated these from polymorphemic words, formed by the combination of two morphemes, which could be either free or bound morphemes used in several compound words.

In the study, which directly addressed the question of whether both a morpheme and a polymorphemic word level are necessary, Taft and Zhu (1995) attempted to distinguish between four possible variations of the model based on which levels of representation are incorporated. A naming task was used, in which participants were asked to name single characters that were either the first or second character of a binding word, to investigate whether the naming was being based on the phonology of a whole-word representation or that of a bound element alone. The finding of this experiment was that it took longer to name the second character of a binding word than the first. Reasoning that the naming responses must, therefore, have been to be based on whole-word representations, Taft and Zhu (1995) concluded that the two elements of binding words are not morphemic, having no independent lexical representations, and accordingly binding words are monomorphemic. In the case of bound morphemes that appear in several words,[66] however, the results of a third experiment showed no significant differences in naming latencies for the first or second elements. These results prompted Taft and Zhu (1995) to remark that "morphemes are indeed represented in the lexicon, whether they be free or bound," but for "a unit to be considered to be a morpheme, it seems that there needs to be some sense of its semantic function" (p. 310).

Taft and Zhu (1997b) also considered the question of whether both a morphemic and a polymorphemic word level are necessary for the model. Here, the naming task was somewhat unusual in that participants were asked to name only the first element of a two-character compound word. The logic for this experiment was that if there is a polymorphemic level of representations, then the presentation of a second element from a compound word would prime the naming of the first element of the word, but not if there are only morphemic level representations. The result of this study was that the second character prime did facilitate naming responses for the first element, both for monomorphemic binding words and polymorphemic words, which suggests that there is a polymorphemic level representation.

5.1.3.5 Companion-activation model: Saito (1997)

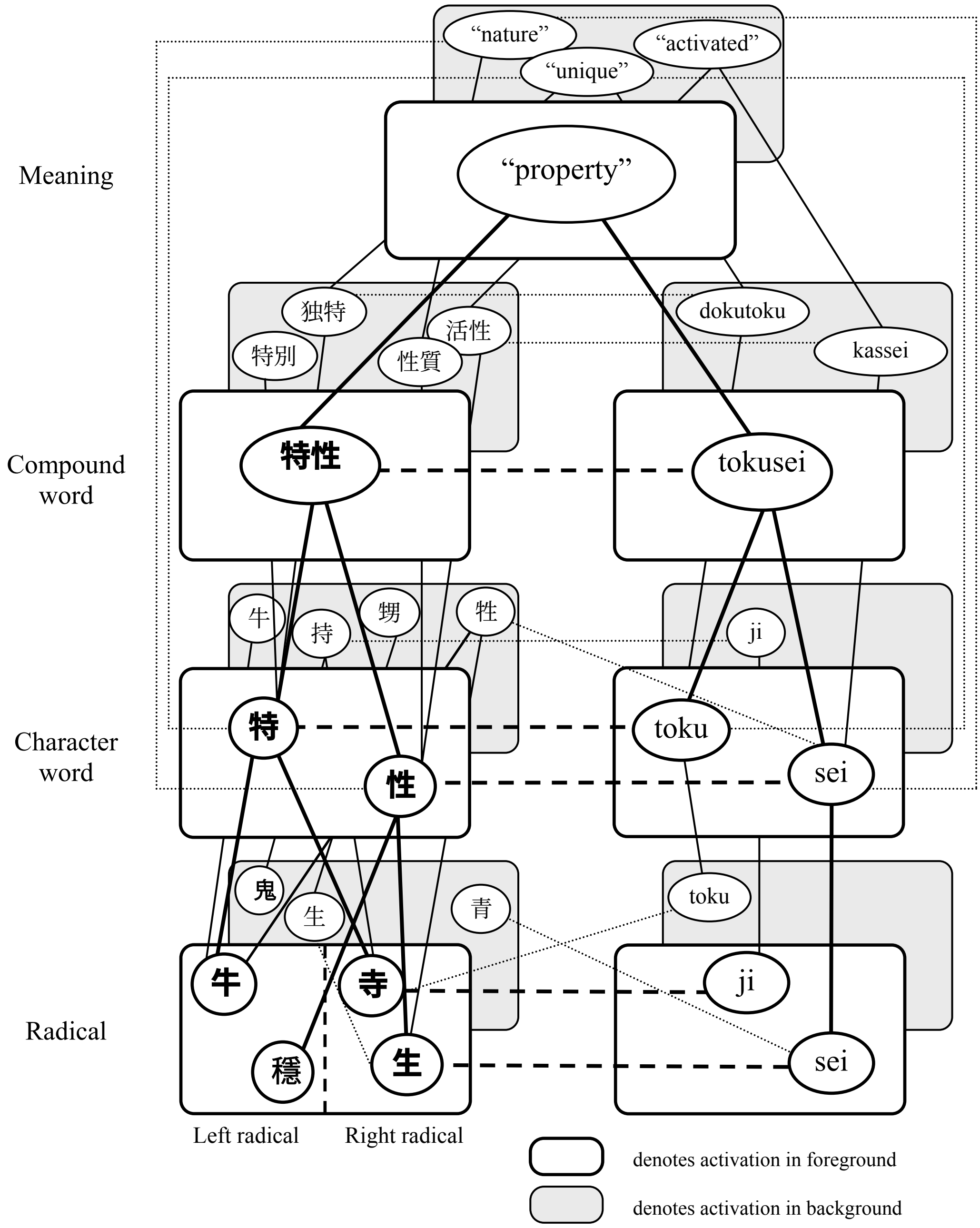


Figure 5.4. Saito’s (1997) companion-activation model (CAM). (Based on Saito, Masuda, & Kawakami, 1998, p. 329).

Although, as we have already noted, Ijuin, et al. (1999) have proposed a distributed connectionist model of word naming for two-kanji compound words, a number of models involving activation mechanisms, as local connectionist models similar to the multilevel interactive-activation framework for Chinese (Taft & Zhu, 1997), have also been proposed for the Japanese mental lexicon. One such model is the companion-activation model advocated by Saito (1997), which is presented in Figure 5.4.

Architecturally similar to the multilevel interactive-activation framework proposed by Taft and Zhu (1997), this model consists of various levels, including radical level representations, single character level representations and two-kanji compound word level representations linked in a hierarchy to semantic representations. However, as Saito, Masuda, and Kawakami (1998) point out, in the companion-activation model, the activation of constituents at the radical and character word levels of the model is separated into two kinds—foreground activation, for the target kanji or word, and background activation for shared companion radicals or words. This notion of background activation is proposed to account for differences in the number of 'neighbors' or radical companions (Saito, Kawakami, & Masuda, 1995) and for differences in the strengths of connections.

5.1.3.7 Multilevel interactive-activation model for Japanese: Tamaoka & Hatsuzuka (1998)

Another local connectionist model for the Japanese mental lexicon is the multilevel interactive-activation version for Japanese proposed by Tamaoka and

Hatsuzuka (1998), depicted in Figure 5.5.

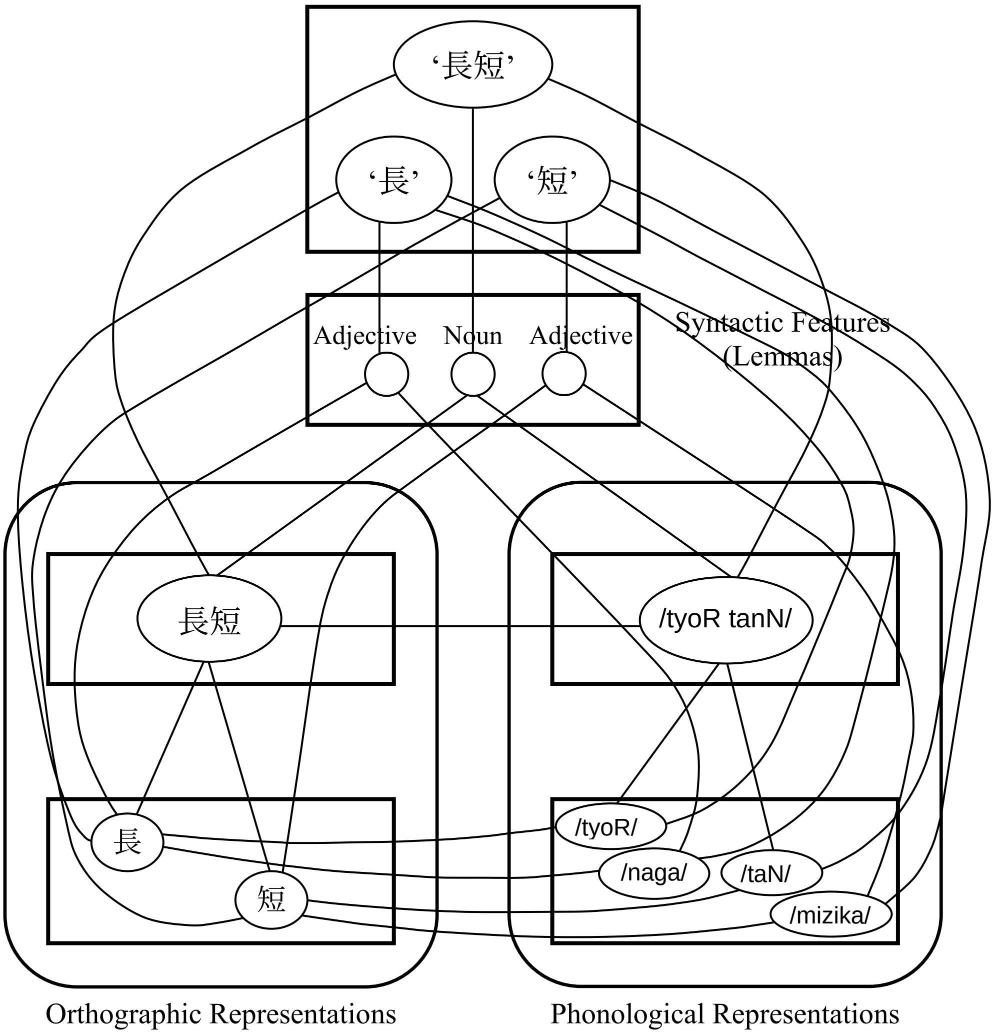


Figure 5.5. Tamaoka and Hatsuzuka’s (1998) version of the multilevel interactive-activation framework for Japanese. (Based on Tamaoka & Hatsuzuka, 1998, p. 312).

Although this is also essentially a Japanese version of the model proposed for Chinese by Taft and Zhu (1997), there is a variation with this version, because Tamaoka and Hatsuzuka (1998) incorporate what they refer to as syntactic feature representations, or lemmas, in their model, which serve as a means of representing word-class information within the model.

5.1.3.8 Lemma unit version for Chinese: Taft, Liu, & Zhu (1999)

Taft, Liu, and Zhu (1999) have recently proposed an important modification to the multilevel interactive-activation framework presented in Figure 5.3. Essentially, the modification is to replace the whole-word level orthographic and phonological representations with 'lemma units,' abstract modality-free units, which function to link orthographic units, phonological units and semantic units. Although Taft, Liu, and Zhu (1999) characterize the lemma units as connections or way-stations that develop when semantic information regularly co-occurs with form information, there are similarities with Levelt's (1989) notion of lemmas[67] and the 'concept nodes' in the model of morphological processing proposed by Schreuder and Baayen (1995), which we discussed in Chapter 4

Taft, Liu, and Zhu's (1999) revision was motivated by two unsatisfactory features of the previous Chinese multilevel interactive-activation framework model. The first was the redundancy in having both orthographic and phonological representations at both the constituent and the compound word levels. This redundancy is eliminated with the lemma units, which mediate links between orthographic and phonological access representations. The second feature was the problem of homographs, that is, orthographic units associated with more that one meaning. Again, incorporating lemma units into the model can overcome this problem, for a homograph orthographic unit can be linked to separate meanings via different lemmas.

These lemma units are similar to the morpheme representations proposed by Liu and Peng (1997) that were also envisaged more as links connecting elements of meaning and form rather than specific representations containing semantic, orthographic, or phonological information. However, in the Taft, Liu, and Zhu (1998) model, there are no explicit representations of morphemes and polymorphemic words, simply lemma units that emerge from the regular co-occurrence of orthographic forms with meaning.

These lemma units also provide an effective method of expressing the contrasts between compound words of varying semantic transparency, which Zwisterlood (1994) sought to capture in the mental lexicon structure that she proposed for Dutch compound words. Zwisterlood (1994) suggested that all compounds would be represented as morphologically complex regardless of semantic make-up, with distinctions between truly opaque compounds and fully or partially transparent compounds defined in terms of whether the semantic representations for the compounds are linked or not to the semantic representations for the constituent elements. With lemma units mediating the links between semantic representations and orthographic representations, such distinctions can be modeled easily.

Schreuder and Baayen (1995) point out that their model of morphological processing assumes that semantic information is stored only once, with the concepts nodes functioning as a means of differentiating and addressing concepts. Accordingly, semantic transparency for a compound is defined as the degree of overlap between the sets of semantic primitives activated by the constituent elements and the set by the compound word itself. In this way, morphological relations are expressed in terms of the shared semantic and syntactic properties that underlie morphological families, rather than in spatial terminology (such as 'clusters').

Taft, Liu, and Zhu (1999) draw on this notion in their account of differences in priming for semantically opaque and transparent compound words, suggesting that activation in the lemmas can be reset. When an opaque compound word is presented, because there is little or no overlap between the meaning of the compound word and the meanings of the elements, activation for irrelevant lemma units is reset to baseline levels. However, in the case of transparent compound words, activation is maintained which can serve to facilitate related words.

The modification of the model to replace explicit morpheme and polymorphemic word representation units with abstract lemma units to connect semantic units to orthographic and phonological representations clearly has certain advantages over the previous model.

5.1.3.9 Lemma unit model for Japanese: Joyce (1999; in press)

Joyce (1999, in press) has proposed adapting for the Japanese mental lexicon the recently modified Chinese version of the multilevel interactive-activation framework advocated by Taft, Liu, and Zhu (1999).

As noted in the previous section, Taft, Liu, and Zhu (1999) have recently modified earlier versions of the model which assumed a two-character compound word representation level, by replacing this with lemma units—abstract modality-free units mediating the links between orthographic, phonological, and semantic units—in order to overcome problems associated with representational redundancy, homographs, and varying degrees of semantic transparency. As these are also problematic aspects for Japanese models, the adaptation to Japanese of the lemma unit is undoubtedly an attractive proposal.

The proposal to adopt a Japanese version of the lemma unit model was also motivated and supported by the results from the two constituent-morpheme priming experiments that are presented in the remainder of this chapter which were conducted to investigate the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon from a morphological perspective.

5.2 Constituent morpheme priming for two-kanji compound words Joyce (1999, in press)

As we noted in Chapter 4, the representation of morphological information is a fundamental issue for all models of the mental lexicon (Feldman, 1995; Jarema, Kehayia, & Libben, 1999; Sandra & Taft, 1994; Taft, 1991). This is true not only because of the vast numbers of polymorphemic words that exist in all languages and the relative ease with which language users handle both existing and novel forms (Sandra, 1994), but also because the issue has important implications for processing. The representation and organization of lexical information in the mental lexicon directly determines the nature of lexical retrieval—whether search (e.g., Forster, 1976) or activation (e.g., Taft, 1991, 1994) mechanisms are assumed, as well as the extent of morphological involvement.

As we have seen in the brief review of studies concerned with two-kanji compound words, both search mechanisms and activation mechanisms have been suggested for the Japanese mental lexicon. Hirose (1992, 1994, 1996), for example, has evoked search mechanisms in his hypotheses that compound words is structured so that words sharing the same first kanji are linked in clusters, with the first kanji serving as a retrieval cue, but that words sharing the same kanji as a second element are not. In contrast, Joyce (1999, in press) has proposed adopting the recently modified version of the multilevel interactive-activation model proposed by Taft, Liu and Zhu (1999) for Japanese.

5.2.1 Purpose

As any model of the Japanese mental lexicon must in some way capture the morphological relations that exist between polymorphemic words, one may ask how well these two proposals cope with the diversity inherent in the morphological structure of two-kanji compound words, which was discussed in Chapter 3.

Although about nine main principles or relationships are generally accepted as underlying the formation of two-kanji compound words (see Nomura, 1988; also, Kageyama, 1982; Tamamura, 1985), this constituent-morpheme priming study will focus on the five principles that were used as experimental conditions in the two conducted experiments.[68] The first three principles, which are syntactic in nature, are modifier + modified (e.g., 山 'mountain' + 桜 'cherry' in 山桜 *yamazakura* 'mountain cherry'), verb + complement (e.g., 登 'climb' + 山 'mountain' in 登山 *tozan* 'mountain climbing,' and complement + verb (e.g., 外 'outside' + 食 'eat' in 外食 *gaishoku* 'eating out'). The last two principles, which are semantic in nature, are associative pairs (e.g., 男 'man' + 女 'woman' in 男女 *danjo* 'men and women') and synonymous pairs (山 'mountain' and 岳 'mountain' in 山岳 *sangaku* 'mountains' (Joyce & Ohta, 1999).

From a representational perspective, such diversity clearly poses problems for Hirose's (1992) hypothesis that compound words are linked in clusters based on the first kanji. For instance, if clusters are only based on a shared first kanji, then semantically-related compounds sharing a common second kanji, such as modifier + modified compounds like 山桜 *yamazakura* 'mountain cherry' and 夜桜 *yozakua* 'cherry blossoms at night, will not be linked. Another difficulty arises from the reversed syntactic structures of verb + complement and complement + verb. Although Hirose does not discuss native Japanese compound words like 山登り, presumably the clustering based on the first kanji extends to these compound words as well. However, this would lead to a strange situation where 登山 and 山登り would not be linked in the mental lexicon because of the reversed order of elements, even though they are almost identical in meaning. Even if native Japanese compound words are treated separately, there are still many semantically-related Sino-Japanese compound words, such as 殺人 *satsujin* 'murder' ('kill' + 'person') and 毒殺 *dokusatsu* 'poison' ('poison' + 'kill'), which would not be linked according to

Hirose's hypothesis. On the other hand, the Japanese lemma-unit model has little problem with this morphological diversity. This is because instead of grouping morphologically related compound words in 'clusters,' the relations underlying morphological families are modeled by the connections between representations.

Looking at the results of constituent-morpheme priming, which we discussed in Chapter 4, however, the situation in terms of lexical retrieval is perhaps less clear. The advantage for the first constituent of modifier + modified words in both Greek and Polish reported by Kehayia, Jarema, Tsapkini, Perlak, Ralli, & Kadzielawa (1999) would seem to match Hirose's (1992, Exp. 1) results.[69] However, Monsell (1985), Sandra (1990), and Zwisterlood (1994) did not find a first-constituent advantage in their experiments, for they all obtained similar levels of facilitation in first and second constituent conditions. Although the Japanese lemma-unit model, where retrieval is based on an activation mechanism, predicts similar levels of priming from both constituent kanji, such a finding would be incompatible with the search mechanisms that Hirose has evoked given their reliance on the first kanji as a retrieval cue.[70]

The storage of lexical information and the mechanism of lexical retrieval are mutually defining aspects of the mental lexicon. Given this and the difficulties with Hirose's (1992, 1994, 1996) hypotheses in coping with the diversity in compound-word structure, one may wonder if the pattern of facilitation found by Hirose (1992) would be observed once compound-word structure is considered. As these two proposals make different predictions concerning constituent-morpheme priming, the pattern of facilitation for two-kanji compound words is examined in two priming experiments that control for compound-word morphology by including five word-formation principles as experimental conditions.

5.2.2 Survey 1:
Two-kanji compound word classification survey
Joyce and Ohta (1999)

5.2.2.1 Purpose

Although the word formation principles underlying the morphology of two-kanji compound words discussed in Chapter 3 are fairly well accepted, the task of classifying a given two-kanji compound word under the appropriate problem is not always without problems. For instance, compounds of the form verb + noun can be difficult to classify, for this structure appears in both the verb + object relationship of 登山 and the modifier + modified relationship, as in 寝室 *shinshitsu* '(lit 'sleep' + 'room') bedroom,' which means a room for sleeping in rather than the act of sleeping in a room.

The aim of this survey was, therefore, to obtain evaluations from native Japanese speakers concerning the appropriateness of classifying a corpus of two-kanji compound words according to the five word formation principles shown in Table 5.4 overleaf, in order to control for the diversity in the morphology of two-kanji compound words in the constituent morpheme priming experiments.[71]

Table 5.4

Examples of the five word-formation principles for two-kanji compound words examined in this study

Principle	Pronunciation	Constituent meanings	Compound meaning
Modifier + modified			
山桜	*yamazakura*	'mountain' + 'cherry'	mountain cherry
国道	*kokudō*	'country' + 'road'	national road
Verb + complement			
登山	*tozan*	'climb' + 'mountain'	mountain climbing
殺人	*satsujin*	'kill' + 'person'	murder
Complement + verb			
外食	*gaishoku*	'outside' + 'eat'	eating out
毒殺	*dokusatsu*	'poison' + 'kill'	kill by poison
Associative pairs			
男女	*danjo*	'man' + 'woman'	men and women
左右	*sayū*	'left' + 'right'	left and right
Synonymous pairs			
山岳	*sangaku*	'mountain' + 'mountain'	mountains
変化	*henka*	'change' + 'change'	change

5.2.2.2 Method

A corpus of 1,000 two-kanji compound words was created, based mainly on a list of basic vocabulary for Japanese language teaching (National Language Research Institute, 1984). After collecting all the two-kanji compound words in this basic vocabulary list (excluding proper nouns, compound words involving numbers, and words written with kanji outside the Jōyō kanji list), the set was supplemented with a number of compound words from a kanji dictionary for elementary school students (Ishii, 1996), particularly associative pairs, which were poorly represented in the National Language Research Institute's (1984) list. In total, approximately 3,000 words were tentatively classified according to the 5 word-formation principles. Finally, the corpus of 1,000 compounds was taken from this set by selecting 200 compound words for each word-formation principle.

In order to keep the survey task as simple as possible, respondents were asked to evaluate the appropriateness of classifying a list of words according to a single principle. That is, respondents were asked to rate 100 words as examples of a particular word-formation principle on a 7-point scale, with 1 representing bad examples and 7 good examples. When a respondent evaluated a compound word as a bad example of a particular word-formation principle (i.e., with an evaluation of 1 or 2), they were also asked to suggest an alternative classification from one of the other 4 principles, if they felt one of those was more appropriate.

Ten separate lists of 100 words each (2 lists for each principle) were prepared, and each list received 10 evaluations, making a total of 100 completed questionnaire forms. Seventy native Japanese speakers (average age = 27.8, range 18-56, *SD* 7.2) were asked to complete the 100 forms (each person completed between 1 to 5 lists, with the average number completed being 1.42, *SD* 1.01).

5.2.2.3 Results

The average evaluation scores for each word-formation principle are plotted as cumulative percentage curves in Figure 5.6.

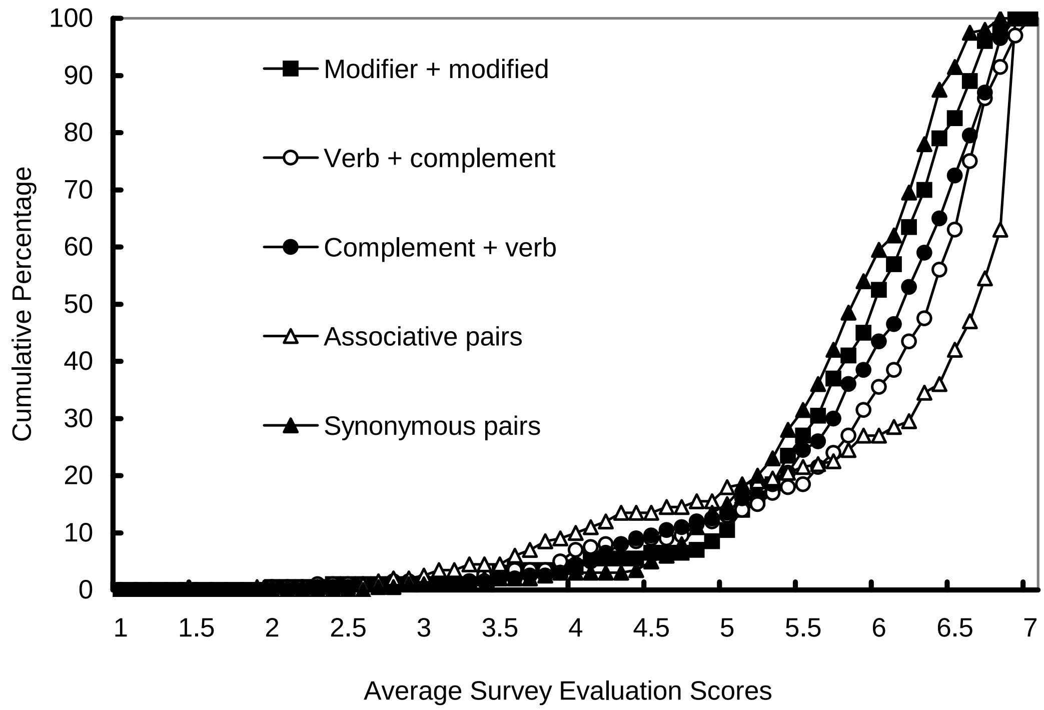


Figure 5.6. Results of classification survey plotted as cumulative percentage curves for word-formation principles.

With this kind of graph, a normal distribution would appear as a straight diagonal line from 0% to 100%, but here the distributions are clearly skewed in favour of higher ratings. This indicates that the majority of these words were highly rated as examples of the relevant principles. For instance, compounds receiving a mean rating of 5.5 or over account for 73% of the modifier +

modified compounds, 81.5% of the verb + complement compounds, 75.5% of the complement + verb compounds, 78.5% of the associative pairs, and 68.5% of the synonymous pairs.

5.2.2.4 Discussion

This classification survey was conducted, because although the word formation principles underlying the morphology of two-kanji compound words are fairly well accepted, the task of classifying a given two-kanji compound word under the appropriate problem is not always without problems. However, although some items received somewhat mixed evaluations, generally there seems to have been fairly good agreement among respondents. For example, although the verb + noun forms of 寝室 'bedroom' and 決心 *kesshin* '(lit. 'decide' + 'heart/mind') determination, resolution' were both tentatively assigned to the modifier + modified principle, 寝室 was rated as a good example with a score of 6.8, whereas 決心 was rated as a bad example with a low score of 2 and 9 out of the 10 respondents suggesting that this would be more appropriately classified as an example of the verb + complement principle.

Based on these results, and on subsequent supplementary surveys, all the two-kanji compound words used as stimulus items in the experiments reported in this thesis had classification evaluation scores of 5.5 or over on a 7-point scale.

5.2.3 Experiment 1: Primed lexical decision task

5.2.3.1 Purpose

The purpose of this experiment is to investigate whether the pattern of facilitation obtained by Hirose (1992) would be found once the morphological structure of the two-kanji compound words was controlled using five word-formation principles.

5.2.3.2 Method

Participants. Forty-two native Japanese students (average age 20.1, SD = 3.25) of the University of Tsukuba participated in the experiment as volunteers.

Design and Materials. A 3 x 5 two-factor design was used, with both factors as within-subject variables. The three prime conditions are first constituent, second constituent, and unrelated prime, as shown in Table 5.5.

Table 5.5

Prime-Target Conditions in Experiment 1

Condition	Prime kanji	Target compound
First constituent	外	外国
Second constituent	国	外国
Unrelated kanji	勢	外国

Table 5.6

Mean Classification Scores and Familiarity Scores for Target Compound Words Used in Experiment 1

Word formation principle	Classification score	Familiarity score
Modifier + Modified	6.32 (0.29)	6.47 (0.23)
Verb + Complement	6.56 (0.25)	6.24 (0.30)
Complement + Verb	6.51 (0.24)	6.13 (0.27)
Associative Pairs	6.85 (0.09)	6.29 (0.33)
Synonymous Pairs	6.23 (0.33)	6.42 (0.22)

Note. Both surveys used a 7-point scale; in the case of the classification survey, one indicated bad examples and seven indicating good examples, and in the familiarity survey, one indicated low familiarity and seven indicating high familiarity words. The figures in parenthesis are the standard deviations.

Based on the results of the two-kanji compound word classification survey described in the previous section, the five word-formation conditions are modifier + modified, verb + complement, complement + verb, associative pair, and synonymous pair. For each of the five word-formation principles, 18 compound words were selected with evaluation scores of 5.5 or over on a 7-point scale for both a classification evaluation score and a familiarity evaluation score. The two-kanji compound word targets for Experiment 1 are shown with classification and familiarity scores in Appendix A. The mean scores for both criteria over the five word-formation principles are shown in Table 5.6.

For the lexical decision task, non-word combinations of two kanji were generated from the word classification survey corpus by randomizing the second constituent kanji. These combinations were checked to eliminate any real words generated by chance by consulting an electronic dictionary (Microsoft/ Shogakukan, 1997), and based on the results of a survey[72] to control for word-like-ness, 90 nonword items were selected.

To counterbalance the target compound words over the three prime conditions, three presentation lists were prepared. Participants were assigned evenly to these lists, which were randomized for each participant. After first excluding kanji used in compound and non-word stimulus items, the kanji for the unrelated condition were randomly assigned from the most frequent 1,000 Jōyō kanji (National Language Research Institute, 1976).[73]

Apparatus. Super Lab Pro (Cedrus Corporation, Version 1.05), running on a personal computer (Dell, Dimension XPS D333), controlled the presentation of stimulus items and recorded lexical decisions collected via a response box (Cedrus Corporation, RB-600). Stimuli at a font size of 36 points were displayed on the computer screen at a viewing distance of approximately 50 cm.

Procedure. The procedure used for this experiment is very similar to that of Hirose's (1992) Experiment 1, with a stimulus onset asynchrony (SOA) of 3000 ms, as shown in Figure 5.7. The only difference is that participants were asked to only look at the prime, rather than name it, which could bias them towards the activation of phonological information. At the start of a trial, a plus symbol (+) was displayed in the center of the screen as a fixation point for 1,000 ms. After this, a blank screen for 500 ms was followed by a prime kanji for 1,000 ms. Following a second blank screen for 500 ms, two plus symbols (+ +) as fixation points for the target compound words were displayed for 1,000 ms. After a third blank screen for 500 ms, a target compound word was displayed until the participants made a lexical decision by pressing a key on a response box.

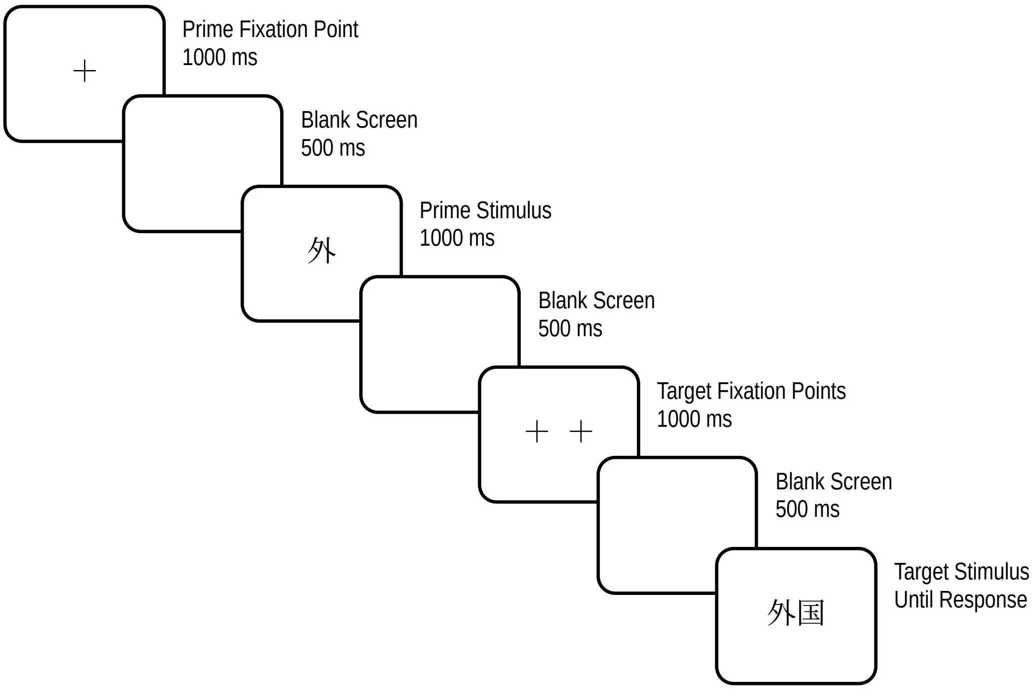


Figure 5.7. Presentation sequence during a trial in Experiment 1.

Participants were instructed to press a green button for a compound word and a red button for a non-word as quickly and as accurately as possible. The instructions to the participants were printed on a card, which they were asked to read, as shown in Appendix B. After briefly repeating the instructions verbally to confirm that they had been understood, there was practice session of ten trials, and the whole experiment took about 20 to 25 minutes to complete.

5.2.3.3 Results

Analyses of variance (ANOVAs) were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). Error responses were excluded from the analysis of reaction times. Responses were also removed if the standard score for a reaction time was outside the range of +/- 2.5 calculated from the mean response time for a given participant. In all, 6.4% of the 3780 responses were excluded because of these procedures. Table 5.7 presents the mean reaction times and error rates for Experiment 1.

Error analysis. The overall error rate was low at 4.4%. Although there were significant main effects of prime in both the subject and item analyses, F_1 (2, 82) = 27.00, $p < .0001$; F_2 (2, 170) = 30.71, $p < .0001$, the main effect of word-formation principle was only significant in the subject analysis, F_1 (4, 164) = 3.26, $p < .013$, with no significant interaction in either analysis.

Planned comparisons using Tukey's Honestly Significant Difference (HSD) test (all HSDs reported in this thesis are at the 0.05 level) for the main effect of prime revealed that although errors in both the first constituent and the second constituent conditions were significantly lower than in the unrelated condition, the differences between the first constituent and second constituent conditions were not significant, HSD = 0.134 and HSD = 0.288 for the subject and item analyses respectively. Planned comparisons for the main effect of word-formation principle revealed that errors in the modifier + modified and the verb + complement conditions were significantly lower than in the complement + verb condition, HSD = 0.159.

Table 5.7

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 1

Word Formation Principle		Error rate	Reaction times		RT diff
Modifier + modified					
	First constituent	0.4	529	(70)	+54
	Second constituent	1.6	544	(77)	+39
	Unrelated	5.6	583	(77)	
Verb + complement					
	First constituent	2.0	539	(81)	+55
	Second constituent	3.2	558	(81)	+36
	Unrelated	11.5	594	(81)	
Complement + verb					
	First constituent	4.0	559	(78)	+48
	Second constituent	2.8	551	(88)	+56
	Unrelated	9.1	607	(87)	
Associative pairs					
	First constituent	3.2	543	(86)	+54
	Second constituent	2.0	539	(79)	+58
	Unrelated	6.7	597	(84)	
Synonymous pairs					
	First constituent	4.0	539	(90)	+63
	Second constituent	2.0	534	(84)	+68
	Unrelated	8.7	602	(94)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unrelated condition reaction time in each word formation principle condition.

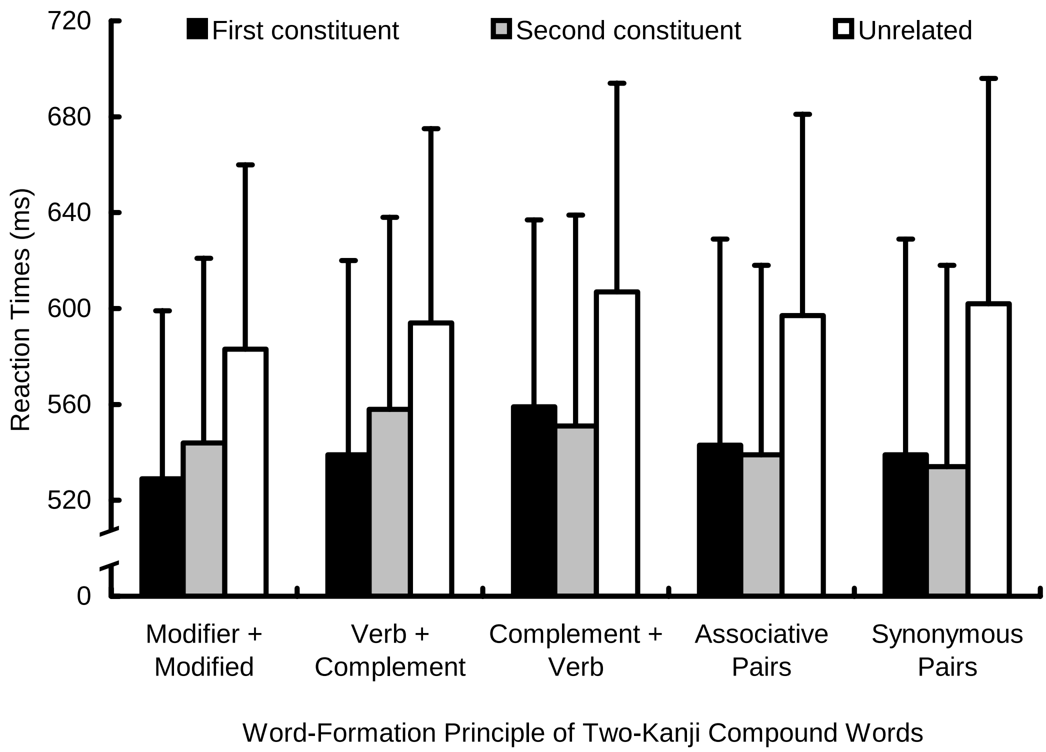


Figure 5.8. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 1.

Reaction times analysis. In the subject analysis, there were significant main effects for prime, F_1 (2, 82) = 102.22, $p < .0001$, and for word-formation principle, F_1 (4, 164) = 5.34, $p < .0001$, with significant interaction, F_1 (8, 328) = 2.45, $p < .014$. In the item analysis, there was only a significant main effect of prime, F_2 (2, 170) = 56.37, $p < .0001$. Reaction times for Experiment 1 are shown in Figure 5.8.

Planned comparisons for the main effect of prime for each word-formation condition in the subject analysis revealed that although reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unrelated condition for all conditions, the only word-formation condition with a significant difference between first constituent and second constituent conditions was the verb + complement condition where the first-element was faster, F_1 (2, 82) = 28.84, $p < .0001$ at modifier + modified, F_1 (2, 82) = 38.53, $p < .0001$ at verb + complement, F_1 (2, 82) = 35.21, $p < .0001$ at complement + verb, F_1 (2, 82) = 35.00, $p < .0001$ at associative pairs, and F_1 (2, 82) = 34.55, $p < .0001$ at synonymous pairs, with HSD = 18.09 in all cases. Planned comparisons for the main effect of prime in the item analysis revealed that reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unrelated condition for all conditions, HSD = 14.51.

Planned comparisons were conducted for the main effect of word-formation principle for each prime condition. In the first constituent condition, reaction times were significantly faster in the modifier + modified condition compared to the complement + verb condition, F_1 (4, 164) = 4.13, $p < .003$, HSD = 20.33. In the second constituent condition, reaction times were significantly faster in the synonymous pair condition compared to the verb + complement condition, F_1 (4, 164) = 3.63, $p < .007$, HSD = 20.33. In the unrelated condition, reaction times were significantly faster in the modifier + modified condition compared to the complement + verb condition, F_1 (4, 164) = 2.99, $p < .021$, HSD = 20.33.

5.2.3.4 Discussion

Experiment 1 sought to examine the pattern of facilitation in constituent-morpheme priming by controlling the word-formation principle of two-kanji compound words. The results showed priming for both the first and the second constituent conditions. However, only the verb + complement condition matched the prediction from Hirose's (1992, 1994, 1996) hypotheses of significantly faster reaction times for the first constituent condition. In line with the prediction from the Japanese lemma-unit model, the differences between the constituent primes were not significant in the other four word-formation conditions. The question of how the Japanese lemma-unit model could possibly account for the verb + complement pattern of priming will be taken up later.

While the subject analyses indicated an effect of word-formation principle, no effect was indicated by the item analyses. Moreover, planned comparisons failed to show a consistent pattern in the reaction times over all three prime conditions, which one would expect if word-formation principle were indeed a variable influencing lexical retrieval.

5.2.4 Experiment 2: Automatic primed lexical decision task

However, one aspect of the present experiment that needs to be investigated further is the SOA of 3000 ms. Tamaoka and Hatsuzuka (1998) have criticized Hirose's (1992) experiment for its very long SOA, claiming that at such a long delay participants may adopt strategies in their lexical-decision-making. If this were so, then the results of Hirose (1992) and the present experiment might not be a true reflection of the time required for the lexical retrieval of the target stimulus items.

Tamaoka and Hatsuzuka (1995) avoided the problem of strategy-adoption arising from a long SOA by simply using unprimed lexical decision and naming tasks in their experiments.[74] Treating kanji frequency as a characteristic of the underlying morpheme, their findings of word and kanji frequency effects support the position that the frequency of the morphemic elements, represented by the kanji of a compound word, plays a role in word recognition. However, using an unprimed lexical decision task, as Tamaoka and Hatsuzuka (1995) have done, can tell us nothing about the arrangement of compound word representations in the mental lexicon in terms of their semantic and morphological relationships.

As noted in Chapter 4, the priming paradigm has been extremely influential in the study of word recognition. This is because it represents a potentially powerful tool with which to investigate how words might be related in the mental lexicon (McRae & Boisvert, 1998; Neely 1991). Although the possibility that the participants were able to develop response strategies in making their lexical decision in Experiment 1 because of the long SOA of 3,000 ms is a serious problem, completely rejecting this useful paradigm would seem unwarranted. A more appropriate course of action is to reduce the SOA to levels that eliminate the possibility of participants adopting such strategies. Based on Neely's (1977) seminal study on the distinction between automatic and expectancy-based processing, described in Chapter 4, and on a recent study by McRae and Boisvert (1998) on automatic priming with semantically similar English words, we may expect a short SOA of 250 ms to be free from strategy-adoption.[75]

5.2.4.1 Purpose

Experiment 2 investigates whether the pattern of priming found in Experiment 1 could still be obtained after replacing the very long SOA of 3,000 ms (based on the presentation procedure used in Hirose, 1992, Exp. 1) with a shorter SOA of 250 ms, which has been demonstrated to involve automatic priming.[76] If a similar pattern does emerge at this shorter SOA, then it would indicate that the results of Experiment 1 do reflect the time required to achieve lexical access and make an appropriate response, and that they were not influenced by strategy-adoption by the participants.

5.2.4.2 Method

Participants. Forty-five native Japanese students (average age 19.6, SD = 2.18) of the University of Tsukuba participated in the experiment as volunteers. None of these participants took part in Experiment 1.

Design and Materials. The 3 x 5 two-factor design with both factors as within-subject variables for Experiment 2 is the same as for Experiment 1.

The short SOA of 250 ms made it possible to increase statistical reliability by including more stimulus items without adding to the burden on the participants. Accordingly, the 18 compound words for each word-formation condition in Experiment 1 were supplemented with 12 additional items selected according to the same criteria from the same corpus (Joyce & Ohta, 1999). Thus, there were 30 compound words in total for each word-formation condition in Experiment 2, which are listed in Appendix C. The mean criteria scores for these stimulus items, which are similar to those of the Experiment 1 items, are shown in Table 5.8. The 90 non-words for the lexical decision task in this experiment were the same as those used in Experiment 1.

Table 5.8

Mean Classification Scores and Familiarity Scores for Target Compound Words Used in Experiment 2

Word formation principle	Classification score	Familiarity score
Modifier + Modified	6.27 (0.25)	6.24 (0.37)
Verb + Complement	6.56 (0.21)	6.02 (0.37)
Complement + Verb	6.48 (0.25)	6.00 (0.30)
Associative Pairs	6.83 (0.12)	6.07 (0.40)
Synonymous Pairs	6.27 (0.32)	6.27 (0.32)

Note. Both surveys used a 7-point scale; in the case of the classification survey, one indicated bad examples and seven indicating good examples, and in the familiarity survey, one indicated low familiarity and seven indicating high familiarity words. The figures in parenthesis are the standard deviations.

Three presentation lists were prepared to counterbalance the stimulus items over the three prime conditions. Participants were assigned evenly to these lists, which were randomized for each participant.

Apparatus. This was identical to that used in Experiment 1.

Procedure. The procedure in this experiment was very similar to that used by McRae and Boisvert (1998, Exp. 1), and was as shown in Figure 5.9.

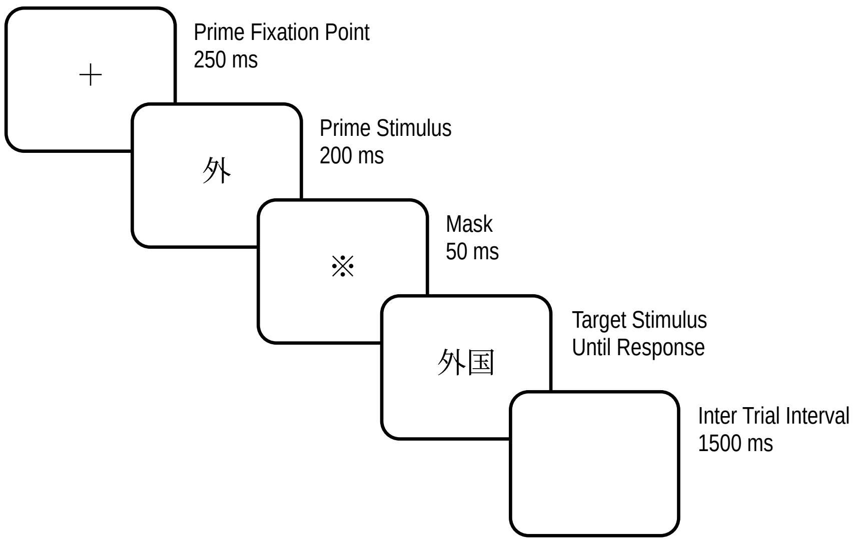


Figure 5.9. Presentation sequence during a trial in Experiment 2.

At the start of a trial, a plus symbol (+) appeared in the center of the screen as a fixation point for 250 ms. This was followed by the single kanji prime, displayed for 200 ms, and then by an asterisk-like symbol (※) as a mask for 50 ms. The target stimulus item was then displayed and remained on the screen until the participant pressed a button on the response box for the lexical decision. There was a 1,500 ms inter-trial interval. Participants were instructed to press a green button for a compound word and a red button for a non-word as quickly and as accurately as possible. The instructions to the participants were printed on a card, which they were asked to read, as shown in Appendix D. After briefly repeating the instructions verbally to confirm that they had been understood, there was practice session of ten trials, and the whole experiment took about 15 to 20 minutes to complete.

5.2.4.3 Results

ANOVAs were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable).[77] Following the same procedures used in Experiment 1, the data was adjusted for error responses and outliners, resulting in 5.9% of the 6750 responses being excluded. Table 5.9 presents the mean reaction times and error rates for Experiment 2.

Error analysis. The overall error rate was very low at 3.4%. Although there were significant main effects of prime in both the subject and item analyses, F_1 (2, 88) = 24.61, $p < .0001$; F_2 (2, 290) = 27.48, $p < .0001$, the main effect of word-formation principle was only significant in the subject analysis, F_1 (4, 176) = 5.79, $p < .0001$, with no significant interaction in either analysis.

Planned comparisons for the main effect of prime revealed that although errors in both the first constituent and the second constituent conditions were significantly lower than in the unrelated condition, the differences between the first constituent and second constituent conditions were not significant, HSD = 0.154 and HSD = 0.216 for the subject and item analyses respectively. Planned comparisons for the main effect of word-formation principle revealed that errors in the synonymous pairs were significantly lower than both the complement + verb condition and associative pairs, and that errors in the modifier + modified condition were lower than for the complement + verb compounds, HSD = 0.203.

Table 5.9

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 2

Word Formation Principle		Error rate	Reaction times		RT diff
Modifier + modified					
	First constituent	2.9	544	(78)	+52
	Second constituent	3.3	549	(72)	+47
	Unrelated	6.2	596	(81)	
Verb + complement					
	First constituent	2.7	537	(82)	+77
	Second constituent	2.0	559	(79)	+55
	Unrelated	5.3	614	(88)	
Complement + verb					
	First constituent	4.9	561	(92)	+62
	Second constituent	4.0	561	(76)	+62
	Unrelated	8.9	623	(80)	
Associative pairs					
	First constituent	4.4	546	(73)	+67
	Second constituent	2.9	558	(91)	+55
	Unrelated	7.6	613	(81)	
Synonymous pairs					
	First constituent	1.8	528	(75)	+89
	Second constituent	0.7	538	(85)	+79
	Unrelated	6.0	617	(94)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unrelated condition reaction time in each word formation principle condition.

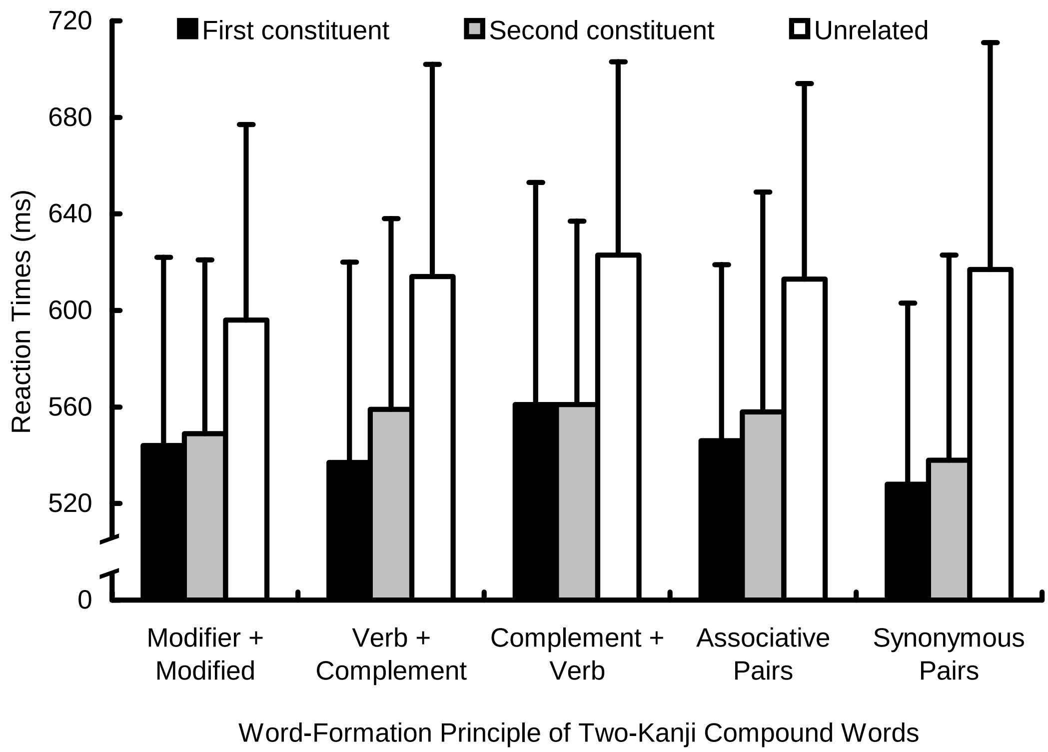


Figure 5.10. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 2.

Reaction times analysis. In the subject analysis, there were significant main effects of prime, F_1 (2, 88) = 162.05, $p < .0001$, and of word-formation principle, F_1 (4, 176) = 7.53, $p < .0001$, with significant interaction, F_1 (8, 352) = 2.92, $p < .004$. In the item analysis, there was only a significant main effect of prime, F_2 (2, 290) = 148.42, $p < .0001$. Reaction times for Experiment 2 are shown in Figure 5.10.

Planned comparisons for the main effect of prime for each word-formation condition in the subject analysis revealed that although reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unrelated condition for all conditions, the only word-formation condition with a significant difference between the first constituent and second constituent conditions was the verb + complement condition where the first-element was faster, F_1 (2, 88) = 36.18, $p < .0001$ at modifier + modified, F_1 (2, 88) = 66.19, $p < .0001$ at verb + complement, F_1 (2, 88) = 45.03, $p < .0001$ at complement + verb, F_1 (2, 88) = 41.79, $p < .0001$ at associative pairs, and F_1 (2, 88) = 79.93, $p < .0001$ at synonymous pairs, with HSD = 17.50 in all cases. Planned comparisons for the main effect of prime in the item analysis revealed that reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unrelated condition for all conditions, HSD = 10.97.

Planned comparisons were conducted for the main effect of word-formation principle for each prime condition. In the first constituent condition, reaction times were significantly faster in the verb + complement and the synonymous pair compounds compared to the complement + verb condition, F_1 (4, 176) = 5.46, $p < .0001$, HSD = 19.32. In the second constituent condition, the reaction times were significantly faster in the synonymous pair condition compared to the verb + complement, the complement + verb and the associative pair compounds, F_1 (4, 176) = 3.51, $p < .009$, HSD = 19.32. In the unrelated condition, reaction times were significantly faster in the modifier + modified condition compared to the complement + verb and synonymous pair compounds, F_1 (4, 176) = 5.00, $p < .001$, HSD = 19.32.

5.2.4.4 Discussion

The results of Experiment 2, with a short SOA of 250 ms, not only match closely the reaction times obtained in Experiment 1, with a long SOA of 3,000 ms, but also provide a very similar pattern of priming, with significant differences between the first constituent and second constituent conditions only in the verb + complement condition. Thus, these results indicate that the reaction times in Experiment 1 were free of participant strategy-adoption, and that the basic pattern of facilitation found in both experiments is due to automatic priming from constituent morphemes.

A main effect of word-formation principle was again indicated in the subject analysis. However, again no consistent pattern was found for this in the significant differences in the planned comparisons. Moreover, the fact that the significant differences also varied across the two experiments cautions against thinking that word-formation principle per se is influencing lexical retrieval.[78] Although Tamaoka and Hatsuzuka (1998) argue that there are differences in processing times for two-kanji compound words according to morphological structure, problems concerning the appropriateness of their control group would appear to undermine the credibility of their results.[79]

5.3 The Japanese mental lexicon: Summary

Two experiments investigated the different predictions concerning constituent-morpheme priming made by the two proposals for the Japanese mental lexicon discussed in this paper. Specifically, they sought to examine the pattern of facilitation from the constituents of two-kanji compound words once word-formation principle was controlled for as an experimental variable.

In all five word-formation conditions, the reaction times associated with both constituent conditions were significantly faster than those for the unrelated condition. However, in all but one word-formation principle, reaction time differences between the first-element and second-element conditions were not significant. That is, the results from these two experiments show that both related prime conditions facilitated responses to the target, and, in the majority of cases, at similar levels.

These results are clearly more consistent with the prediction from the Japanese lemma model than with the pattern of priming obtained by Hirose (1992), suggesting that a model of the Japanese mental lexicon must be capable of accounting for similar levels of facilitation from both constituents. According to the lemma model, the lexical retrieval for a two-kanji compound word is achieved by activation passing to the lemma unit representing the compound word from both the orthographic representation units for the constituent kanji characters. The priming effects found in the present experiments can be accounted for by assuming that lingering activation in lemma units linked to the orthographic unit of the prime give them an advantage over other units. When the prime is presented again as a constituent element of a compound word, this lingering activation leads to faster reaction times compared to the unrelated prime condition.

One aspect of the present results that requires further consideration is the significantly faster reaction times for the first constituent compared to the second constituent in the verb + complement condition indicated in the subject analyses. Although Taft, Zhu, and Peng (1999) suggest that character representations are not positionally sensitive within Chinese compound words, the possible influence of word formation on positional sensitivity was not examined. However, given the central role of frequency as a regulating mechanism within the lemma-unit model, this influence cannot be ruled out. Because the lemma units function as a means of differentiating and addressing concepts, morphological relations are expressed within the model in terms of shared semantic and syntactic properties, such as restrictions on combinations, like verb and direct object, as well as information about word class and element order. Thus, the positional frequency of a kanji is likely to be closely linked to its word class and, in turn, to word formation principles at the compound level. Although this issue needs further investigation, positional sensitivity coupled with the characteristics of verb + complement compound words, which involve only Sino-Japanese elements and, as Kageyama (1982) observes, are limited to mainly verb + direct object combinations, may have been sufficient to produce the significant difference between the constituents in this word-formation principle. This point is discussed further in Section 6.4

Overall, the Japanese lemma-unit model provides a good account of the lexical storage and retrieval of two-kanji compound words. It is also very attractive for its potential to model the complex relationships between meaning, orthography, and phonology that arise from the Japanese writing system's multi-script nature and its dual-reading system, as shown in Figure 5.11 (Note that lower-level orthographic and phonological representations are not included in the figure, which focuses on the connections to lemma units).

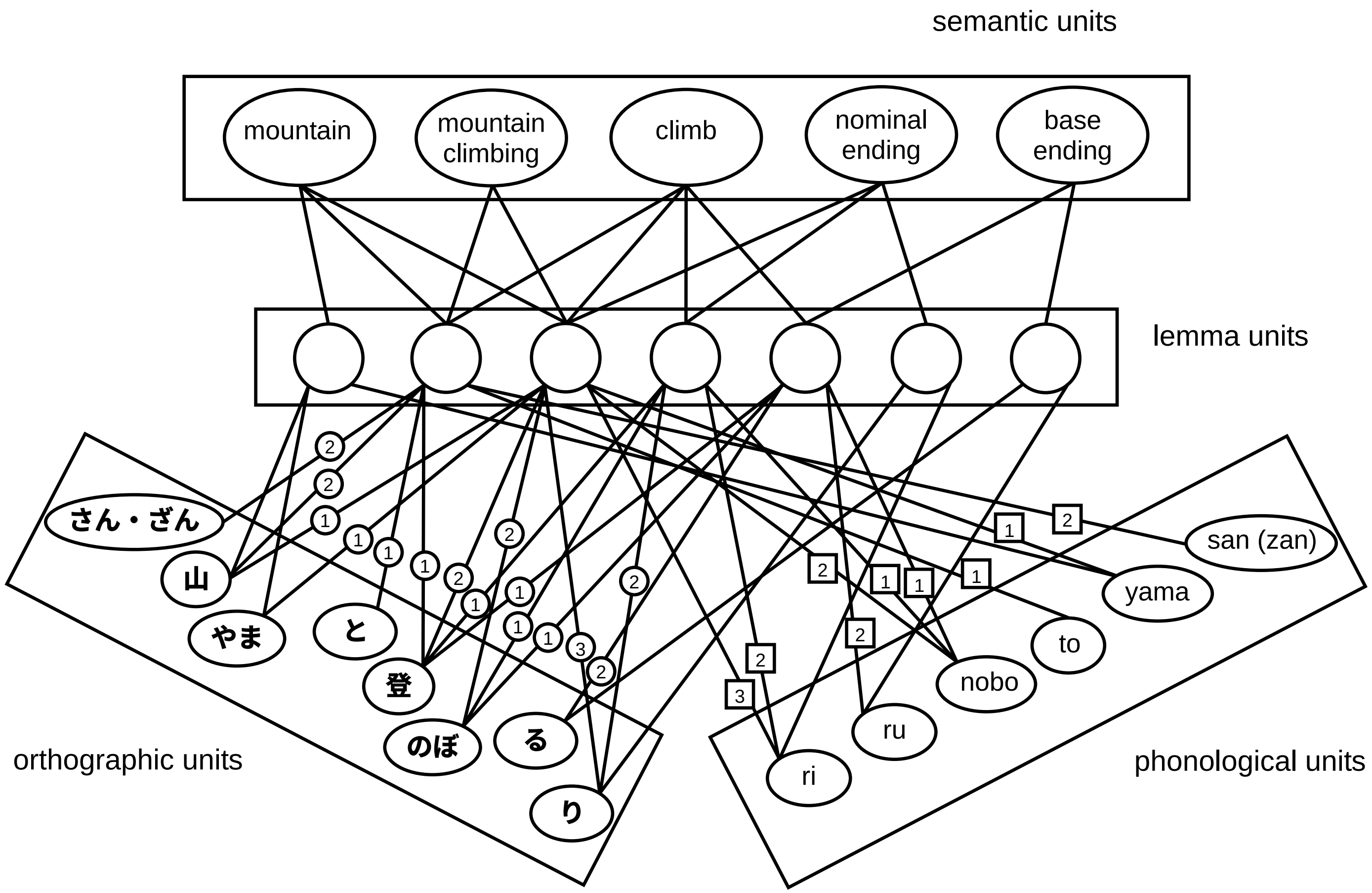


Figure 5.11. Lemma-unit-mediated connections in the Japanese lemma unit model (Joyce, 1999; in press).

Incorporating the lemma units to mediate the connections between semantic units and access representation units provides a simple way of explaining how both kanji and kana map on to meaning, as both 山 and やま *yama* can be linked via a single lemma unit to the meaning 'mountain'. The lemma units also provide an elegant method of capturing the nuances in the use of on-readings and kun-readings. As already mentioned, the lemma units are connections or way-stations that develop when semantic information regularly co-occurs with form information (Schreuder & Baayen, 1995; Taft, Liu, & Zhu, 1999). The regular co-occurrence of the meaning 'mountain' with the orthographic form 山 and the phonological form *yama* will lead to the development of a lemma unit to mediate these relationships. Similarly, the co-occurrences of the meaning 'mountain climbing,' with both the native Japanese word 山登り *yamanobori* and the Sino-Japanese word 登山 *tozan* would lead to the development of two separate lemma units linking the meaning to the appropriate orthographic and phonological forms, as depicted in Figure 5.11.

When multiple units at the same level link to a unit at another level, it is necessary to specify the order of the links (Taft, Liu, & Zhu, 1999). Although this is indicated with numbered connections in Figure 5.11, the ordering of elements is may be conceived of as a function of the lemma units rather than an artifact of the connections. Arguing that the central role of morphology is in computing meaning, Schreuder and Baayen (1995) posit lemma units (concept nodes) as part of a lexical representation that also includes mechanisms of licensing and composition, which operate on activated concept nodes. These mechanisms of licensing and composition appear to be related to the morphological awareness that Hatano (1995) refers to as compounding schemata—the implicit awareness of how kanji can be combined in compound words—acquired through learning to use the Japanese writing system. In his discussions, however, Hatano (1995) suggests that experienced readers of Japanese have two kinds of mental lexicons—the usual lexicon of words and a lexicon of kanji, or rather their corresponding morphemes as the building blocks for compound words. Although the notion of a separate lexicon for kanji is undoubtedly prompted by the bound nature of on-readings, the idea of two separate lexicons is far from appealing. It is also completely unnecessary, because the relationships between semantic representations for morphemes, orthographic representations for kanji, and phonological representations for both on-readings and kun-readings can be adequately explained through the mediation of lemma units.

Although there are some issues that need to be considered further, some of which are investigated in the research that is reported in Chapter 6, the Japanese lemma unit model undoubtedly provides a extremely appealing way of thinking about the relationships that exist between semantic representations, and access representations for orthography and phonology, as well as the morphological information that underlies two-kanji compound words in the Japanese mental lexicon.

Chapter 6

The lemma unit model: Testing the model

From our short review of visual word research in Chapter 4, we have argued that a fundamental issue that all models of the mental lexicon must address is how to represent morphological information concerning the relations that exist between polymorphemic words. Given that the central claim being made in this thesis is that kanji function as a morphographic writing system, clearly the morphology of two-kanji words is a matter of special concern for any models of the Japanese mental lexicon. Chapter 5 discussed the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon from a morphological perspective by examining the ability of two proposals to cope with the diversity that exists in the morphological structure of two-kanji compound words. As the results from two constituent-morpheme priming experiments, in which the word-formation principle of two-kanji compound word targets was controlled, are consistent with the predictions from the proposed Japanese lemma unit model version of the multilevel interactive-activation framework (Joyce, 1999, in press), they provide support for its adoption as a model for the Japanese mental lexicon.

The major feature of the lemma unit model is the incorporation of lemma units that mediate the links between access representations and semantic units. This presents an attractive approach to modeling the complex relations between meaning, orthography and phonology that exist within the Japanese mental lexicon. In this chapter, the Japanese lemma unit model is investigated further through a series of seven experiments that seek to examine various aspects of this model, while retaining a focus on two-kanji compound word morphology. Experiment 3 will investigate the extent to which the results of the constituent-morpheme priming study (Experiments 1 and 2) might be task specific. In Experiments 4 and 5, the possibility that the lemma unit model offers to integrate both kana and kanji processing within a single model is examined, while Experiments 6 and 7 consider the model's potential to account for the dual reading system of on-readings and kun-readings. In addition, by employing recently compiled constituent-morpheme frequency data (Joyce & Ohta, in press), Experiment 8 examines the issue of positional sensitivity for verbal constituent-morphemes of two-kanji compounds words. Finally, in Experiment 9, the patterns of facilitation found in the constituent-morpheme study (Exp. 1 and 2) and the positional sensitivity experiment (Exp. 8) are investigated further by reversing the prime-target relationship of the previous experiments.

6.1 Experiment 3: Primed naming task

As discussed in Chapter 4, the two most commonly used experimental tasks in visual word recognition research are the lexical decision task and the naming task. Although arguably reflecting post-access processing rather than lexical retrieval itself, the two tasks have sometimes provided different patterns of results (Balota, 1994; Balota & Chumbley, 1984; Seidenberg, 1990).

As already noted, Shen and Forster (1999) report only finding phonological priming in the naming task which explicitly emphasizes phonological processing, although they obtained orthographic priming effects in both masked-priming lexical decision and naming tasks. In an experiment that manipulated the frequencies of the constituents of two-kanji compound words, Tamaoka and Hatsuzuka (1995) observed a reversed pattern of results for these two tasks. Although these experiments differ from the primed and automatic primed lexical decision task experiments, the findings raise the question of to what extent the results from the constituent-morpheme priming studies might be task specific.

6.1.1 Purpose

The purpose of the primed naming task experiment is to investigate whether the patterns of facilitation in the primed and automatic primed lexical decision tasks (Joyce, 1999; in press) will be obtained when the experimental task is naming.

6.1.2 Method

Participants. Forty-five native Japanese students (average age 19.3, SD = 1.2) of the University of Tsukuba participated in the experiment as volunteers. None of these participants took part in Experiments 1 and 2.

Design and Materials. A 3 x 4 two-factor design was used, with both factors as within-subject variables. Although essentially the same design as for

Experiments 1 and 2, there are two differences in the design of Experiment 3. The first difference is in the prime conditions, where, instead of an unrelated kanji as a baseline condition, a blank screen was presented in an unprimed condition. This removes the possibility that the priming effects may be inflated by inhibition from an unrelated kanji (for discussion of neutral baselines, see De Groot, Thomassen, & Hudson, 1982).

The other difference was in the number of word-formation principles employed. Although five principles were used in Experiments 1 and 2, associative pairs were not included in this and all subsequent experiments. The main reason for their exclusion is that too few compound words of this kind were of sufficiently high familiarity to match with the other principles. In Experiments 1 and 2, familiarity was controlled based on the results of an familiarity evaluation survey (Joyce and Ohta, 1999). However, following the publication of the more reliable familiarity data within the NTT database (Amano & Kondo, 1999), that data was used to control for familiarity effects in this and all subsequent experiments. However, few of the associative pair compounds surveyed by Joyce and Ohta (1999) were found to be of sufficiently high familiarity compared to the other principles, and so this principle was not included in subsequent experiments.[80] For each of the four word-formation principles, 30 compound words were selected with classification scores (Joyce & Ohta, 1999) and familiarity scores (Amano & Kondo, 1999) of 5.5 or over on a 7-point scale for both criteria. The two-kanji compound word targets for Experiment 3 are shown with classification and familiarity scores in Appendix E. The mean scores for both criteria over the four word-formation principles are shown in Table 6.1.

Table 6.1

Mean Classification Scores and Familiarity Scores (NTT database) for Target Compound Words Used in Experiment 3

Word formation principle	Classification score	Familiarity score
Modifier + Modified	6.59 (0.33)	5.93 (0.32)
Verb + Complement	6.56 (0.43)	5.93 (0.24)
Complement + Verb	6.48 (0.37)	5.93 (0.27)
Synonymous Pairs	6.22 (0.34)	5.93 (0.25)

Note. Both surveys used a 7-point scale. The figures in parenthesis are the standard deviations.

Three presentation lists were prepared to counterbalance the stimulus items over the three prime conditions. Participants were assigned evenly to these lists, which were randomized for each participant.

Apparatus. Naming tasks usually measure naming latencies with a voice key that triggers a timer when the sound intensity exceeds a preset level. However, as Kello and Kawamoto (1998) point out, because voice keys are sensitive to intensity rather than frequency, they may not always be accurate for words with initial phonemes that produce low-intensity high-frequency energy (p. 371). Thus, depending on the sound-pressure level of the initial phoneme of a word, there may be a considerable delay before the amplitude exceeds the threshold level. Sakuma, Fushimi, and Tatsumi (1997) have examined this issue for Japanese, and they found, for example, a mean error of 113ms for the voiceless fricative /s/ as the initial phoneme of a word, when they compared recorded speech wave data to voice key triggers. To avoid, therefore, the problem with using a voice key to measure naming latencies, naming latencies in this experiment were measured from waveform data recorded using a PowerLab data-recording unit (4SP: ADInstruments).[81] Participant naming responses into a microphone were recorded on one channel. A second channel recorded a signal from a luminance sensor monitoring a corner of the computer screen. When a ※ mask appeared at the center of the computer screen within a trial of the experiment, a black patch would also appear in the corner of the screen (hidden from the participants by a plastic cover). The luminance-sensor signal, which was interrupted by the black patch and was restored when the target stimulus is presented, provided an indicator of target stimulus onset. By comparing these two signals using the Chart software application (version 4.0.1: ADInstruments), it was possible to accurately measure naming latency, as shown in Figure 6.1 overleaf which is a screen print of Chart showing a waveform for a participant response. In this experiment, the presentation of stimulus items was controlled by SuperLab Pro (version 2.0, Cedrus Corporation) running on a personal computer (Performance, Gateway). The microphone input and the signal from a luminance sensor were collected via a PowerLab data-recording unit (4SP, ADInstruments), and recorded by the Chart software application (version 4.0.1, ADInstruments) running on a laptop personal computer (DynaBook DB65P/4RA, Toshiba Corp.), which was connected to the PowerLab data-recording unit. A photograph showing the apparatus for this experiment is included in Appendix F. Stimuli at a font size of 36 points were displayed on the computer screen at a viewing distance of approximately 50 cm.

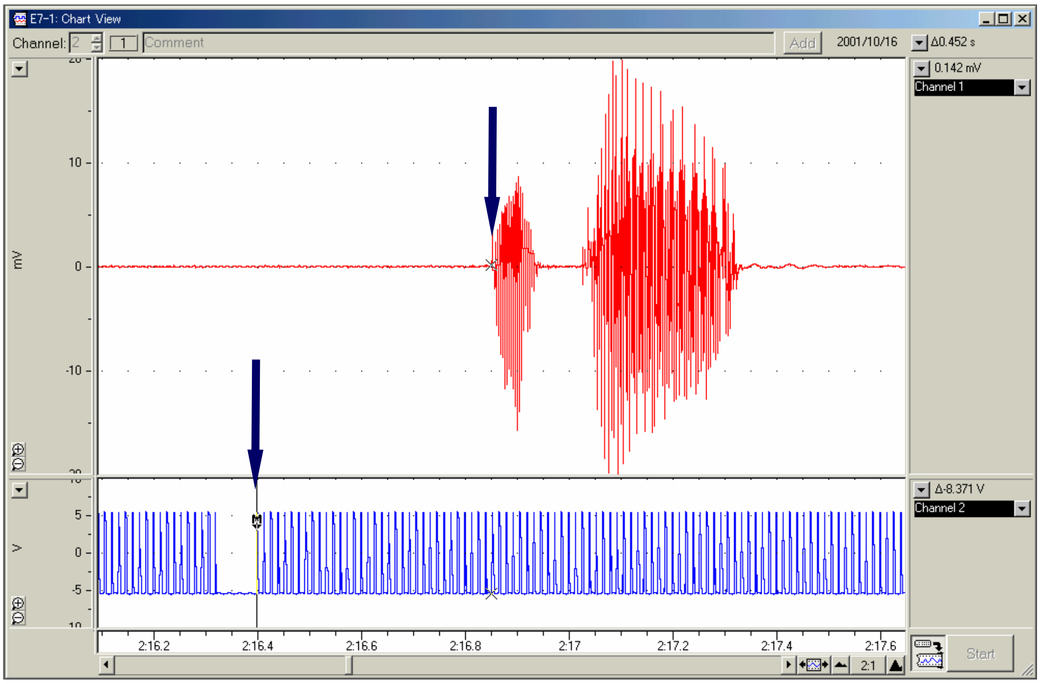


Figure 6.1. Screen print of Chart (version 4.0.1, ADInstruments) showing waveform of participant response in Channel 1 pane (upper section) and luminance-sensor signal in Channel 2 pane (lower section). The naming latency for this response (/ryokō/ 旅行) of 452 ms is measured from the marker (M) at target stimulus onset in the Channel 2 pane to the cursor (X) set at the onset of voicing in the Channel 1 pane.

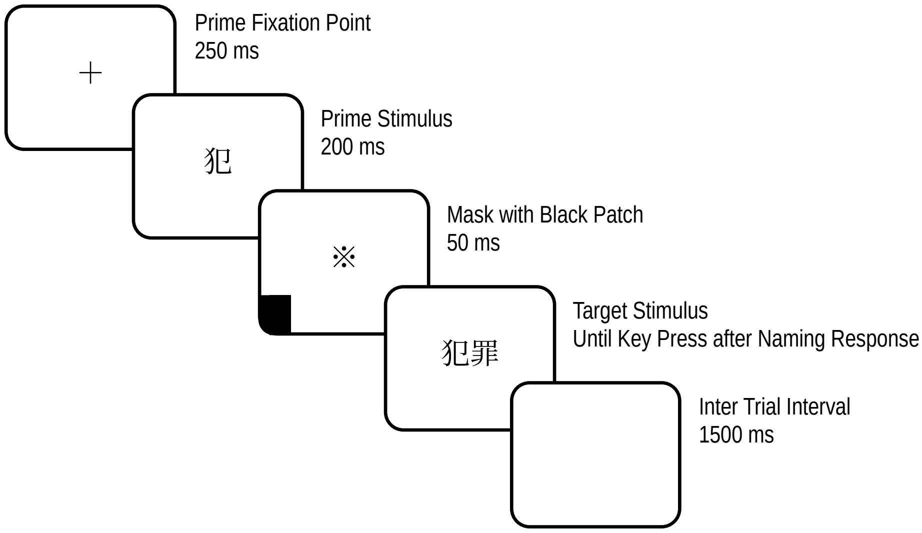


Figure 6.2. Presentation sequence during a trial in Experiment 3.

Procedure. The procedure in this experiment was similar to that used in Experiment 2, and was as shown in Figure 6.2. At the start of a trial, a plus symbol (+) appeared in the center of the screen as a fixation point for 250 ms. This was followed by the single kanji prime, displayed for 200 ms, and then by an asterisk-like symbol (※) as a mask for 50 ms. When the mask appeared at the center of the computer screen, a black patch would also appear in the corner of the screen (hidden from the participants by a plastic cover), which was monitored by the luminance sensor. The target stimulus item was then displayed at the center of the computer screen. After naming the target, the participant pressed a button on a response box to start the next trial. There was a 1,500 ms inter-trial interval. Participants were instructed to read the target word aloud as quickly and as accurately as possible into a microphone, positioned in front of them and adjusted to their height. The instructions to the participants were printed on a card, which they were asked to read, as shown in Appendix G. After briefly repeating the instructions verbally to confirm that they had been understood, there was practice session of 12 trials, and the whole experiment took between 10 to 15 minutes to complete.

6.1.3 Results

ANOVAs were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). For two items, the number of error responses was very high (over 15% for both), and so those items were removed from all analyses. Error responses were excluded from the analysis of reaction times. If the standard score for a reaction time was outside the range of +/- 2.5 for a given participant, the reaction time was adjusted to that equivalent to a standard score of +/- 2.5. Table 6.2 presents the mean reaction times and error rates for Experiment 3.

Error analysis. The overall error rate was low at 0.92%. No significant differences were found in the analyses of the error data; for prime in the both subject and item analyses, p = .572 and p = .611 respectively, and for word-formation principle in the both subject and item analyses, p = .182 and p = .334 respectively.

Table 6.2

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 3

Word Formation Principle		Error rate	Reaction times		RT diff
Modifier + modified					
	First constituent	1.1	543	(94)	+21
	Second constituent	0.7	564	(82)	0
	Unprimed	1.1	564	(100)	
Verb + complement					
	First constituent	1.6	520	(90)	+36
	Second constituent	1.1	549	(89)	+7
	Unprimed	1.3	556	(98)	
Complement + verb					
	First constituent	1.1	538	(87)	+25
	Second constituent	0	553	(86)	+10
	Unprimed	0.9	563	(100)	
Synonymous pairs					
	First constituent	0.2	522	(99)	+32
	Second constituent	1.1	543	(92)	+11
	Unprimed	0.7	554	(106)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unprimed condition reaction time in each word formation principle condition.

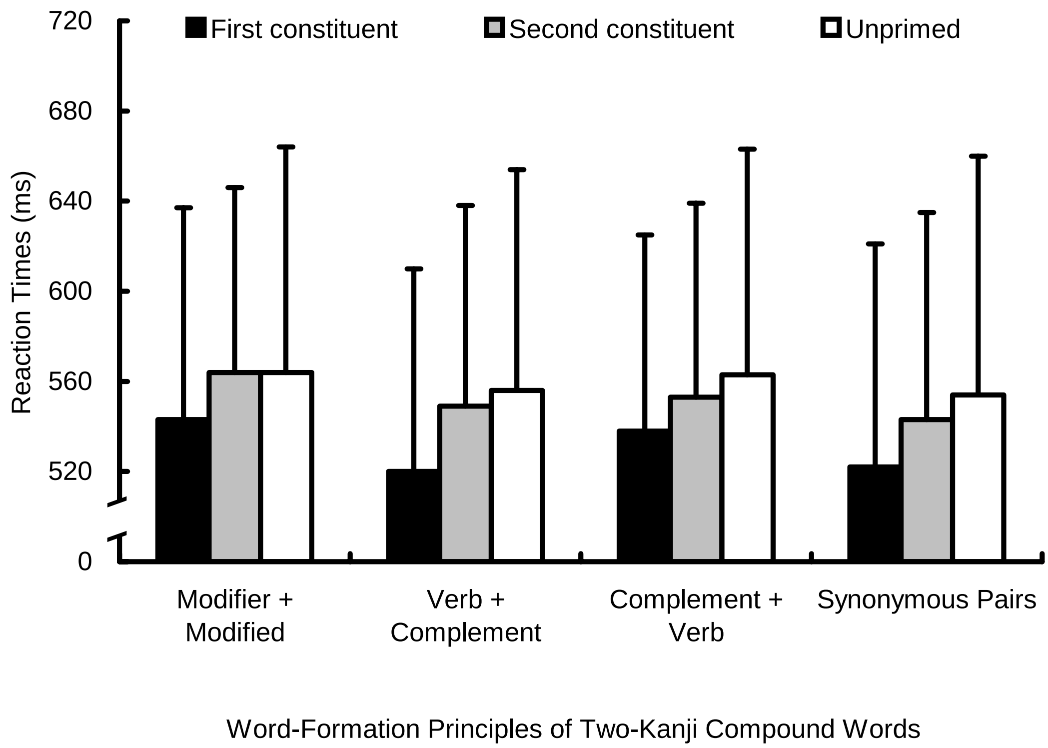


Figure 6.3. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 3.

Reaction times analysis. In the subject analysis, there was a significant main effect of prime, F_1 (2, 88) = 28.29, p < .0001, and of word-formation principle, F_1 (3, 132) = 15.97, p < .0001, although there was no significant interaction, p = .238. In the item analysis, there was only a significant main effect of prime, F_2 (2, 228) = 40.13, p < .0001. Reaction times for Experiment 3 are shown in Figure 6.3.

Planned comparisons for the main effect of prime in the subject analysis revealed that reaction times in the first constituent condition were significantly faster than both the second constituent condition and the unprimed condition for all word-formation principle conditions, with HSD = 9.60 in all cases. Planned comparisons for the main effect of prime in the item analysis revealed that reaction times in the first constituent condition were significantly faster than both the second constituent condition and the unprimed condition for all word-formation principle conditions, with HSD = 7.75 in all cases.

Planned comparisons for the main effect of word-formation revealed that average reaction times were significantly slower in the modifier + modified condition compared to the verb + complement condition and the synonymous pairs; that reaction times were significantly slower in the verb + complement condition than in the complement + verb condition; and that reaction times were slower in the verb + complement condition than in the synonymous pairs, with HSD = 7.53 in all cases.

6.1.4 Discussion

The results from Experiment 3 differ from the results of the primed lexical decision experiments (Exp. 1 and 2). However, arguably, the difference may not be a reflection of differences in lexical retrieval in the naming and the lexical decision tasks, and could stem from differences in post-access processing.

Significant priming was found in the first constituent condition compared to both the second constituent and unprimed conditions. As the SOA of 250 ms in this experiment is the same as that used in Experiment 2, it is reasonable to assume that the lexical retrieval processing of the prime was the same in both experiments. The priming in the first constituent condition suggests that in the processing of the prime, at least, the on-reading of the kanji had become available, and being consistent with the initial part of the pronunciation for the two-kanji compound word target facilitated its naming. In the second constituent condition, however, although in most cases reaction times were slightly faster compared to the unprimed conditions, the differences failed to reach significant levels for all word-formation conditions. Again, it is reasonable to assume that the lexical retrieval processing of the second constituent primes was the same as for the first constituent primes, and that in this processing of the prime, its on-reading had become available, but because this was not consistent with the initial part of the pronunciation for the two-kanji compound word which was required in the naming task, the phonological activation did not facilitate the naming of the two-kanji compound word target. These results suggest that in the naming task, where phonological processing is emphasized, when naming the two-kanji compound word targets, it is much easier to extend the phonological information from the first constituent prime than it is to reverse the phonological information from the second constituent prime to a word-final position. Because of this, the second constituent prime condition did not significantly facilitate the naming of the two-kanji compound word targets.[82]

6.2 Cross-script priming (Kana orthography version) Joyce (2001)

One important aspect of the lemma unit model is that the inclusion of lemma units provides an extremely promising approach to modeling the complex relations that exist between meaning, orthography and phonology that arise from the Japanese writing system's multi-script nature and its dual-reading system (Joyce, 1999, in press). In proposing the lemma unit model for the Japanese mental lexicon, Joyce suggests that an important feature of the model is its potential to unify both kanji and kana processing within a single integrated model.

As mentioned in Chapter 5, there have been many comparative studies of kanji and kana processing, which stressing the functional differences between the two scripts have interpreted findings such as those of Shimamura (1987), which we discussed earlier, of faster word comprehension with kanji words and faster word naming with kana as evidence of differences in processing. We noted also that there is evidence of similarities between the two scripts, for just as Besner and Hildebrandt (1987) have reported evidence of direct visual processing of kana, other studies, such as Wydell, Patterson, and Humphreys (1993) have provided evidence that phonological information is involved in kanji processing.

Shafiullah and Monsell (1999) suggest that decoding kana and kanji involves different resources based their finding of switching costs in the results of a series of naming and semantic categorization task experiments where participants had to switch between reading words in kanji and kana at predictable intervals.[83] While the implications of such switching costs for the normal reading of mixed kanji and kana writing are not clear, still, the fact that words that are normally in kanji can still be retrieved when written in kana suggests that the two scripts function together as elements of an integrated system rather than as totally separate systems.

Joyce (1999, in press) suggests that theoretically the incorporation of lemma units within the multilevel interactive-activation structure can provide a means of accounting for both kanji and kana processing within a single model. The purpose of the cross-script priming studies was to provide evidence for the existence of connections from the kana representations of morphemes to the lemma units representing compound words normally represented orthographically as two-kanji compound words.

A second purpose of these studies was to further explore the implications of the Japanese writing system's dual-reading system for the lemma unit model by using as primes both on-readings, in Experiment 4, and kun-readings, in Experiment 5, for the constituents of compounds that are normally represented orthographically as two-kanji compound words. Within the Japanese lemma unit model, the different characteristics of on-readings and kun-readings are explained in terms of the numbers of lemma units the readings are connected to and by the strengths of these connections, where the weightings will be a product of frequency. To illustrate the possible differences, we shall consider the numbers of connections that may exist with respect to the kanji 犯 'commit (a crime),' which has an on-reading of *han* and a kun-reading of *okasu*. This kanji has a familiarity rating of 6.0 in the NTT database (Amano & Kondo, 1999) and is listed in order of frequency at 849 with 3,535 tokens in a recent survey of newspaper kanji frequencies (Yokoyama, Sasahara, Nozaki, & Long. 1998). According to recently compiled constituent morpheme frequency data for two-kanji compound words (Joyce & Ohta, in press), which is introduced in Section 6.4.1, this kanji appeared in newspapers has the first constituent of six two-kanji compounds words and as the second constituent of 16 compound words. Although we cannot assume that all these compound words will exist in the average mental lexicon, in addition to connections for its on-reading and its kun-reading, this kanji may have connections to at least 22 lemma units. Looking at the on-reading *han*, this is shared by at least 33 kanji,[84] consisting of 21 Jōyō kanji (including 9 kyōiku kanji), 9 JIS level 1 kanji, and 3 JIS level 2 kanji. Among these 33 kanji, the kanji 犯 has the sixth highest familiarity rating. However, counting all the compound words that have the on-reading *han* as part of their pronunciation in the constituent-morpheme frequency data (Joyce & Ohta, in press), there is a total of 564, where kanji with this reading are the first constituents (308 compound words) or the second constituents (256 compound words). Again, while not all these will exist in the average mental lexicon, still, the number of connections to lemma units for compound words contrasts sharply with the possible 22 connections from the kanji 犯. Looking at the kun-reading *okasu*, there are seven words with this reading (four of these are inflected forms of a base verb), of which 犯す has the highest familiarity score of 5.94. These facts suggest that the levels and patterns of facilitation in the two cross-script priming experiments (Experiments 4 and 5) presenting kana orthography primes of the on-readings and kun-readings associated with the constituent-morphemes of two-kanji compound words might be different from those found in Experiments 1 and 2.

6.2.1 Experiment 4: On-reading primes

6.2.1.1 Purpose

The purpose of this experiment is to investigate whether the on-readings for the constituent-morpheme kanji of two-kanji compound word presented in kana orthography will facilitate recognition of the two-kanji compound word. Such a finding would support the existence of connections between kana representations and lemma units within the lemma unit model.

6.2.1.2 Method

Participants. Forty-five native Japanese students (average age 20.7, SD = 2.71) of the University of Tsukuba participated in the experiment as volunteers. None of the participants took part in the previous experiments.

Design and Materials. The 3 x 4 two-factor design with both factors as within-subject variables in Experiment 4 is the same as that for Experiment 3.

Fujii (1996) observes that there are 296 syllable types as on-readings for the 1945 Jōyō kanji. Moreover, 229 of these have a two mora structure, of which the second syllable is always one of just seven syllables, イ, ウ, キ, ク, チ, ツ, ン (p. 100). Because of this rather limited variety in on-readings, and because many kanji share the same on-readings, it was not possible to make this experiment so that an on-reading represented in kana was only used once for one particular kanji. To control for the effects of repeated presentations of the same on-reading primes to different kanji as constituents of two-kanji compound word targets as far as possible two measures were taken.

The first measure was to ensure that the average number of kanji associated with a given on-reading was matched closely over the four word-formation principles. Thus, for each of the four word-formation principles, 30 compound words were selected with classification scores (Joyce & Ohta, 1999) and familiarity scores (Amano & Kondo, 1999) of 5.5 or over on a 7-point scale for both criteria, which were also closely matched over the conditions in terms of the average number of kanji per on-reading. The two-kanji compound word targets for Experiment 4 are shown with classification and familiarity scores in Appendix H. The mean scores for both criteria over the four word-formation principles, together with the average number of kanji associated with an on-reading, are shown in Table 6.3 overleaf.

The second measure to limit the effects of repeated presentations of the same on-reading primes was to restrict the maximum number of kanji sharing an on-reading to two within a particular word-formation principle and to six over all four principles. The stimuli were divided into two sets, with a maximum of three kanji shared an on-reading in each. For both sets, three presentation lists were prepared to counterbalance the stimulus items over the three prime conditions. Each participant was run on two presentation lists with a break between blocks, and by counterbalancing the combinations of all six presentation lists, which were randomized for each participant, each participant would, at most, only see a prime repeated twice, once in each block.

Table 6.3

Mean Classification Scores and Familiarity Scores (NTT database) for Target Compound Words and Mean Number of Kanji per On-reading Primes Used in Experiment 4

Word formation principle	Classification score	Familiarity score	Number of kanji per on-reading prime
Modifier + Modified	6.62 (0.23)	5.93 (0.31)	18.23
Verb + Complement	6.72 (0.22)	5.93 (0.20)	18.33
Complement + Verb	6.57 (0.32)	5.93 (0.25)	18.35
Synonymous Pairs	6.38 (0.25)	5.93 (0.24)	18.25

Note. Both surveys used a 7-point scale. The figures in parenthesis are the standard deviations.

For the lexical decision task, 72 non-word combinations of two kanji were selected from the set of items created for Experiment 1.

Apparatus. Super Lab Pro (version 2.0, Cedrus Corporation), running on a personal computer (Performance, Gateway), controlled the presentation of stimulus items and recorded lexical decisions collected via a response box (Cedrus Corporation, RB-600). Stimuli at a font size of 36 points were displayed on the computer screen at a viewing distance of approximately 50 cm.

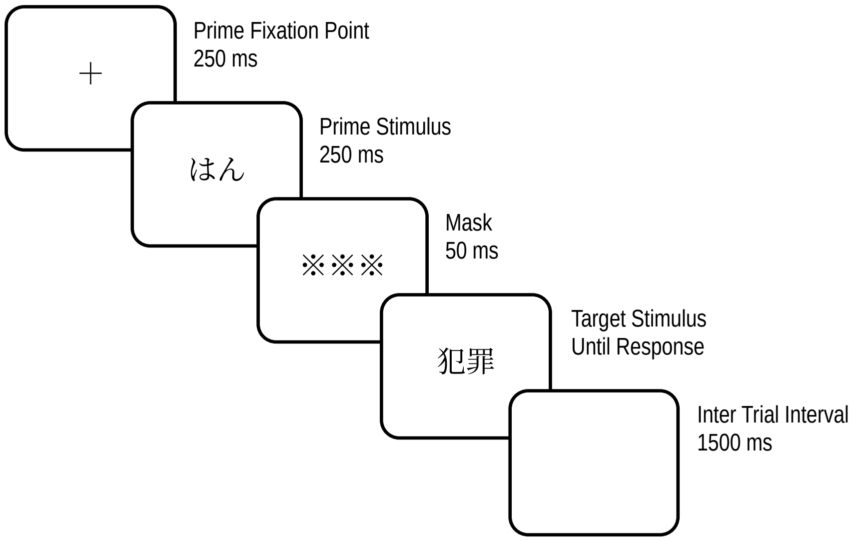


Figure 6.4. Presentation sequence during a trial in Experiment 4.

Procedure. The presentation procedure for Experiment 4 was similar to that in Experiment 3, with a kanji prime replaced by the kana orthography on-reading prime, as shown in Figure 6.4.

At the start of a trial, a plus symbol (+) appeared in the center of the screen as a fixation point for 250 ms. This was followed by the on-reading prime in kana displayed for 250 ms, and then by a string of three asterisk-like symbols (※※※) as a mask for 50 ms. The target stimulus item was then displayed and remained on the screen until the participant pressed a button on the response box for the lexical decision. There was a 1,500 ms inter-trial interval. Participants were instructed to press a green button for a compound word and a red button for a non-word as quickly and as accurately as possible. The instructions to the participants were printed on a card, which they were asked to read, as shown in Appendix I. After briefly repeating the instructions verbally to confirm that they had been understood, there was practice session of ten trials, and the whole experiment took between 15 to 20 minutes to complete.

6.2.1.3 Results

ANOVAs were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). Error responses were excluded from the analysis of reaction times. If the standard score for a reaction time was outside the range of +/- 2.5 for a given participant, the reaction time was adjusted to that equivalent to a standard score of +/- 2.5. Table 6.4 presents the mean reaction times and error rates for Experiment 4.

Error analysis. The overall error rate was very low at 2.78%. Analyses only revealed significant main effects of word-formation principle in both the subject and item analyses, F_1 (3, 132) = 8.45, $p < .0001$; F_2 (3, 116) = 3.04, $p < .05$.

Planned comparisons for the main effect of word-formation principle in the subject analysis revealed that errors in the modifier + modified condition were higher than all the other word-formation principle conditions, HSD = 0.157 in all cases. Planned comparisons for the main effect of word-formation principle in the item analysis revealed that errors in the modifier + modified condition were higher than in the verb + complement condition, HSD = 0.39.

Table 6.4

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 4

Word Formation Principle		Error rate	Reaction times		RT diff
Modifier + modified					
	First constituent	3.8	588	(136)	+34
	Second constituent	5.6	609	(148)	+11
	Unprimed	4.2	622	(148)	
Verb + complement					
	First constituent	1.1	570	(121)	+25
	Second constituent	2.0	591	(131)	+4
	Unprimed	2.4	595	(124)	
Complement + verb					
	First constituent	2.0	583	(146)	+24
	Second constituent	4.0	598	(133)	+9
	Unprimed	2.4	607	(133)	
Synonymous pairs					
	First constituent	1.6	576	(144)	+21
	Second constituent	2.4	584	(129)	+13
	Unprimed	1.8	597	(118)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unprimed condition reaction time in each word formation principle condition.

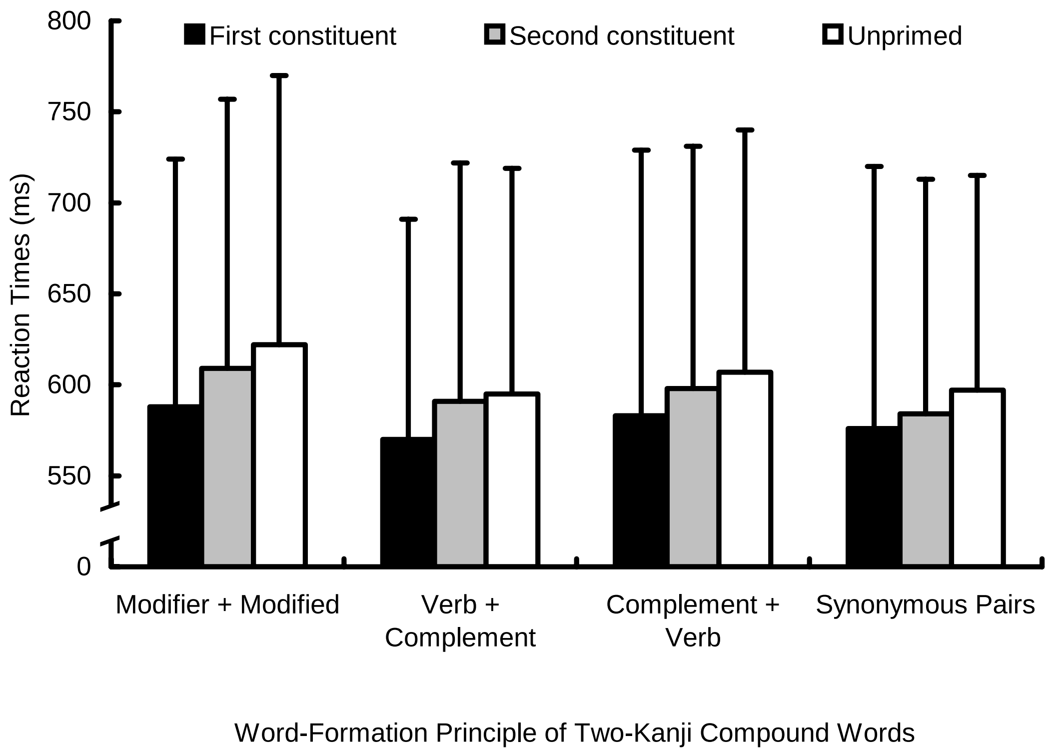


Figure 6.5. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 4.

Reaction times analysis. In the subject analysis, there were significant main effects of prime, F_1 (2, 88) = 9.13, $p < .0001$, and of word-formation principle, F_1 (3, 132) = 7.38, $p < .0001$, although there was no significant interaction, $p = .822$. In the item analysis, there was only a significant main effect of prime, F_2 (2, 232) = 8.51, $p < .0001$. Reaction times for Experiment 4 are shown in Figure 6.5.

Planned comparisons for the main effect of prime in the subject analysis revealed that reaction times in the first constituent condition were significantly faster than both the second constituent condition and the unprimed condition for all word-formation principle conditions, with HSD = 14.79 in all cases. Planned comparisons for the main effect of prime in the item analysis revealed that reaction times in the first constituent condition were significantly faster than both the second constituent condition and the unprimed condition for all word-formation principle conditions, with HSD = 15.25 in all cases.

Planned comparisons for the main effect of word-formation principle in the subject analysis revealed that reaction times in the modifier + modified condition were significantly slower than in the verb + complement and the synonymous pairs, with HSD = 13.56 in all cases.

6.2.1.4 Discussion

Although there were slightly faster reaction times in the second constituent condition, significant differences compared to the unprimed condition were only found in the first constituent condition. The results of this experiment are similar to those in the primed naming experiment (Exp. 3). This suggests that when phonological information is emphasized by presenting only the on-reading for the constituent morpheme, it would seem to be easier to extend this for first constituent prime than it is to reverse the phonological information from the second constituent prime to a word-final position.

However, the presence of priming in the first constituent condition supports the suggestion for the lemma unit model that kana representations are linked to the lemma units, and that activation from kana representations can prime lemma units for compound words usually represented orthographically with kanji.

6.2.2 Experiment 5: Kun-reading primes

6.2.2.1 Purpose

The presence of priming in the first constituent condition in Experiment 4 from kana orthography on-reading primes, leads us to expect facilitation from kana orthography kun-reading primes as well. However, with the greater semantic information that we may assume to be available with kun-readings compared to the purely phonological information of the on-readings, the pattern of facilitation in this experiment is likely to be different in this experiment.

The purpose of this experiment is to investigate whether the kun-readings for the constituent-morpheme kanji of two-kanji compound word presented in kana orthography will facilitate recognition of the two-kanji compound word, and if so, how the pattern of facilitation would differ from on-reading primes. Again, such finding would support the existence of connections between kana representations and lemma units within the lemma unit model, and would also suggest that these connections link the allomorphs of the constituent-morphemes of two-kanji compound words to lemma units for the two-kanji compound words.

6.2.2.2 Method

Participants. Forty-five native Japanese students (average age 22, SD = 4.8) of the University of Tsukuba participated in the experiment as volunteers. None of the participants took part in the previous experiments.

Design and Materials. The 3 x 4 two-factor design with both factors as within-subject variables in Experiment 5 is the same as that for Experiment 4.

The problem of repeated presentations of a prime was not an issue in this experiment, because it was possible to counterbalance such primes over the different prime conditions across the three presentation lists. However, in contrast to the one or two syllable on-readings, there was greater variation in the lengths of the kun-readings. Because this is related to word class, with kun-readings tending to be longer for adjectives and verbs than for nouns, it was not possible to perfectly match for this, but the kun-readings for first and second constituents were match as closely as possible, while maintaining controls for average familiarity and classification scores. For each of the four word-formation principles, 30 compound words were selected with classification scores (Joyce & Ohta, 1999) and familiarity scores (Amano & Kondo, 1999) of 5.5 or over on a 7-point scale for both criteria, which were also closely matched over the conditions in terms of the average length of the kun-readings. The two-kanji compound word targets for Experiment 5 are shown with classification and familiarity scores in Appendix J. The mean scores for both criteria over the four word-formation principles, together with the average lengths of the kun-readings for the constituents, are shown in Table 6.5.

For the lexical decision task, 72 non-word combinations of two kanji were selected from the set of items created for Experiment 1.

Table 6.5

Mean Classification Scores and Familiarity Scores (NTT database) for Target Compound Words and Mean Length of Kun-reading Primes as Function of Position Used in Experiment 5

Word formation principle	Classification score	Familiarity score	Mean length of kun-reading prime	
			First	Second
Modifier + Modified	6.62 (0.23)	5.93 (0.28)	2.8	2.3
Verb + Complement	6.64 (0.37)	5.93 (0.24)	3.0	2.4
Complement + Verb	6.39 (0.44)	5.93 (0.26)	2.8	3.0
Synonymous Pairs	6.18 (0.31)	5.93 (0.31)	3.0	3.3

Note. Both surveys used a 7-point scale. The figures in parenthesis are the standard deviations. First = first constituent and second = second constituent.

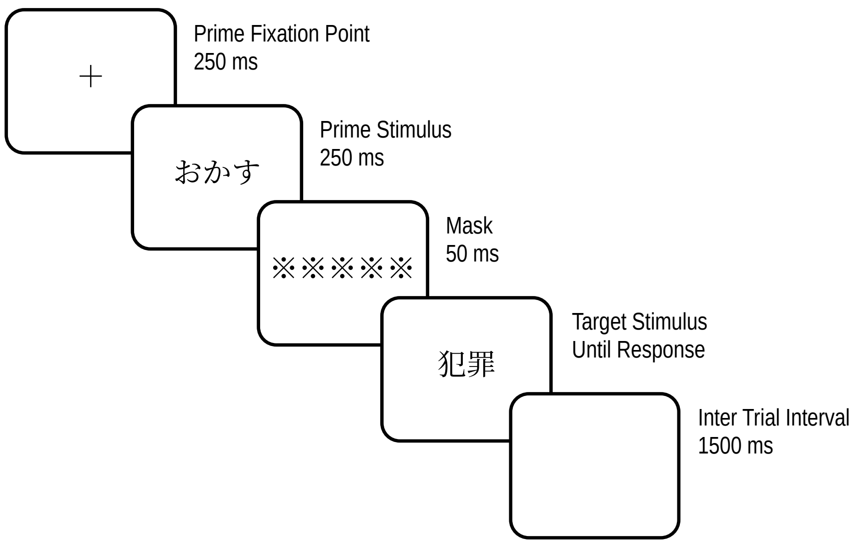


Figure 6.6. Presentation sequence during a trial in Experiment 5.

Apparatus. The apparatus used in Experiment 5 is the same as that used in Experiment 4.

Procedure. The presentation procedure for Experiment 5 is similar to Experiment 4, with a kana orthography on-reading prime being replaced by a kana orthography kun-reading prime, as shown in Figure 6.6.

At the start of a trial, a plus symbol (+) appeared in the center of the screen as a fixation point for 250 ms. This was followed by the on-reading prime in kana displayed for 250 ms, and then by a mask for 50 ms consisting of a string of five asterisk-like symbols (※※※※※). The target stimulus item was then displayed and remained on the screen until the participant pressed a button on the response box for the lexical decision. There was a 1,500 ms inter-trial interval. Participants were instructed to press a green button for a compound word and a red button for a non-word as quickly and as accurately as possible. The instructions to the participants were printed on a card, which they were asked to read, as shown in Appendix K. After briefly repeating the instructions verbally to confirm that they had been understood, there was practice session of 10 trials, and the whole experiment took between 10 to 15 minutes to complete.

6.2.2.3 Results

ANOVAs were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). Error responses were excluded from the analysis of reaction times. If the standard score for a reaction time was outside the range of +/- 2.5 for a given participant, the reaction time was adjusted to that equivalent to a standard score of +/- 2.5. Table 6.6 presents the mean reaction times and error rates for Experiment 5.

Error analysis. The overall error rate was very low at 2.70%. Analyses only revealed significant main effects of word-formation principle, F_1 (3, 132) = 5.74, $p < .0001$; F_2 (3, 116) = 3.20, $p < .05$.

Planned comparisons for the main effect of word-formation principle in the subject analyses revealed that errors in the synonymous pairs condition were lower in both the modifier + modified and the verb + complement conditions, HSD = 0.176. Planned comparisons for the main effect of word-formation principle in the item analysis revealed that errors in the synonymous pairs were lower than in the modifier + modified condition, with HSD = 0.35.

Reaction times analysis. In the subject analysis, there were significant main effects of prime, F_1 (2, 88) = 31.80, $p < .0001$, and of word-formation principle, F_1 (3, 132) = 5.91, $p < .005$, although there was no significant interaction, $p = .189$. In the item analysis, there was only a significant main effect of prime, F_2 (2, 232) = 21.74, $p < .0001$. Reaction times for Experiment 5 are shown in Figure 6.7.

Planned comparisons for the main effect of prime in the subject analysis revealed that reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unprimed condition for all conditions. However, reaction times in the first constituent were also significantly faster than in the second constituent condition for all word-formation conditions, with HSD = 12.64 in all cases.

Table 6.6

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 5

Word Formation Principle		Error rate	Reaction times		RT diff
Modifier + modified					
	First constituent	4.7	583	(89)	+36
	Second constituent	4.0	600	(86)	+19
	Unprimed	2.4	619	(78)	
Verb + complement					
	First constituent	2.2	563	(78)	+53
	Second constituent	3.1	594	(89)	+22
	Unprimed	2.7	616	(81)	
Complement + verb					
	First constituent	4.4	589	(79)	+29
	Second constituent	3.1	596	(88)	+22
	Unprimed	2.4	618	(87)	
Synonymous pairs					
	First constituent	0.9	559	(88)	+51
	Second constituent	1.1	585	(100)	+25
	Unprimed	1.3	610	(81)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unprimed condition reaction time in each word formation principle condition.

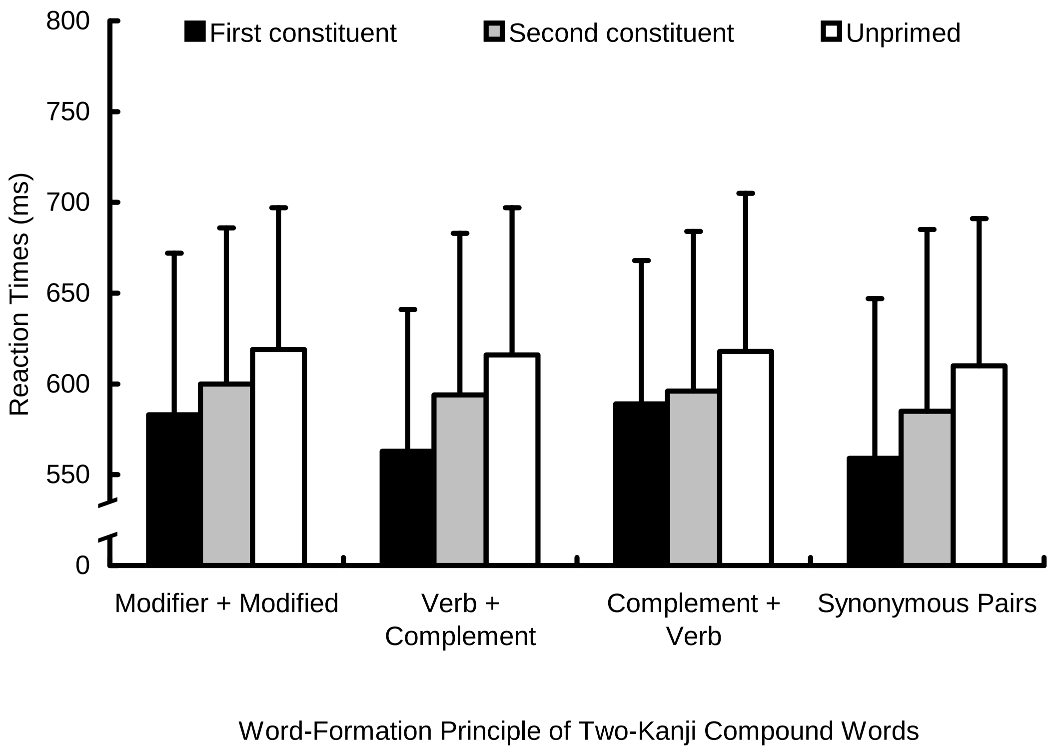


Figure 6.7. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 5.

Similarly, planned comparisons for the main effect of prime in the item analysis revealed that reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unprimed condition for all conditions. However, reaction times in the first constituent were also significantly faster than in the second constituent condition for all word-formation conditions, with HSD = 14.87 in all cases.

Planned comparisons for the main effect of word-formation principle in the subject analysis revealed that reaction times in the synonymous pairs were faster than in both the modifier + modified and the complement + verb conditions, with HSD = 11.79 in all cases.

6.2.2.4 Discussion

The results of this experiment show significant priming in both the first constituent and second constituent conditions compared to the unprimed condition, as well as significant differences between the first constituent and the second constituent condition.

In contrast to the bias towards phonological information in the on-reading cross-script priming experiment, with the greater semantic information that can be assumed to be available with kun-readings, both constituent conditions primed lexical decisions for two-kanji compound words. It is not, however, clear why there was greater facilitation in the first constituent conditions compared to the second constituent condition, for there is no orthographic or phonological overlap between the primes and the constituent morphemes in the two-kanji compound words.[85]

The presence of cross-script priming in both Experiment 4 with on-reading primes and in Experiment 5 with kun-reading primes clearly supports the lemma unit model, where both kanji and kana processing are seen as being integrated within a single system. The differences between the two experiments may be taken as a reflection of the differences in the characteristics of on-readings and kun-readings, which we discussed in the introduction to the cross-script studies. These differences are explained by the lemma unit model in terms of differences in the number of connections to lemma units that exist for the two kinds of reading, where on-readings are shared by more kanji that for kun-readings, and by the strengths of these connections, which would be the product of frequency.

6.3 Cross-modality priming (Auditory priming) Joyce (2001)

In addition to providing an interesting approach to unifying both kanji and kana processing within a single integrated model, as suggested in the cross-script priming study (Experiments 4 and 5), the lemma unit model is capable of accounting for the complexity in terms of phonological representations due to the Japanese writing system's dual-reading system of on-readings and kun-readings. Here, we continue to investigate this aspect of the Japanese mental lexicon, by conducting cross-modality versions of the cross-scripts experiments.[86] In Experiment 6, on-readings for the constituent morphemes of two-kanji compound words are presented auditorily as primes to visually presented two-kanji compound words. Similarly, in Experiment 7, kun-readings for the constituent morphemes of two-kanji compound words are presented auditory as primes to visually presented two-kanji compound words.

The purposes of these cross-modality priming experiments is to provide evidence for the existence of connections from phonological representations for both on-readings and kun-readings for the constituent morphemes to lemma units representing compound words, and to further examine the levels and patterns of facilitation from auditorily presented on-readings and kun-readings that are associated with the constituent-morphemes of two-kanji compound words

6.3.1 Experiment 6: On-reading primes

6.3.1.1 Purpose

The purpose of this experiment is to investigate whether the auditorily presented on-readings for the constituent-morpheme kanji of two-kanji compound word will facilitate recognition of the two-kanji compound word when presented visually. Such a finding would support the existence of connections between phonological access representations and lemma units within the lemma unit model.

6.3.1.2 Method

Participants. Forty-five native Japanese students (average age 21, SD = 3.9) of the University of Tsukuba participated in the experiment as volunteers. None of the participants took part in the previous experiments.

Design and Materials. The 3 x 4 two-factor design with both factors as within-subject variables in Experiment 6 is the same as that for Experiment 4.

The two-kanji compound word targets used in Experiment 6 are the same as those used in Experiment 4. The counterbalancing of the combination of the six sets of presentation lists was also the same, to ensure that each participant would, at most, only hear a prime repeated twice, once in each block of trials.

Wave files for the auditory primes were created from the sound files in the

NTT database (Amano & Kondō, 1999), which were edited for included silence using sound editing software (Sound Forge XP, version 4.5c).[87] As the average duration of these sound files 342.9 ms (SD = 70.3, range 158 ms to 519 ms), a sound file with 340 ms of silence was prepared for the unprimed condition.

Apparatus. The apparatus used in Experiment 6 is the same as that used in Experiment 4.

Procedure. The presentation procedure for Experiment 6 is similar to Experiment 4, but with the on-reading prime being auditorily presented through headphones, as shown in Figure 6.8.

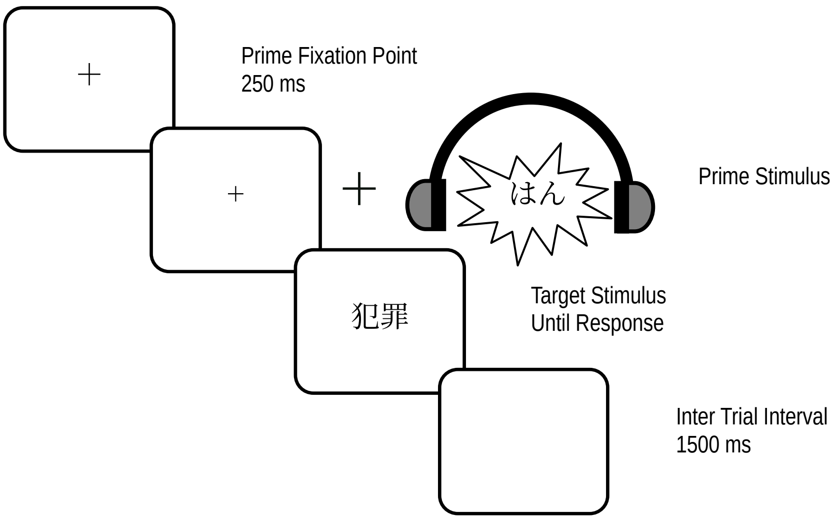


Figure 6.8. Presentation sequence during a trial in Experiment 6.

At the start of a trial, a plus symbol (+) appeared in the center of the screen as a fixation point for 250 ms. This remained on the screen as the on-reading prime was presented auditorily through headphones. Presentation of the target stimulus item at the center of the computer screen immediately followed offset of the auditory prime. This remained on the screen until the participant pressed a button on the response box for the lexical decision. There was a 1,500 ms inter-trial interval. Participants were instructed to press a green button for a compound word and a red button for a non-word as quickly and as accurately as possible. The instructions to the participants were printed on a card, which they were asked to read, as shown in Appendix L. After briefly repeating the instructions verbally to confirm that they had been understood, there was practice session of ten trials, and the whole experiment took between 15 to 20 minutes to complete.

6.3.1.3 Results

ANOVAs were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). Error responses were excluded from the analysis of reaction times. If the standard score for a reaction time was outside the range of +/- 2.5 for a given participant, the reaction time was adjusted to that equivalent to a standard score of +/- 2.5. Table 6.7 presents the mean reaction times and error rates for Experiment 6.

Error analysis. The overall error rate was very low at 2.63%. Analyses only revealed significant main effects of word-formation principle, F_1 (3, 132) =

6.89, $p < .0001$; F_2 (3, 116) = 54.58, $p < .0001$.

Planned comparisons for the main effect of word-formation principle in the subject analysis revealed that errors in the modifier + modified condition were higher than in the verb + complement condition and in the synonymous pairs condition, HSD = 0.176. Similarly, planned comparisons for the main effect of word-formation principle in the item analysis revealed that errors in the modifier + modified condition were higher than in the verb + complement condition and in the synonymous pairs condition, HSD = 0.40.

Reaction times analysis. In the subject analysis, although there was no significant effects for prime, $p = .159$, there was a significant effect of word-formation principle, F_1 (3, 132) = 6.915, $p < .0001$, with no significant interaction. In the item analysis, no significant effects were found. Reaction times for Experiment 6 are shown in Figure 6.9.

Planned comparisons for the main effect of word-formation principle in the subject analysis revealed that reaction times in the modifier + modified condition were slower than in other conditions, with HSD = 10.9 in all cases.

Table 6.7

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 6

Word Formation Principle	Error rate	Reaction times		RT diff
Modifier + modified				
First constituent	3.1	604	(127)	+3
Second constituent	4.7	614	(129)	-7
Unprimed	5.3	607	(107)	
Verb + complement				
First constituent	2.0	587	(110)	+5
Second constituent	0.2	604	(139)	-12
Unprimed	2.2	592	(127)	
Complement + verb				
First constituent	2.7	582	(101)	+22
Second constituent	2.9	596	(126)	+8
Unprimed	2.4	604	(126)	
Synonymous pairs				
First constituent	1.8	593	(125)	-3
Second constituent	2.7	588	(111)	+2
Unprimed	1.6	590	(114)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unprimed condition reaction time in each word formation principle condition.

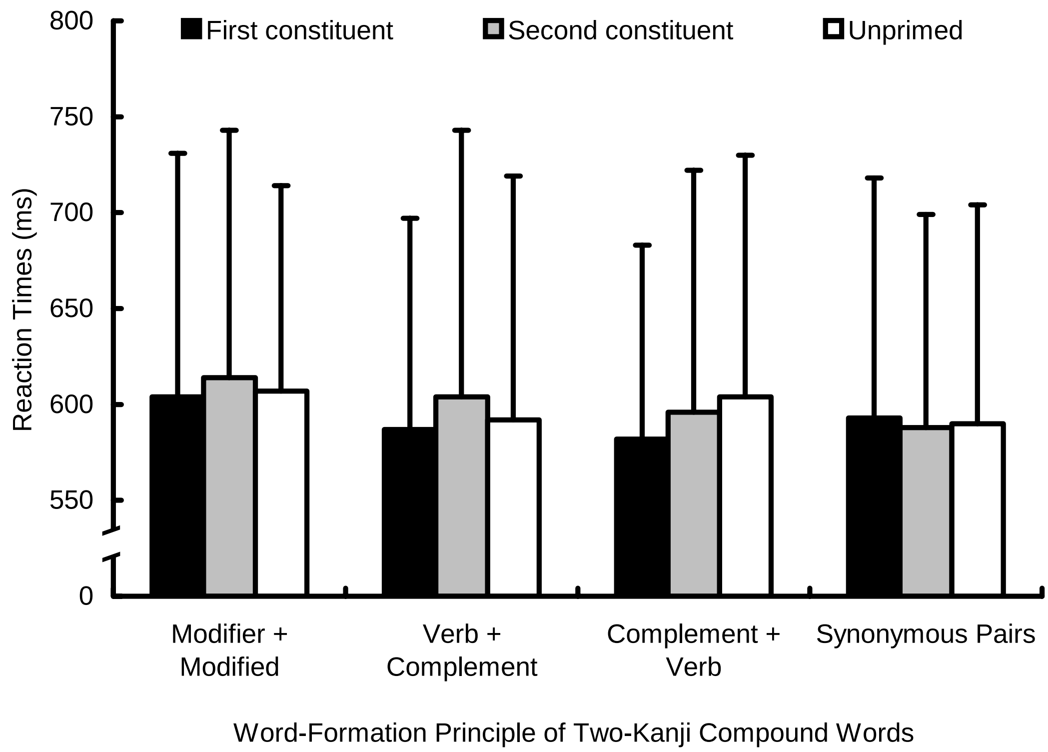


Figure 6.9. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 6.

6.3.1.4 Discussion

The results of this experiment show no priming effects in the first constituent and second constituent conditions when the on-reading of a constituent morpheme for a two-kanji compound word was auditorily presented.

The lack of priming in this experiment with on-reading primes is probably related to the fact that an on-reading will be shared by many kanji, which means that the auditory presentation of a on-reading is not very effective in priming the visual word recognition of a two-kanji compound word with that reading for one of its constituents.

6.3.2 Experiment 7: Kun-reading primes

6.3.2.1 Purpose

The purpose of this experiment is to investigate whether the auditorily presented kun-readings for the constituent-morpheme kanji of two-kanji compound word will facilitate recognition of the two-kanji compound word when presented visually. Such a finding would support the existence of connections between phonological access representations and lemma units within the lemma unit model.

6.3.2.2 Method

Participants. Forty-five native Japanese students (average age 19.3, SD = 0.6) of the University of Tsukuba participated in the experiment as volunteers. None of the participants took part in the previous experiments.

Design and Materials. The 3 x 4 two-factor design with both factors as within-subject variables in Experiment 7 is the same as that for Experiment 5.

The two-kanji compound word targets used in Experiment 7 are the same as those used in Experiment 5.

Wave files for the auditory primes were created from the sound files in the NTT database (Amano & Kondō, 1999), which were edited for included silence using sound editing software (Sound Forge XP, version 4.5c). As the average duration of these sound files 497 ms (SD = 131, range 220 ms to 808 ms), a sound file with 500 ms of silence was prepared for the unprimed condition.

Apparatus. The apparatus used in Experiment 7 is the same as that used in Experiment 5.

Procedure. The presentation procedure for Experiment 7 is similar to Experiment 5, but with the kun-reading prime being auditorily presented through headphones, as shown in Figure 6.10.

At the start of a trial, a plus symbol (+) appeared in the center of the screen as a fixation point for 250 ms. This remained on the screen as the kun-reading prime was presented auditorily through headphones. Presentation of the target stimulus item at the center of the computer screen immediately followed offset of the auditory prime. This remained on the screen until the participant pressed a button on the response box for the lexical decision. There was a 1,500 ms inter-trial interval. Participants were instructed to press a green button for a compound word and a red button for a non-word as quickly and as accurately as possible. The instructions to the participants were printed on a card, which they were asked to read, as shown in Appendix M. After briefly repeating the instructions verbally to confirm that they had been understood, there was practice session of ten trials, and the whole experiment took between 15 to 20 minutes to complete.

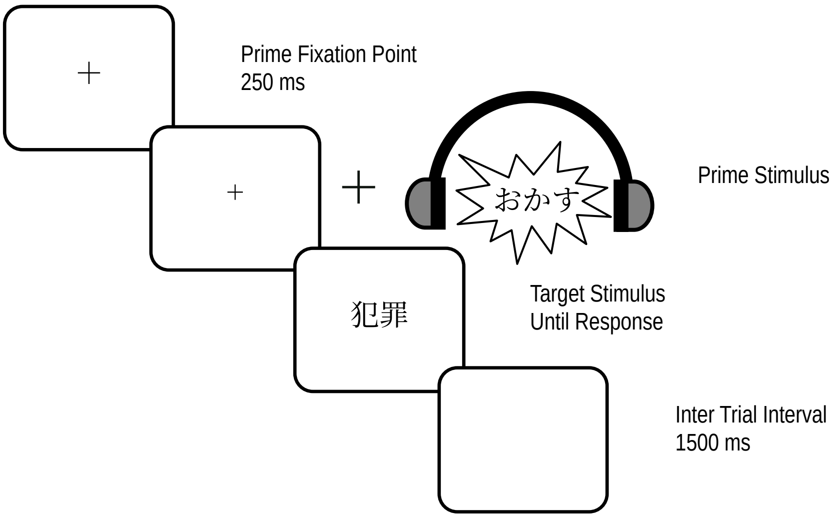


Figure 6.10. Presentation sequence during a trial in Experiment 7.

6.3.2.3 Results

ANOVAs were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). Error responses were excluded from the analysis of reaction times. If the standard score for a reaction time was outside the range of +/- 2.5 for a given participant, the reaction time was adjusted to that equivalent to a standard score of +/- 2.5. Table 6.8 presents the mean reaction times and error rates for Experiment 7.

Table 6.8

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 7

Word Formation Principle		Error rate	Reaction times		RT diff
Modifier + modified					
	First constituent	3.1	604	(127)	+3
	Second constituent	4.7	614	(129)	-7
	Unprimed	5.3	607	(107)	
Verb + complement					
	First constituent	2.0	587	(110)	+5
	Second constituent	0.2	604	(139)	-12
	Unprimed	2.2	592	(127)	
Complement + verb					
	First constituent	2.7	582	(101)	+22
	Second constituent	2.9	596	(126)	+8
	Unprimed	2.4	604	(126)	
Synonymous pairs					
	First constituent	1.8	593	(125)	-3
	Second constituent	2.7	588	(111)	+2
	Unprimed	1.6	590	(114)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unprimed condition reaction time in each word formation principle condition.

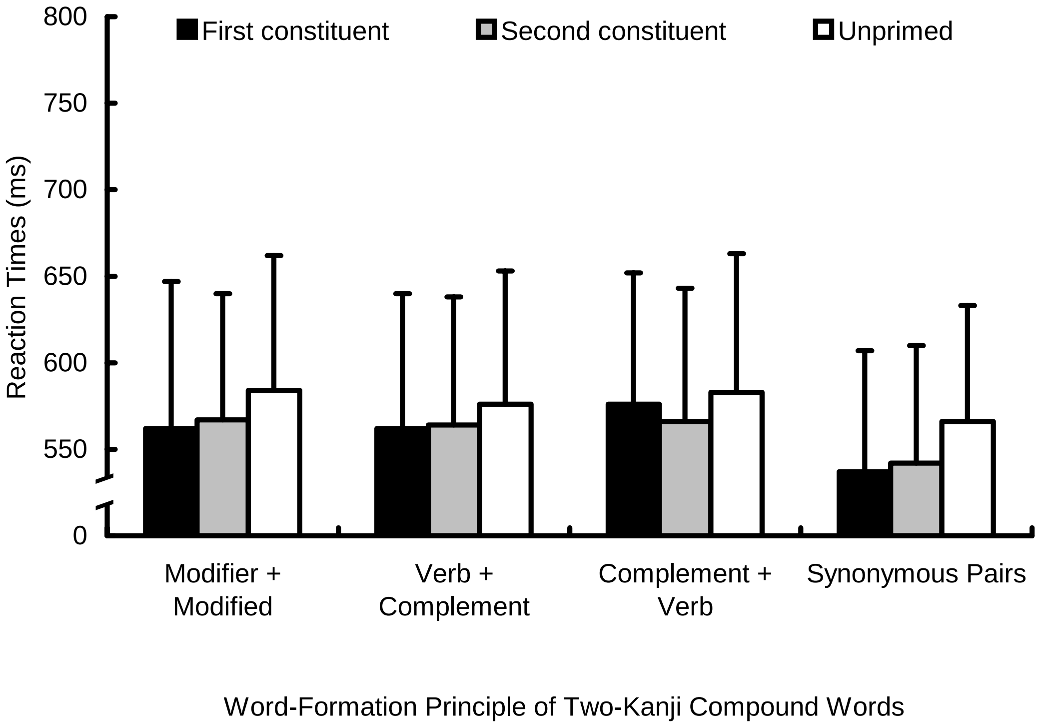


Figure 6.11. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 7.

Error analysis. The overall error rate was very low at 3.13%. Analyses only revealed significant main effects of word-formation principle in the subject analysis, F_1 (3, 132) = 3.943, $p < .05$.

Planned comparisons for the main effect of word-formation principle in the subject analyses revealed that errors in the synonymous pairs condition were lower in both the modifier + modified and the verb + complement conditions, HSD = 0.158

Reaction times analysis. In the subject analysis, there were significant main effects of prime, F_1 (2, 88) = 9.22, $p < .0001$, and of word-formation principle,

F_1 (3, 132) = 15.57, $p < .0001$, although there was no significant interaction, $p = .491$. In the item analysis, there was only a significant main effect of prime, F_2 (2, 232) = 5.96, $p < .005$. Reaction times for Experiment 7 are shown in Figure 6.11.

Planned comparisons for the main effect of prime in the subject analysis revealed that reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unprimed condition for all conditions. However, there were no significant differences in reaction times between the first constituent and the constituent conditions, with HSD = 11.34 in all cases. Similarly, planned comparisons for the main effect of prime in the item analysis revealed that reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unprimed condition for all conditions. However, there were no significant differences in reaction times between the first constituent and the constituent conditions, with HSD = 12.26 in all cases.

Planned comparisons for the main effect of word-formation principle in the subject analysis revealed that reaction times in the synonymous pairs were faster than in all other conditions, with HSD = 10.98 in all cases.

6.3.2.4 Discussion

The results of this experiment show significant priming in both the first constituent and second constituent conditions compared to the unprimed condition, but no significant differences between the first constituent and the second constituent condition.

Similar to the cross-script priming experiment with kun-reading primes (Experiment 5), with the availability of semantic information from the auditorily presented kun-reading primes, both constituent conditions primed lexical decisions for two-kanji compound words.

The presence of cross-modality priming in this experiment supports the lemma unit model, where phonological access codes are assumed to link to lemma units. The fact that priming was not found for auditorily presented on-readings but was found for both constituent conditions in this experiment is consistent with the pattern of results for the two cross-script priming experiments (Experiments 4 and 5), and further underscores the differences in the characteristics of on-readings and kun-readings.

6.4 Positional sensitivity

One aspect of the results from Experiments 1 and 2 that needs to be investigated further is the significantly faster reaction times for the first constituent compared to the second constituent in the verb + complement condition. This was the only word-formation principle condition where the priming from the constituents was significantly different.

As we shall seen shortly, Taft, Zhu, and Peng (1999) argue that character representations are not positionally sensitive within Chinese compound words. However, the frequency with which a kanji will appear in a given position within a two-kanji compound is likely to be closely related to its word class, for this will, in turn, determine the kind of word-formation principles in which the kanji will be involved. Given that frequency has a central role as a regulating mechanism within the lemma-unit model, the influence of word class and two-kanji compound word morphology cannot be ruled out. Within the model, lemma units function as a means of differentiating and addressing concepts, and morphological relations are expressed in terms of shared semantic and syntactic properties, which would include restrictions on the combinations of morphemes, such as verb and direct object, as well as information about word class and element order. The positional frequency of a kanji within two-kanji compound words is likely to be a product of its word class and the kinds of word-formation principles it will be involved in. We may, therefore, reasonably speculate about whether the significant difference between the constituent conditions in the verb + complement condition is due to positional sensitivity coupled with the characteristics of verb + complement compound words, which, as noted earlier, only involve Sino-Japanese elements and are limited to mainly verb + direct object combinations.

The purpose of the positional sensitivity study was to investigate this possibility, by examining the pattern of facilitation for verb + complement and the reversed order of complement + verb two-kanji compound words, once the positional frequency of the constituent verb is controlled for with. If the results in Experiments 1 and 2 were due only to the frequencies of the verb in the verb + complement compound words, then we would not expect to find significantly faster reaction times compared to the second constituent when the positional frequency of the verb is low.

6.4.1 Survey 2: Constituent morpheme frequency data Joyce and Ohta (in press)

6.4.1.1 Purpose

The aim of this survey was to obtain cumulative frequency data for the constituent morphemes of two-kanji compound words. This data was used in controlling for the positional frequency of the constituent verbs in both verb + complement and complement + verb word-formation principles in Experiment 8 which examines the possible effect of positional sensitivity.

6.4.1.2 Method

Although it is relatively easy to look up word frequency and word familiarity data from existing surveys (Amano & Kondō, 1999, 2000; National Language Research Institute, 1962, 1976; but see also Yokoyama, Sasahara, Long, & Tanimoto, 2001), more specific kinds of data such as cumulative frequencies must be calculated from raw word frequency data. Compiling this kind of data can, however, be very time-consuming, particularly when the raw data is not in electronic form. The cumulative frequency data for the constituent morphemes of two-kanji compound words collated in this survey was compiled from the newspaper frequency data in the NTT database (Amano & Kondō, 2000) by surveying two-kanji compound words for a six-year period (1993-1998).

All two-kanji compound words were extracted from the NTT newspaper frequency data (Amano & Kondō, 2000) for the six years 1993-1998. Although the NTT database includes newspaper data for 14 years (1985-1998), the large increases in article coverage both in 1987 and 1993 make it impossible to reliably average the type and token counts across all 14 years. Following the inclusion of the Sunday editions and sport page articles in 1993, the average coverage for the six-year period 1993-1998 was 76,556 articles (S.D. 1,476), representing an average increase of 12,069 articles compared to the previous five-year period (Amano & Kondō, 2000, p. xxv).

After eliminating a number of two-kanji compound words composed of kanji outside the 2,965 JIS level 1 set, the items were grouped according to their constituent kanji. Cumulative type and token frequencies were calculated for these constituent family groups, with separate counts for all compound words and for compound words excluding proper nouns based on the NTT database's classification of word class.

6.4.1.3 Results

Table 6.9 shows the annual type and token counts, together with the total type and token counts for the six-year period, for all compounds and compound words excluding proper nouns.

Table 6.9

Annual Type and Token Counts, Together with Total Type and Token Counts for the Six-year Period (1993-1998)

	Types		Tokens	
Year	All compound words	Excluding proper nouns	All compound words	Excluding proper nouns
1993	42,474	24,457	3,984,370	3,472,674
1994	42,928	24,763	4,034,259	3,510,526
1995	43,641	25,116	4,176,237	3,628,669
1996	44,008	25,600	4,283,915	3,739,755
1997	43,985	25,595	4,177,732	3,653,244
1998	43,442	25,348	4,218,128	3,707,865
	69,528	35,647	24,874,641	21,712,733

Based on this corpus, cumulative type and token frequency data for the constituent morphemes of two-kanji compound words (excluding proper nouns)[88] was calculated for the 2,965 JIS level 1 kanji.[89] This data includes the total type count for the six-year period, the average annual type count and the average annual token counts as a function of constituent position. A sample of the first twenty kanji with these counts is shown in Table 6.10.

Table 6.10

Example Kanji with Cumulative Type and Token Frequency Data

	Types	Tokens		Types	Tokens
	K1・K2	K1 ・ K2		K1・ K2	K1・ K2
亜	10(5.5)・3(3.0)	63.5 ・ 55.2	茜	3(0.7)・0(0)	0.8 ・ 0
唖	1(1.0)・2(1.0)	9.2 ・ 2	穐	0(0)・0(0)	0 ・ 0
娃	0(0)・0(0)	0 ・ 0	悪	75(52)・24(18)	2448 ・ 479
阿	7(3.8)・0(0)	12.3 ・ 0	握	3(2.2)・2(1.2)	203 ・ 515
哀	18(13)・1(1.0)	138 ・ 46.3	渥	0(0)・0(0)	0 ・ 0
愛	45(35)・17(14)	1027 ・ 434	旭	2(1.3)・0(0)	15.3 ・ 0
挨	1(1.0)・0(0)	25.8 ・ 0	葦	4(1.2)・1(0.2)	3.2 ・ 0.2
姶	0(0)・0(0)	0 ・ 0	芦	0(0)・0(0)	0 ・ 0
逢	3(1.7)・0(0)	2.5 ・ 0	鯵	0(0)・0(0)	0 ・ 0
葵	1(0.8)・2(0.8)	1.0 ・ 0.8	梓	0(0)・0(0)	0 ・ 0

Note. These figures represent total type, average type (in parentheses) and average token counts, as a function of constituent position (first kanji (K1) and second kanji (K2))

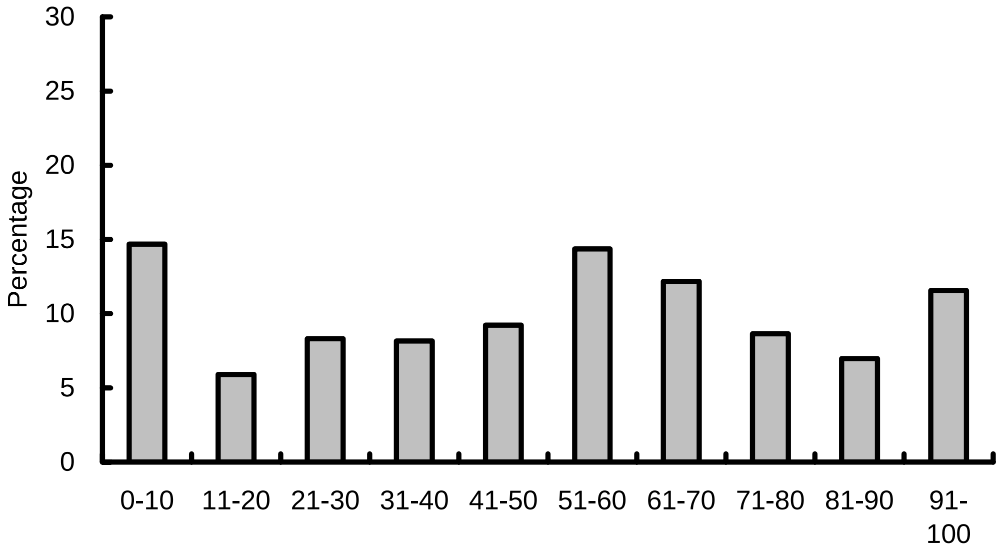
6.4.1.4 Discussion

The aim of this survey was to obtain cumulative frequency data for the constituent morphemes of two-kanji compound words. This data was used in controlling for the positional frequency of the constituent verbs in both verb + complement and complement + verb word-formation principles in the positional sensitivity experiment.

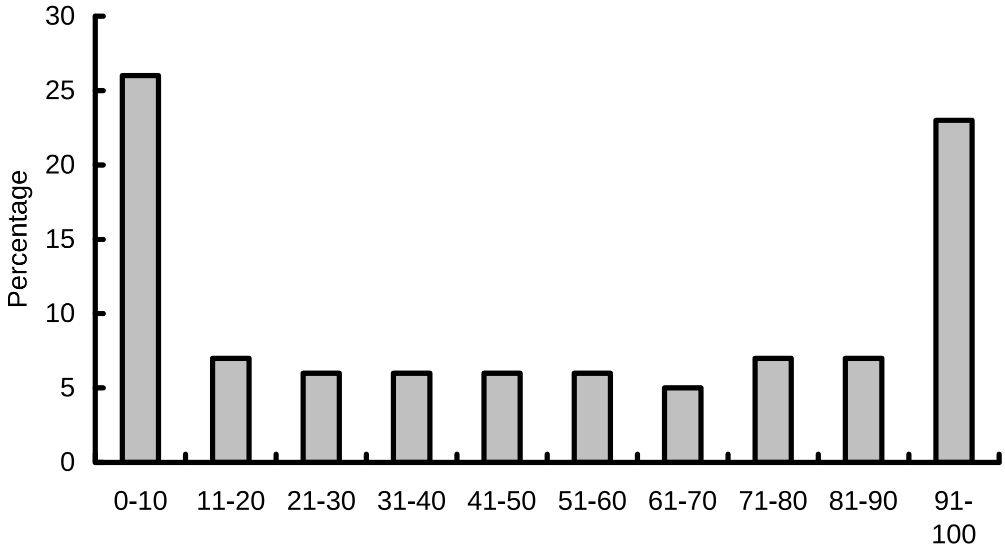
Discussing Chinese character and word recognition in terms of a multilevel interactive-activation framework, Taft, Zhu, & Peng (1999) argue that while the representations of radicals (including both semantic radicals and phonetic marker components) are positionally specific, positional information is not so important for character-level representations. On the assumption that responses to characters consisting of transposable radicals (i.e., the radicals 口 and 木 in 呆 and 杏) would be slower due to activation of both characters if radical representations are position-free, Taft, et al. regarded the lack of interference in both character decision and naming as evidence that positional information is part of the radical representation. Extending the same logic to two-character compound words consisting of transposable characters (such as 社会 and 会社 in Japanese), Taft, et al. interpreted interference in lexical decision as indicating that characters have position-free representations. However, while it may be the case that many characters have position-free representations, it is not clear whether this is true for all characters. As we have suggested, because two-kanji compound words are formed according to a number of word formation principles, such as modifier + modified, verb + complement and synonymous pairs, the positional frequencies of constituent kanji are likely to be closely linked to word class. For example, consistent with modifiers proceeding modified elements, 赤 'red' is the first constituent of 65 types but only the second constituent of 3 types, which have cumulative token counts of 2,531 and 63 respectively. Positional frequency information of this kind is assumed in the lemma-unit model, although the locus of this kind of cumulative frequency information is open to debate (see Baayen, et al. 1997; De

Jong, et al., 2000). Frequency effects may be an attribute of access representations or of lemma units, which function to differentiate and address concepts and are seen as representing information about word class and element order.

In a survey of radical frequencies based on the 2,965 JIS level 1 kanji, Saito, Kawakami, and Masuda (1995) report that for left-right (hen + tsukuri) structured complex kanji, 82% of the 857 radical components are positionally fixed (11% on the left and 71% on the right only), with the remaining 18% being free-floating appearing in both the left and right positions. From the cumulative type and token counts presented in this paper, it is possible to calculate ratios of positional frequency or positional sensitivity for the constituent kanji of two-kanji compound words. Taking the cumulative type counts, for instance, to calculate how often a kanji appears as the first constituent of two-kanji compound words, 以 would have a positional ratio of 100% (appearing as first kanji in 17 types, but never as second constituent), whereas 務 would have a positional ratio of 0% (never as first constituent, but appearing as second constituent in 63 types). Figure 6.12 shows the distributions of positional frequency ratios based on (a) cumulative type counts and (b) cumulative token counts for 2,743 kanji.[90]



(a) Positional Frequency Ratio based on Type Counts



(b) Positional Frequency Ratio based on Token Counts

Figure 6.12. Distribution of positional frequency ratios for 2,743 kanji based on (a) cumulative type counts and (b) cumulative token counts.

In contrast to the strong trend observed by Saito et al. for most radicals (82%) to be positionally-fixed within kanji, only 610 kanji (22%) have absolute ratio values, indicating that these kanji always appear in the same position within two-kanji compound words (342 kanji (12%) have a 100% ratio and 268 kanji (10%) have a 0% ratio). However, as the figures show the distributions for positional frequency ratios are not normal, particularly in the case of ratios based on token counts. While it seems reasonable to assume that characters with mid-range positional ratios have position-free representations, the question of whether characters with more extreme positional ratio scores also have position-free representations would seem to warrant further investigation, which is examined in Experiment 8.

6.4.2 Experiment 8: Positional sensitivity experiment

6.4.2.1 Purpose

The purpose of this experiment is to investigate further results of Experiments 1 and 2, where the reaction times for the first constituent compared to the second constituent in the verb + complement condition only. Specifically, this experiment compares the pattern of facilitation for verb + complement compound words and for compound words with the reversed order of complement + verb, once the positional frequency of the constituent verb is controlled for. If the results in Experiments 1 and 2 were due only to the frequencies of the verb in the verb + complement compound words, then we would not expect to find significantly faster reaction times compared to the second constituent when the positional frequency of the verb is low.

6.4.2.2 Method

Participants. Forty-five native Japanese students (average age 20.2, SD = 3.7) of the University of Tsukuba participated in the experiment as volunteers. None of the participants took part in the previous experiments.

Design and Materials. A 3 x 4 two-factor design was used, with both factors as within-subject variables. The three prime conditions are first constituent, second constituent, and unprimed.

Table 6.11

Word-Formation Principle Conditions in Experiment 8

Condition	Example compounds	
LPR-Verb + Complement	止血	迎春
HPR-Verb + Complement	預金	被災
Complement + LPR-Verb	順延	空輸
Complement + HPR-Verb	専務	共済

Note. LPR = low positional ratio and HPR = high positional ratio

Based on the results of the constituent-morpheme frequency data (Joyce & Ohta, in press), positional frequency ratios were calculated for verbs, and these were used in selecting two-kanji compounds words representing the word-formation principles of verb + complement and complement + verb, where the verb had either a low positional ratio (LPR) of less than 20% or a high positional ratio (HPR) of over 80%. The four word-formation principle conditions are shown in Table 6.11

A problem for the selection of two-kanji compound word target stimulus items was the fact that a verbal kanji appropriate for a particular condition would also be appropriate for an opposing condition. For example, although 止 has a low positional frequency for the first constituent position and so the compound word 止血 can be classified as an example of the LPR-verb + complement condition, 止 has a high positional ratio for the second constituent condition being compound words where it appears in that position possible candidates for the complement + HPR-verb condition. Given this problem with the selection of stimulus items, as well as the fact that searches of the NTT database (Amano & Kondō, 1999) had yielded relatively few suitable compound word with a familiarity of over 5.5, it was necessary for this experiment to extend the familiarity range to over 5.0. With this modification to the selection criteria, for each of the four word-formation principle conditions, 18 compound words were selected with a classification scores of 5.5 or over on a 7-point scale, a familiarity score of over 5.0 or over on a 7-point scale and an appropriate positional ratio for a given condition. The two-kanji compound word targets for Experiment 8 are shown with classification and familiarity scores in Appendix N. The mean scores for both criteria over the four word-formation principles, together with the mean positional ratios, are shown in Table 6.12.

Table 6.12

Mean Classification Scores and Familiarity Scores (NTT database), as well as Mean Positional Ratios for Target Compound Word Used in Experiment 8

Word formation principle	Classification score	Familiarity score	Positional ratio
LPR-Verb + Complement	6.4 (0.99)	5.58 (0.29)	10%
HPR-Verb + Complement	6.5 (0.84)	5.58 (0.31)	96%
Complement + LPR-Verb	6.2 (1.24)	5.58 (0.34)	10%
Complement + HPR-Verb	6.3 (1.01)	5.58 (0.21)	96%

Note. Both survey scores used a 7-point scale. The figures in parenthesis are the standard deviations.

Apparatus. The apparatus used in Experiment 8 is the same as that used in Experiment 7.

Procedure. The procedure used in Experiment 8 is the same as that used in Experiment 2.

6.4.2.3 Results

ANOVAs were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). Error responses were excluded from the analysis of reaction times. If the standard score for a reaction time was outside the range of +/- 2.5 for a given participant, the reaction time was adjusted to that equivalent to a standard score of +/- 2.5. Table 6.13 presents the mean reaction times and error rates for Experiment 8.

Error analysis. The overall error rate was low at 5.75%. Analyses only revealed significant main effects of prime, F_1 (2, 88) = 5.816, $p < .004$, F_1 (2, 128) = 7.571, $p < .001$,

Planned comparisons for the main effect of prime in the subject analyses revealed that errors in the first constituent condition were lower than in the unprimed condition, HSD = 0.168

Reaction times analysis. In the subject analysis, there were significant main effects of prime, F_1 (2, 88) = 68.71, $p < .0001$, and of word-formation principle, F_1 (3, 132) = 5.97, $p < .001$, as well as significant interaction, F_1 (6, 264) = 3.76, $p < .001$. In the item analysis, there was a significant main effect of prime, F_2 (2, 128) = 49.40, $p < .0001$, as well as significant interaction, F_2 (6, 128) = 2.47, $p < .027$. Reaction times for Experiment 8 are shown in Figure 6.13.

Planned comparisons for the main effect of prime for each word-formation condition revealed that reaction times in both the first constituent and the second constituent conditions were significantly faster than in the unprimed condition for all word formation conditions. There were no significant differences between the first constituent and the second constituent conditions in the low positional frequency ratio conditions, F_1 (2, 88) = 13.29, $p < .0001$ at LPR-verb + complement; F_1 (2, 88) = 24.76, $p < .0001$ at complement + LPR-verb.

Table 6.13

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 8

Word Formation Principle		Error rate	Reaction times		RT diff
LPR-Verb + Complement					
	First constituent	0.7	580	(117)	+56
	Second constituent	6.7	602	(71)	+34
	Unprimed	8.5	636	(102)	
HPR-Verb + Complement					
	First constituent	3.3	563	(99)	+73
	Second constituent	7.0	592	(93)	+44
	Unprimed	7.8	636	(96)	
Complement + LPR-Verb					
	First constituent	6.3	589	(101)	+80
	Second constituent	4.4	598	(103)	+71
	Unprimed	7.8	669	(116)	
Complement + HPR-Verb					
	First constituent	3.3	595	(106)	+39
	Second constituent	3.7	568	(92)	+66
	Unprimed	5.6	634	(98)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the unprimed condition reaction time in each word formation principle condition.

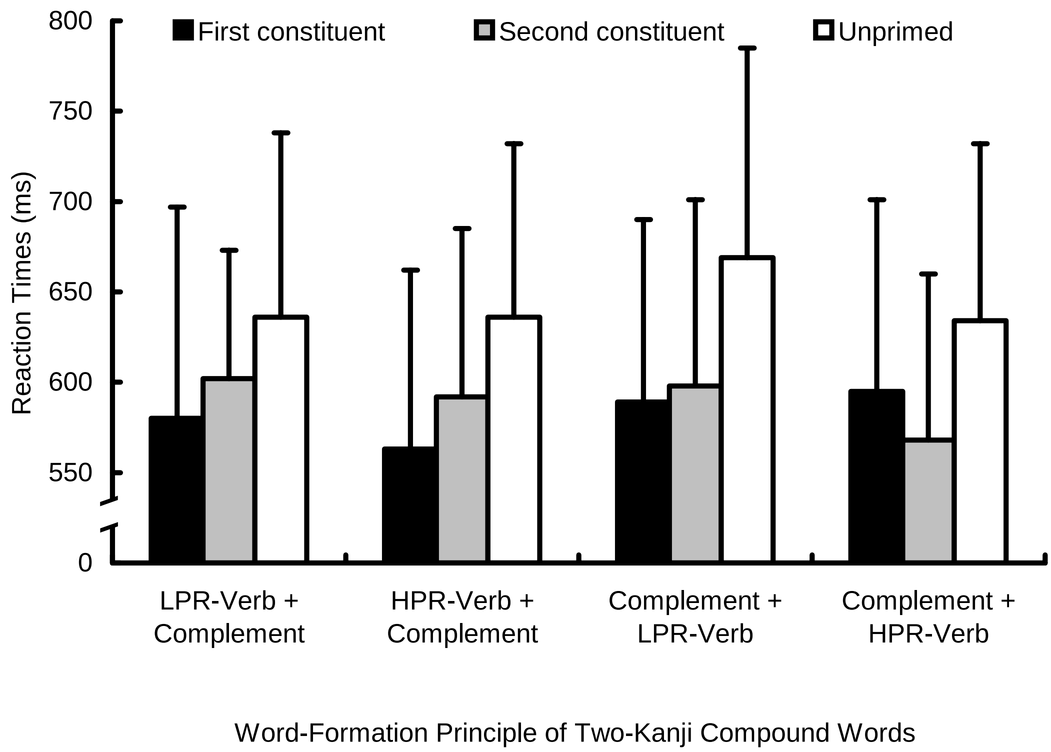


Figure 6.13. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 8.

However, when the positional frequency was high, there were significant differences between the first constituent and second constituent conditions, with the first constituent being faster than the second constituent in the HPR-verb + complement condition, F_1 (2, 88) = 28.43, p < .0001, but with the second constituent being faster than the first in the complement + HPR-verb, F_1 (2, 88) = 22.64, p < .0001, with HSD = 25.73 in all cases.

6.4.2.4 Discussion

The positional sensitivity experiment sought to investigate the finding from the constituent-morpheme priming studies (Experiments 1 and 2) where the only word-formation principle condition was the verb + complement condition with faster reactions times for the first constituent compared to the second constituent, by examining the pattern of facilitation for verb + complement and the reversed order of complement + verb two-kanji compound words, once the positional frequency of the constituent verb is controlled for.

The results from this experiment for the two verb + complement conditions would seem to suggest that positional frequency may have been the cause of the significant differences in the constituent-morpheme studies because the differences in this experiment were only significant when the positional frequency ratio of the verb was high. However, the fact that a significant difference was found in the opposite direction, with faster reaction times in the second constituent condition than the first constituent condition, in the reversed order of complement + verb when the positional frequency ratio of the verb was high suggests that actually semantic information from the verb is influencing the pattern of priming in constituent-morpheme when the positional frequency of the verb is high.

6.5 Experiment 9: Compound-word priming

6.5.1 Purpose

The results of the constituent-morpheme priming studies (Experiments 1 and 2) and the results of the positional sensitivity experiment (Experiment 8) provide support for the lemma unit model that accounts for constituent-morpheme priming in terms of the activation passing from orthographic representations to lemma units. The results also suggest that the weightings for connections for verbal morphemes may be positional sensitivity, which, in turn, can be seen as being related to word-formation principles.

To further examine the influence of compound-word morphology, in the compound-word priming experiment the prime-target relation in the previous experiments is reversed. That is, this experiment seeks to examine the patterns of facilitation from compound words in naming the constituent morphemes. If facilitation is found in this experiment, it cannot be attributed to the activation of phonological information, because there would be no phonological overlap between the compound-word primes, which would be read according to the on-readings of the constituents, and the target words which would be the kun-readings of the constituent kanji.

6.5.2 Method

Participants. Sixty native Japanese students (average age 19.7, SD = 2.2) of the University of Tsukuba participated in the experiment as volunteers. None of the participants took part in the previous experiments.

Design and Materials. A 4 x 4 two-factor design was used, with both factors as within-subject variables. The four target conditions are shown in Table 6.14. The four word-formation principle conditions were modifier + modified, verb + complement, complement + verb, and synonymous pairs.

Table 6.14

Target Conditions in Experiment 9, Compound-word Priming

Condition	Compound prime	Target word
First constituent - Primed	犯罪	犯す
First constituent - Unprimed		犯す
Second constituent - Primed	犯罪	罪
Second constituent - Unprimed		罪

Procedure. The procedure for this experiment, as shown in Figure 6.14, is similar to that used in Experiment 3, and used the same equipment and methods to recording wave-form data for the naming responses from which to measure reactions times.

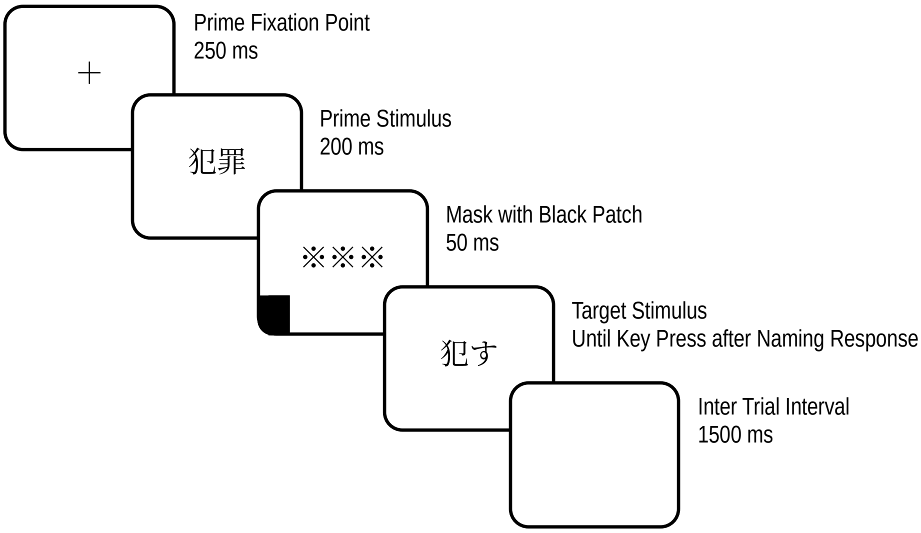


Figure 6.14. Presentation sequence during a trial in Experiment 9.

6.5.3 *Results*

Analyses of variance (ANOVAs) were carried out for the effects of prime and word-formation principle both by subject (both factors as within-subject variables) and item (prime as a within-subject and principle as a between-subject variable). Error responses were excluded from the analysis of reaction times. If the standard score for a reaction time was outside the range of +/- 2.5 for a given participant, the reaction time was adjusted to that equivalent to a standard score of +/- 2.5. Table 6.15 presents the mean reaction times and error rates for Experiment 8.

Error analysis. The overall error rate was very low at 1.22%. Analyses a revealed significant main effect of word-formation principle, F_1 (3, 177) = 4.33,

$p < .006$, and significant interaction, F_1 (9, 531) = 2.84, $p < .003$.

Planned comparisons for the main effect of word-formation principle in the subject analyses revealed that in the first constituent primed condition errors in the modifier + modified condition were higher than in both the verb + complement and the verb + complement conditions, F_1 (3, 177) = 5.13, $p < .002$; and in the first constituent unprimed condition errors in the modifier + modified condition were higher than in both the verb + complement and the verb + complement conditions, F_1 (3, 177) = 3.87, $p < .01$, HSD = 0.138 in all cases.

Reaction times analysis. The analysis of reaction times revealed that a significant main effect for prime, F_1 (3, 177) = 70.16, $p < .0001$, with significant interaction, F_1 (9, 531) = 11.32, $p < .0001$. Reaction times for Experiment 9 are shown in Figure 6.15.

Planned comparisons were conducted for the main effect of prime for each word-formation principle. In the modifier + modified condition, reaction times were significantly faster in the first constituent primed condition compared to the first constituent unprimed condition, F_1 (3, 177) = 40.38, $p < .0001$. In the verb + complement condition, reaction times were significantly faster in the first constituent primed condition compared to the first constituent unprimed condition, F_1 (3, 177) = 36.66, $p < .0001$. In the complement + verb condition, reaction times were significantly faster in the first constituent primed condition compared to the first constituent unprimed condition, and reaction times were significantly faster in the second constituent primed condition compared to the second constituent unprimed condition, F_1 (3, 177) = 21.70, $p < .0001$.

Table 6.15

Error Rates (as Percentages) and Mean Reaction Times (in milliseconds) as a Function of Word Formation Principle and Prime-Target Relation in Experiment 9

Word Formation Principle	Error rate	Reaction times		RT diff
Modifier + modified				
First constituent - Primed	3.1	516	(82)	+54
First constituent - Unprimed	2.9	570	(85)	
Second constituent - Primed	0	588	(89)	+3
Second constituent - Unprimed	1.7	591	(88)	
Verb + complement				
First constituent - Primed	0.7	513	(87)	+76
First constituent - Unprimed	0.5	589	(93)	
Second constituent - Primed	1.2	560	(88)	+17
Second constituent - Unprimed	0.2	577	(89)	
Complement + verb				
First constituent - Primed	0.2	550	(88)	+33
First constituent - Unprimed	0.7	583	(93)	
Second constituent - Primed	0.9	536	(88)	+55
Second constituent - Unprimed	1.4	591	(93)	
Synonymous pairs				
First constituent - Primed	1.4	528	(97)	+58
First constituent - Unprimed	1.2	586	(84)	
Second constituent - Primed	1.9	557	(102)	+44
Second constituent - Unprimed	1.4	601	(87)	

Note. Standard deviations are given in parentheses. RT difference = differences calculated from the matched unprimed condition reaction time.

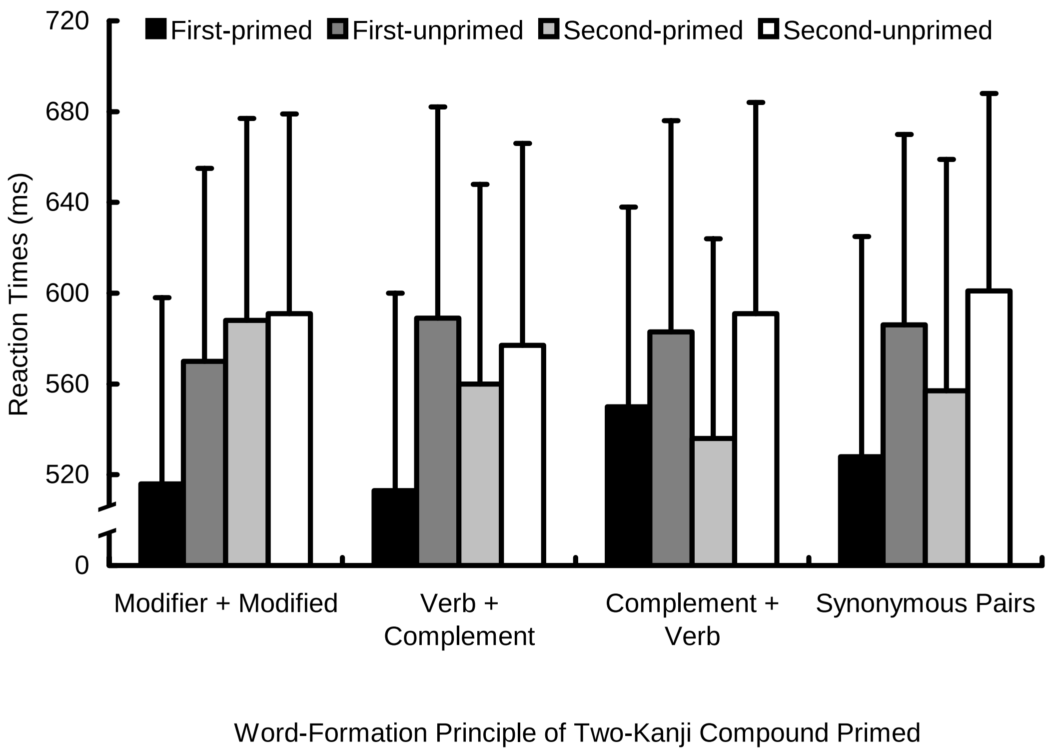


Figure 6.15. Mean reaction times (in milliseconds) as a function of word formation principle by prime-target relation in Experiment 9.

In the synonymous pairs, reaction times were significantly faster in the first constituent primed condition compared to the first constituent unprimed condition, and reaction times were significantly faster in the second constituent primed condition compared to the second constituent unprimed condition, F_1 (3, 177) = 36.60, $p < .0001$, HSD = 20.165 in all cases.

6.5.4 Discussion

This experiment sought to further investigate the influence of compound-word morphology, by reversing the prime-target relation that existed in the previous experiments. Exploiting the dual-reading system of the Japanese writing system, this experiment was able to remove any possibility of facilitation from phonological overlap, which would not be present with compound-word primes because these would be read according to the on-readings of the constituents, whereas the target words would be read according to the kun-readings of the constituent kanji.

The presentation of the compound words facilitated the naming latencies for the first constituent regardless of word-formation principle. However, the presentation of the compound words only facilitated the naming latencies for the second constituent in the complement + verb and the synonymous pair conditions. This result would seem to add further support for the notion that verbal morphology is important. The larger priming effect for the second constituent in the complement + verb condition seem to mirror the finding from the positional sensitivity experiment where priming from the second constituent prime was significantly greater than the priming from the first constituent condition in the case of the complement + HPR-verb condition. Moreover, as most of the synonymous pair compound words are combinations of verbal morphemes, the similar levels of priming for both the first constituents and the second constituents also suggests that verbal morphology is important, especially, as the semantic overlap between the compound word primes and the constituent targets would be greater in this word-formation principle.

6.6 The lemma unit model: Summary

This chapter has investigated the lemma unit model further through a series of experiments that examined various aspects of this model for the Japanese mental lexicon.

Experiment 3 investigated the extent to which the results in the first two experiments might be task specific, by changing the experimental task from lexical decision to naming. Although priming was still found in the first constituent condition, no significant differences were found in the second constituent condition, suggesting that the first constituent advantage found by Hirose (1982) may have been due to a bias towards phonological information that is emphasized in the naming task.

Experiments 4 and 5 were cross-script priming studies that investigated the possibility that the lemma unit model offers to integrate both kana and kanji processing within a single model. Experiment 4 investigated whether the on-readings for the constituent-morpheme kanji of two-kanji compound word presented in kana orthography would facilitate recognition of the two-kanji compound word target. Similar to Experiment 3, priming was only found in the first constituent condition, which may also be attributed to the emphasis on phonological information with on-reading primes. Experiment 5 investigated whether the kun-readings for the constituent-morpheme kanji of two-kanji compound word presented in kana orthography would facilitate recognition of the two-kanji compound word target. The results of this experiment showed significant priming in both the first constituent and second constituent conditions, although there were also significant differences between the first constituent and the second constituent conditions. In contrast to the bias towards phonological information in Experiment 4, with the semantic information that is available with kun-readings, both constituent conditions primed lexical decisions for two-kanji compound words, although it is not clear why there was greater facilitation in the first constituent conditions compared to the second constituent condition.

Experiments 6 and 7 were cross-modality priming studies that further investigated the lemma unit model's ability to account for the dual reading system of on-readings and kun-readings. Experiment 6 investigated whether the auditorily presented on-readings for the constituent-morpheme kanji of two-kanji compound word would facilitate recognition of the two-kanji compound word when presented visually. No priming was found in this experiment suggesting that because an on-reading is shared by many kanji, an auditorily presented on-reading is ineffective as a prime to the visual word recognition of a two-kanji compound word with that reading for one of its constituents. Experiment 7 investigated whether the auditorily presented kun-readings for the constituent-morpheme kanji of two-kanji compound word would facilitate recognition of the two-kanji compound word when presented visually. The results of this experiment showed significant priming in both the first constituent and second constituent conditions, but no significant differences between the first constituent and the second constituent condition. Similar to Experiment 5, with the availability of semantic information from the auditorily presented kun-reading primes, both constituent conditions primed lexical decisions for two-kanji compound words. The priming obtained in the series of cross-script and cross-modality priming experiments supports the lemma unit model, where connections are assumed to exist between orthographic representations, including both kanji and kana representations, and lemma units on the one hand, and between phonological representations, including both on-readings and kun-readings, and lemma units on the other hand.

One aspect of the results from Experiments 1 and 2 that required further investigation was the significantly faster reaction times for the first constituent compared to the second constituent in the verb + complement condition. As the positional frequency of a kanji within two-kanji compound words is likely to be a product of its word class and the kinds of word-formation principles it is involved in, it is possible that the pattern of facilitation in the verb + complement condition is due to a positional frequency, or positional sensitivity, of the verbal constituents. To investigate this possibility, the pattern of facilitation for verb + complement and the reversed order of complement + verb two-kanji compound words was examined, once the positional frequency of the constituent verb is controlled for. Survey 2 compiled cumulative frequency data for the constituent morphemes of two-kanji compound words from the newspaper frequency data in the NTT database (Amano & Kondō, 2000) by surveying two-kanji compound words for a six-year period (1993-1998). Based on the results of this survey, Experiment 8 compared the pattern of facilitation for verb + complement compound words and for compound words with the reversed order of complement + verb, when the verb had either a low positional ratio (LPR) or a high positional ratio (HPR). The finding of this experiment that reaction times were significantly faster for HPR verbal constituents compared to the other constituent in both the verb + complement and the reversed pattern of complement + verb supports the notion that morphological information is an important aspect in the lexical representation of two-kanji compound words, particularly for verbal constituents.

Experiment 9 examined further the influence of compound-word morphology, by reversing the prime-target relation of the previous experiments, to investigate the patterns of facilitation from two-kanji compound words primes on the naming of constituent targets. The results of this experiment showed that the presentation of the compound words facilitated the naming latencies for the first constituent regardless of word-formation principle. However, the presentation of the compound words only facilitated the naming latencies for the second constituent in the complement + verb and the synonymous pair conditions. This result provides further support for the notion that verbal morphology is important.

Although the results from this series of experiments raise further questions, they are generally consistent with the lemma unit model, which is undoubtedly an extremely appealing way of thinking about the complex relations between meaning, orthography and phonology that exist within the Japanese writing system.

Part 3

Conclusions and future research

Chapter 7

Conclusion:
Morphology and the mental lexicon

The central question addressed in this thesis has been *how do kanji function as a writing system?* The answer that has been argued for is that *kanji function as a morphographic writing system*, and, moreover, this is reflected in the organization of the Japanese mental lexicon.

In seeking to understand the significance of this question and its answer, this thesis has explored the linguistic issues surrounding the classification of the Japanese writing system and has conducted a psycholinguistic investigation of the Japanese mental lexicon. The shift in perspective in this thesis from external form to internal representation parallels the transition from printed word to lexical retrieval that is at the heart of visual word recognition and mental lexicon research. This shift in perspective also reflects the importance of classifications of writing systems for our understanding of the literate mental lexicon.

In summarizing the main points of this research, this concluding chapter briefly retraces the discussion from the four broad areas of this thesis, which given the importance of morphology in the organization of the mental lexicon has investigated the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon from a morphological perspective.

7.1 The Japanese writing system

After noting the contribution of its multi-script nature and dual reading system to the impression of the Japanese writing system as being extremely complex, Chapter 2 suggested that a major factor for this portrayal has been the heated debate and controversy surrounding the classification of the Japanese writing system and, in particular, kanji, within a linguistic taxonomy of writing systems.

It was argued that the central question in this thesis of how kanji function as a writing system is actually a very serious issue, turning not only on how we understand writing systems to have developed, but also on how we conceive speech and writing to be related, and even on what we believe language to be. The central issue is not simply about how kanji function as a writing system, but rather *how do writing systems function*—what are the theoretical possibilities for writing systems? We noted that to constitute a full writing system, the graphic symbols must be capable of 'conveying any and all thought' (DeFrancis (1989), which brings us to the single most important issue for understanding writing and how different types of writing systems function, namely, how does writing relate to language?

In contrast to the 'language is speech' view that holds that language should be conceived of in terms of speech, and that writing is merely a means of transcribing speech, with the graphic unit being defined as primarily representing units of speech (Bloomfield, 1933; Daniels, 2001; DeFrancis, 1989; Miller, 1967, 1986; Sproat, 2000; Unger, 1987), this thesis has argued for an alternative understanding of the relationships between language, speech and writing. While it does not deny that speech is the primary medium of expressing language for hearing persons, the abstract entity view of language does not treat sound as a defining characteristic of language, and does not privilege speech over writing, or even sign, as mediums of expressing language (Coulmas, 1989; Crystal, 1987; Garman, 1990; Halliday, 1985; Lyons, 1981; Morioka, 1968; Sandler & Lillo-Martin, 2001; Steinberg, Nagata & Aline, 2001).

The discussions of writing systems in Chapter 2 also noted that while the rebus principle has unquestionably been the key to the development of partial writing systems into full writing systems, the application of this principle in the creation of phonetic compound Chinese characters did not lead to the development of a cenemic writing system. Just as it is important to realize that kanji are a pleremic writing system, of semantically-informed graphic units, it is also essential to realize that 'pure' word writing systems do not, and cannot exist, because of the sheer numbers of symbols that would be required to constitute a full writing system. From this, comes the realization that pleremic writing systems must be morphographic writing system—writing systems where the orthographic units represent morphemes, which are the minimal units of meaning in a language.

The preference for the term morphographic over the widely used term logographic may seem to be overly pedantic, especially given the generally inclusive interpretation that appears to be accorded to the term logography recently and might be dismissed as just a matter of emphasis. However, as this thesis has argued, that would be a mistake because the issue is much very much more than merely a question of labels. The shift in thinking that this terminological revision involves is of fundamental significance not just for our understanding of kanji in the Japanese writing system, but of writing systems in general, and, in particular, for our notions about the organization of the literate mental lexicon.

7.2 Japanese word formation

In pursuing an answer to the central question of how kanji function as a writing system, the thesis has taken the two-kanji compound word as its principle frame of reference. The emphasis on the two-kanji compound word, rather than on the single kanji character, stemmed from a belief that to understand how a writing system functions it is essential to identify the principles that govern the combination of graphic units.

Following directly from the central claim that kanji function as a morphographic writing system, the discussions of Japanese word formation processes, and, in particular, of the morphology of two-kanji compound words in Chapter 3 provided evidence of the morphographic nature of kanji. As the discussion of the word-formation principles for two-kanji compound words illustrated compounding is the joining of morphemes into larger word units, with the graphemes representing the union visually.

7.3 Visual word recognition

While the basis for the claim that kanji function as a morphographic writing system made in Part 1 of the thesis was essentially linguistic, derived from our discussions of the issues surrounding the classification of the Japanese writing system, given that linguistic classifications of writing systems and psycholinguistic research into visual word recognition are intricately linked concerns, Part 2 of the thesis sought to examine the implications of this claim for organization of the Japanese mental lexicon.

As background to the psycholinguistic investigation of the Japanese mental lexicon presented in this thesis, Chapter 4 briefly reviewed visual word recognition and mental lexicon research. The discussion there focused on the two central issues for the mental lexicon of lexical retrieval and lexical representation. These were shown to be mutually defining aspects of the mental lexicon. The mechanisms of lexical retrieval that are assumed will limit and constrain the representation and arrangement of lexical information, as with, for example, the ordered arrangement of lexical representations based on shared features that a serial search model such as that proposed by Forster (1976, 1989) requires. Similarly, the representation of lexical information in the mental lexicon will greatly determine what kind of mechanisms can be assumed as underlying lexical retrieval, in the way, for example, that the hierarchical organization of lexical representations in the multilevel interactive-activation framework proposed by Taft (1991, 1994) presupposes activation mechanisms of lexical retrieval.

In reviewing a number of models concerned primarily with lexical representation, we observed that all allow for some form of morphological involvement in the lexical retrieval of polymorphemic words, which underscores the language-universal importance of morphology for the organization of the mental lexicon.

7.4 The Japanese mental lexicon

The lexical retrieval and representation of two-kanji compound words are fundamental issues for any model of the Japanese mental lexicon, which must be capable of accounting for the diversity in the morphological structure of two-kanji compound words, discussed in Chapter 3.

After a highly selective review of psycholinguistic research on kana and kanji processing, which focused on a few studies from both the Chinese and Japanese literature primarily concerned with two-kanji compound words, Chapter 5 considered two proposals for models of the Japanese mental lexicon that differ in terms of how they account for lexical retrieval and representation. The first proposal was Hirose's (1992, 1994; 1996) hypotheses that evoke search mechanisms, while the second proposal was Joyce's (1999; in press) suggestion for a Japanese lemma-unit version of the multilevel interactive-activation framework, which assumes activation mechanisms. These two proposals were compared in terms of their ability to cope with the diversity inherent in the morphology of two-kanji compound words. As the proposals make different predictions concerning constituent-morpheme priming, the patterns of facilitation from the constituents of two-kanji compound words were examined in two experiments with varying SOAs, with five word-formation principles as experimental conditions. In all five word-formation conditions, the reaction times associated with both constituent conditions were significantly faster than those for the unrelated condition. However, in all but one word-formation principle, reaction time differences between the first-element and second-element conditions were not significant. That is, the results from these two experiments show that both related prime conditions facilitated responses to the target, and, in the majority of cases, at similar levels. This pattern of facilitation is more consistent with the predictions from the Japanese lemma-unit model, supporting this as a model of the Japanese mental lexicon.

The validity of the Japanese lemma unit model was further investigated in the series of experiments presented in Chapter 6, which sought to examine various aspects of this model. While retaining their focus on the lexical retrieval and representation of two-kanji compound words from a morphological perspective, these experiments also investigated task-effects, the possibility that the lemma unit model offers to account for both kana and kanji processing within a single integrated model, and the potential of the model to model the dual reading system of on-readings and kun-readings. In addition, the positional sensitivity of verbal constituents was examined in an experiment that controlled for the positional frequency of a given verbal constituent within two-kanji compound words. Providing explanation for the pattern of facilitation in the verb + complement word-formation conditions in the constituent-morpheme priming studies (Experiments 1 and 2), the results of this experiment, which received additional support from the compound-word prime experiment, suggests that verbal morphological is an important aspect of the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon.

In concluding, I would like to make a few brief observations of a theoretical nature concerning the Japanese lemma unit version of the multilevel interactive-activation framework as an example of the connectionist modeling of cognitive processes. The first point that I would like to draw attention to is the structural similarity that exists between the schematic representation of the abstract entity view of language that was presented in Figure 2.5, and the way that semantic representations at one level and access representations (for the mediums of expression of speech and writing) at another level are presented in many of the connectionist models that have been discussed in this thesis, such as the interactive-activation and competition (IAC) model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), the distributed activation model (Seidenberg & McClelland, 1989), the multilevel interactive-activation framework (Taft, 1991; 1994), the distributed activation model for Japanese (Ijuin, et al, 1999), the companion activation model (CAM) (Saito, 1997), and the multilevel interactive-activation model for Japanese (Tamaoka & Hatsuzuka, 1989), as well as the Japanese lemma unit model (Joyce, 1999, in press). This is not, I would suggest, a matter of mere coincidence, but reflects an important understanding about the nature of language.[90]

A second point I would make relates to the distinction between localist and distributed connectionist models. Developing out of the multilevel interactive-activation framework that Taft (1991; 1994) has advocated and developed to account for the lexical retrieval of Chinese words (Taft, Liu, & Zhu, 1999; Taft & Zhu, 1995, 1997a, 1997b; Taft, Zhu, Peng, 1999), the Japanese lemma unit model is presented here as a localist connectionist model. However, while this is more for explanatory convenience, rather than reflecting a resolute commitment to the localist approach over the distributed approach to connectionist modeling (see for arguments in favor of the localist position, see Page, 2000),[91] in terms of their ability to account for the dual reading system of on-readings and kun-readings of the Japanese writing system, the lemma units in the Japanese lemma unit model would seem to resemble the layer of hidden units that Ijuin, et al (1999) implemented in their connectionist model of word naming for two-kanji compound words, which was presented in Section 5.1.3.2. To the extent that irregularities in the pronunciations of two-kanji compound words are a result of the hybrid compounding, mentioned in Section 3.1.1, where elements from different lexical stratums are combined, the kind of morphological information concerning allomorphic realizations as on-readings and kun-readings that the lemma unit attempt to provide would seem to be important for accounting for the skilled Japanese reader's ability to cope with the inconsistency in two-kanji compound word pronunciations.

The final theoretical observation to make relates to the division and representation of semantic, morphological, and syntactic information within the Japanese lemma unit model. Primarily proposed as a model of visual word recognition that is capable of providing a reasonable account of the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon from a morphological perspective, issues relating to how the balance of between semantic, morphological, and syntactic information should be implemented within the Japanese lemma unit model has not been discussed in detail. The notion of morphology that is incorporated in the form of the Japanese lemma unit model presented in this thesis actually comes to resemble Anderson's (1992) interpretation of morphology, noted earlier, as the study of the relations between words, which in the model are realized by the links mediated by the lemma units, rather than the study of the minimal signs combined in complex words.

However, it is not certain whether it will be possible to implement more complete models of the mental lexicon that attempt to account for the higher cognitive processes involved in comprehension in the future without a more principled division of semantic and syntactic information. Although we are still far from complete connectionist models of the human mental lexicon, if connectionist models, such as the Japanese lemma unit model that has been proposed here are to be extended into larger-scale models that can account for higher-level cognitive processing, further research into the complex nature of semantic representations will be essential.

While it is also a limited model concerned with the resolution of lexical ambiguity, the parallel distributed processing that Tsuzuki, Kawamoto, and Yukihiro (1999) have proposed for homophone Japanese nouns, as shown in Figure 7.1, includes representations for syntactic information.

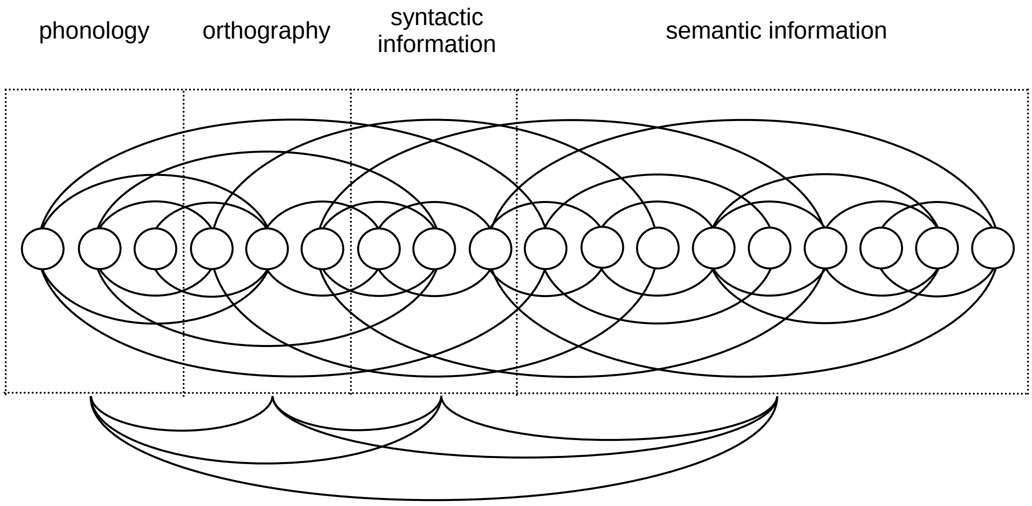


Figure 7.1. Tsuzuki, Kawamoto, and Yukihiro's (1999) parallel distributed processing model of lexical ambiguity resolution.

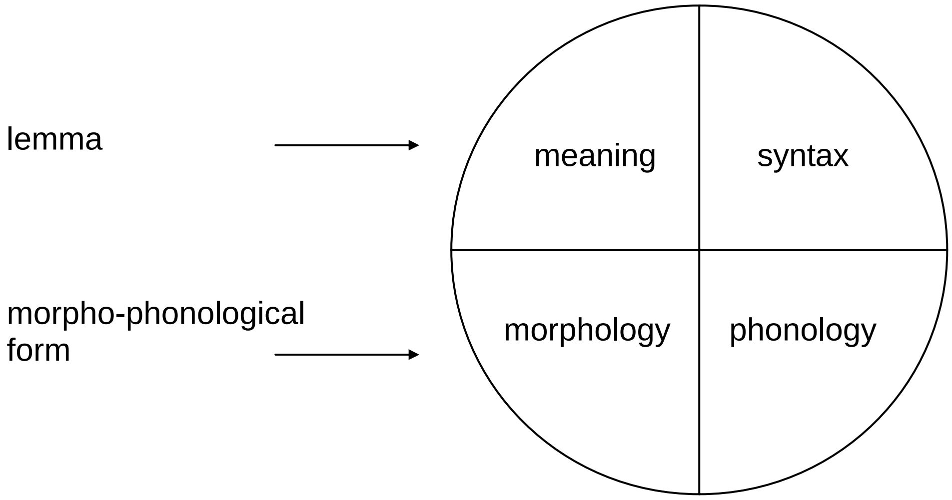


Figure 7.2. Levelt's (1989) notion of a lexical entry consisting of a lemma and morpho-phonological form (based on figures in Levelt, 1989, p. 182 and p. 188).

One possibility for the structure of a more comprehensive connectionist model of the mental lexicon could be to extend Levelt's (1989) notion of a lexical entry, as consisting of four groupings of information, as depicted in Figure 7.2. With access representations for orthography and phonology implemented as in the Japanese lemma unit model presented here, and with semantic and syntactic units divided as separate groupings of information, a layer of units in the center, similar to the proposed lemma units, could function as an interface between the different kinds of information which would constitute a lexical entry, while capturing the important aspects of lexical retrieval and representation that can be modeled by the Japanese lemma unit version of the multilevel interactive-activation framework.

While the theoretical issues that I have touched on here briefly are all important areas that need to be considered further, these are essentially issues relating to the connectionist modeling of cognitive processes that will be addressed in the future when attempts are made implement in a simulation the Japanese lemma unit model proposed in this thesis.

In investigating the lexical retrieval and representation of the two-kanji compound words within the mental lexicon from a morphological perspective, this research has provided important insights not only for our understanding of kanji, but also into the nature of writing systems, the organization of the literate mental lexicon, and of language itself. It can be no coincidence that pleremic writing systems must be morphographic in nature and that morphology is so central in the organization of the mental lexicon.

Chapter 8

Future research: Different procedures, different words, and different tasks

This thesis has presented a psycholinguistic investigation of the Japanese mental lexicon in support of the central claim made here that kanji function as a morphographic writing system. The psycholinguistic investigation examined the implications of this for the organization of the Japanese mental lexicon by investigating the lexical retrieval and representation of two-kanji compound words from a morphological perspective.

In particular, the Japanese lemma unit model (Joyce, 1999, in press) was investigated in a series of constituent-morpheme priming experiments where the morphology of two-kanji compound words is controlled as an experimental condition. Although the results from these experiments were generally consistent with the Japanese lemma unit model and are supportive of the model as a way of accounting for the complex relationships between meaning, orthography and phonology that exist within the Japanese writing system, they also raise further questions and suggest areas for further investigation in developing this model. In concluding, some of the areas for further research are discussed briefly.

8.1 Different procedures

As noted in Chapter 5, a focus of the psycholinguistic research on the processing of kana and kanji processing has been concerned with determining the balance between semantic activation and phonological activation and their relative time courses in lexical retrieval. One of the main tools that has been used for this is the manipulation of SOAs in priming experiments.

Although the pattern of facilitation found across the reported experiments has been generally consistent, one area for further research would be to attempt to replicate these results in similar experiments where the SOA is varied. For instance, conducted experiments where varying the SOA might be particularly interesting are the cross-script and cross-modality studies (Experiments 4 to 7) (Joyce, 2001). In the two cross-script experiments, the SOA was 300 ms (250 ms prime presentation and 50 ms mask) between the kana orthography on-reading primes (Exp. 4) and kun-reading primes (Exp. 5) and the two-kanji compound word targets. However, in the two cross-modality experiments, the two-kanji compound word targets were visually presented at prime offset, so in the case of the auditorily presented on-reading primes (Exp. 6), the average SOA was 342 ms, whereas in the case of the kun-reading primes (Exp. 7), the average SOA was 497 ms. Although the cross-script and cross modality priming experiments showed certain similarities, such as priming in both kun-reading prime experiments in both constituent conditions, there were also differences, with priming obtained in the first constituent condition in Experiment 4 but no priming in Experiment 6. However, for more direct comparisons of the results from these experiments, it would seem to be necessary to rule out any effects that might be due to the differences in the

SOAs, either by presenting auditory primes at an SOA of 300 ms regardless of the duration of the auditory prime stimulus, or by lengthening the SOAs in the cross-scripts to intervals of 340 ms and 500 ms.

8.2 Different words

There are a number of ways in which the research presented in this thesis could be extended based on using two-kanji compounds word with different characteristics and by using different kinds of polymorphemic words in the basic constituent-morpheme priming paradigm used here.

One potentially interesting area of research would be to see if morphological family effects, which were briefly discussed in Chapter 4, could be found in constituent-morpheme priming for two-kanji compound words.[93] The research on morphological families (Baayen, et al., 1997; Bertram, et al., 2000; Schreuder, & Baayen, 1997) has all been all conducted for languages using alphabetic writing systems. These studies reported an effect of morphological family size, although no effects were observed for cumulative family frequency. Although the data from the constituent-morpheme frequency survey only approaches Krott, et al.'s (2001) definition of a constituent family,[94] it could be used to investigate whether such morphological family effects could be replicated in word recognition experiments with kanji.

One important attribute of polymorphemic words is their degree of semantic transparency.[95] Some of the constituent-morpheme priming studies for European languages (Kehayia, et al., 1999; Monsell, 1985; Sandra, 1990; Zwisterlood, 1994) discussed in Chapter 4, and some of the studies for Chinese

(Liu & Peng, 1997; Taft, Liu, & Zhu 1999) discussed in Chapter 5, have dealt with the issue of semantic transparency for polymorphemic words. Although it was not explicitly conducted as a measure of semantic transparency, the classification scores obtained from the two-kanji compound word classification survey (Joyce & Ohta, 1999) for the two-kanji compound words used as stimulus in the presented experiments may be seen as a kind of indication of semantic transparency. The Japanese native speaker respondents in that survey were asked to evaluate on a 7-point scale the appropriateness of classifying a two-kanji compound word according to a given word-formation principle. To the extent that we can assume that this evaluation task required the respondents to identify the relationship between the constituents of the compound words, it would seem reasonable to treat compound words with high evaluation scores as being fairly transparent. Given that the two-kanji compound word target stimulus used in the conducted experiments all had classification scores of 5.5 or more, we may wonder how far the findings of these experiments may apply to opaque two-kanji compound words. While it is intuitively doubtful that there are two-kanji compound words that are simultaneously representative of a particular word-formation principle and semantic opaque, the influence of semantic transparency on the lexical retrieval and representation of polymorphemic words would certainly seem to warrant further consideration.

Although this thesis has focused on two-kanji compound words, many of the word-formation principles underlying two-kanji compound words are also shared by other polymorphemic words. While the classification of the Japanese lexicon into lexical stratums based on the origin of the words is not without problems, in the main, the two-kanji compound words used in the reported studies may be classified as Sino-Japanese. However, native Japanese word-formation processes are also extremely productive, particularly in the case of modifier + modified and complement + verb word-formation principles. Research using such native Japanese lexical items is, clearly, also extremely important for the development of a coherent model of the Japanese mental lexicon, especially in the light of the results of Experiment 8 which point to the importance of verbal morphology and given the productive native word-formation principle of complement + verb.

8.3 Different tasks

This research has employed the lexical decision task and the naming task in investigating constituent-morpheme priming for two-kanji compound words. Although these are the most commonly used tasks in visual word recognition, it is possible that tasks involving higher-level comprehension could provide further insights into morphological aspects of various Japanese polymorphemic words, including two-kanji compound words.

For instance, Hatano (1995) has discussed the Japanese writing system within the context of the practice account of literacy proposed by Scribner and Cole (1981), which assumes that both how literacy is acquired and what cognitive consequences it has depends on the set of socially organized literacy practices in which people engage, and has argued that there are cognitive consequences on modes of communication from acquiring and engaging in Japanese literacy practices. Specifically, he claims that literacy in Japanese includes the acquisition of compounding schemata, by which new words are made, and cognitive skills, that are required to solve homonymic ambiguity and to infer the meanings of unfamiliar words.

These claims are based on the experimental findings from Hatano, Kuhara, and Akiyama (1981), which sought to examine the claim by Suzuki (1975, 1978) that technical terms written in kanji are semantically more transparent than English words derived from Greek and Latin by testing the ability of college students to match technical terms with their definitions in a cross-cultural study of Japanese and Americans.[96] Three Japanese student groups were asked to match the words with their definitions in Japanese, with the word lists being presented in kanji, hiragana, and English, respectively. The American students were tested using the English definitions and words. The result of this study was that performance for the kanji condition was almost perfect, which contrasted with a correct matching of a little over 50% for the American students with the English words. Performance for the hiragana condition was also very high. According to Hatano (1995), this is because the participants were often able to retrieve the appropriate kanji to fit the phonetic constraints of the hiragana, and thus infer the meaning based on compounding schemata; a skill that Hatano (1995) suggests is a general cognitive consequence of Japanese literacy practices.

As mentioned earlier in discussing the Japanese lemma unit model, the compounding schemata that Hatano (1995) has proposed would seem to be similar to the processes of licensing and composition that Schreuder and Baayen discuss, which suggests that the tasks requiring comprehension, such as those used in the area of conceptual combination (Gagné & Shoben 1997; Hampton, 1991; Wisniewski & Gentner, 1991) may provide useful tools for investigating further the fundamental importance of morphology for understanding the mental lexicon, for as Bertram, et al. (2000) observe, much of the interconnectivity in the human mental lexicon would appear to be based on networks of morphologically related words.

Notes

Chapter 1 Introduction

1 The exceptional cases of 熟字訓 *jukujikun* words and 当て字 *ateji* words, which should be treated as monomorphemic words, are obviously excluded from this characterization of two-kanji compound words as polymorphemic words consisting of two morphemes. These exceptions will be mentioned briefly in Chapter 3, but basically jukujikun words can be defined as "words written with two or more kanji, the reading for which is not formed from the readings of the kanji elements if split" (Satô, 1996, p. 81). Paradis, Hagiwara, and Hildebrandt (1985) give a narrow definition of ateji as words written with two or more characters that are read according to the usual readings of the characters, but are semantically arbitrary from the meanings of the constituent kanji.

2 Anderson (1992) summarizes the classical view of morphemes according to which "morphemes are homogeneous and indivisible atomic units of linguistic form" and "words are exhaustively composed of morphemes" (p. 50). However, Anderson actually presents an alternative interpretation of morphology, as a study of relations between words, rather than a study of the minimal signs combined in complex words. Singleton (2000) defines morphemes as "the smallest elements of any language which have semantic and/or grammatical significance" (p. 34).

3 Yokosawa and Umeda (1988) do not state which dictionary was used in their study, but comment that this figure of 70 percent is based on checking approximately 52,000 headwords, for which the average word length was 2.4 characters. It should be noted, however, that the percentage of two-kanji compound words is not as high in more natural contexts. Looking at an inventory of Japanese common words, for example, Hatta and Kawakami (1996) (as cited in Kess & Miyamoto, 1999, p. 11) claim that compound words represented about 40 percent, although this is still a significant number.

Chapter 2 The Japanese writing system

4 Unger (1984, 1987) has even argued that, in the modern computer age, the complexity of this writing system is an unacceptable burden, regarding this as a major reason why Japan lags behind other advanced countries in terms of information technology (see also, Gottlieb. 1993; Mair, 1991).

5 While it undoubtedly warrants far more research than it has received to date, the question of just how functional literacy in Japanese might be defined is not touched on in this thesis (for various perspectives on literacy for Japanese, see Akita & Hatano, 1999; Hatano, 1995; Makita, 1968; Stevenson, Lee, Stigler, Kitamura, Kimura, & Kato, 1986; Taylor & Taylor, 1995).

It is worth a moment's pause in that context, however, to reflect on the comments of Kess and Miyamoto (1999) regarding the impact on kanji usage of easy accessibility with mechanical devices. "There may even come a time in Japan when there will be two character sets in actual practice; one will be for reading comprehension, a **read-only set**, while the other set will be a **write-only set**, the set that literate Japanese will have to be able to produce in handwriting tasks. … Perhaps the reality is that the Japanese are already there" (p. 30).

6 It should be noted that not all have taken a negative view of the Japanese writing system. Backhouse (1984), for example, comments on the mixture of scripts, remarking that it "makes for a potential flexibility of orthography on a scale that is inconceivable in the case of more familiar writing systems" (p. 220).

7 Although it was claimed as early as the late thirteenth century that writing had existed in Japan from the age of the gods, as Seeley (1991) points out, such claims about 神代文字 *jindai moji* 'God Age Script,' are clearly fabrications "dreamed up by Shintoist scholars who were unwilling to acknowledge that writing was one of the many cultural appurtenances for which Japan was at first dependent on China and Korea" (pp. 3-4).

8 Miller (1967) dates the earliest contact of the Japanese with the

Chinese script and language back to the third century CE. However, Shibatani (1990) suggests that it was earlier, commenting that "it is generally believed that Chinese words were first introduced into Japan during the first century AD, or possibly even before that" (p. 145). It is not clear if he is referring to a spoken or written language, but Kaiser (1993a) also observes that Japanese people may have been exposed to Chinese writing in the form of inscriptions on seals, swords, etc., as early as the first century CE.

9 As Coulmas (1989) comments, "whenever illiterate peoples wanted to write, they usually borrowed the written language of a neighbouring peoples. Only gradually did bilingual individuals grasp the possibility of writing their own language, too" (p. 43).

10 There seem to be a number of reasons why this was so. The first was the status of 漢文 *kanbun* as the official language, which continued to be used by educated men throughout the Heian period (Habein, 1984). Another reason was that the use of katakana as a writing aid "led to the practice of writing content morphemes with Chinese characters, and grammatical morphemes and function words with kana" (Coulmas, 1989, p. 132). Seeley (1991) cites the changing nature of the relationship between kana and syllabic structure and changes to the vocabulary as reasons for this development.

11 However, as Keene (1986) observes, while newspapers generally follow Jōyō kanji usage guidelines, "no such restriction on the number of kanji apply to books, and if one wishes to read a novel by Mishima Yukio, for example, a knowledge of at least 3,000 characters is desirable" (p. 6).

12 Shibatani (1990) observes that the preface to the 古事記 *Kojiki*, the earliest extant text (712 CE) was written like this.

13 Singleton (2000) comments that, "a morpheme may be realized by different forms—its *morphemic alternants* or *allomorphs*—according to the particular environment in which it occurs" (p. 34).

14 Although Nelson (1962) lists this character in his dictionary under entry number 4469, he does not include the *tō'on* reading *ju* or the kun-reading *saki*, but he does cite one other kun-reading, *kabu*, which Coulmas (1989) does not mention.

15 The traditional classification, known as 六書 *Rikusho* 'Six Scripts,' was introduced by 許慎 *Kyoshin* in his dictionary, 説文解字 *Setsumon kaiji* compiled in China around 120 CE, and is still widely used in Japan today (Martin, 1972; Kaiho & Nomura, 1983; Henshall, 1988; Coulmas, 1989; Halpern, 1990; Habein & Mathias 1991; Kaiser, 1993a, Boltz, 1996).

The fifth and sixth groups in the traditional classification are actually principles of usage: 仮借文字 *kasha moji* phonetic loans which include characters that have been used for another meaning on the basis of a shared sound (e.g., 来 *rai* 'wheat' borrowed to write the word *rai* 'come'), and 転注文字 *tenchû moji* derivative characters which are kanji used to represent another meaning by extension of the original meaning (e.g., 令 coming to mean 'governor' from 'command' via 'commander').

16 Sometimes, another distinction is made for those kanji that have been created by the Japanese themselves, referred to as either 国字 *kokuji* 'national characters,' or 和製文字 *wasei moji* 'Japanese-made characters,' (Halpern, 1990). However, the majority of these kanji are based on the semantic compound principle and have no *on*-readings (e.g., 峠 'mountain pass' is a combination of 'mountain,' 'up' and 'down,' while 躾 'discipline, train' is a combination of 'body' and 'beautiful').

17 While recognizing that all classifications will inevitably be subjective to some degree, certainly classifications should avoid making simplistic evaluations of what constitutes an optimal orthography. As Henderson (1984) observes, the issue of an optimal level for a script is ultimately an empirical and psychological matter that turns on a number of questions, including how easily a child can learn the orthography, how easily it facilitates writing and reproduction by machines, how easy it is to read for the mature reader, and how effective it is in conveying information (for further discussion of this issue, see also Haas, 1976, 1983; Rogers, 1995).

18 DeFrancis (1989) actually cites the second edition of Taylor's work published in 1899.

19 Although Daniels (1996, 2000), a former student of Gelb (1952), uses the term *grammatology*, as Coulmas (1996) points out, the term has not gained wide currency. Perhaps a reflection of the fact that the study of writing has been a neglected area of linguistics (Daniels, 2001), it is, however, regrettable that even finding a name for the field is problematic. While Kao, van Galen, and Hoosain (1986) have proposed the term *graphonomics* to refer to handwriting research, Watt (1994) seems to be prepared to extend this to the study of writing systems, due to its lack of negative associations. It is unfortunate that the term *graphology* is associated with the practice of handwriting analysis, for this would be an ideal term for a science of writing, for, being analogous with *morphology* and *phonology*, it would clearly underline the relations between meaning, sound, and visual form.

20 While Diringer (1962) acknowledges that alphabetic writing is technically a subdivision of phonetic writing, he claims that "alphabetic writing has within the past three thousand years assumed such importance as to deserve a category of its own" (p. 24).

21 Sampson (1994) points out this division was intended to be more conjectural in nature, merely speculating on "whether there might ever be a semasiographic system comparable in expressive power to a spoken language" (pp. 119-120) rather than arguing for the existence of such a system (hence the dotted line in the figure).

22 The reader is referred to Sproat (2000) for a balanced stance on DeFrancis' (1989) arguments that no full writing system is semasiographic and his singling out of the Yukaghir 'love letter' cited by Sampson (1985).

23 The term abjad is formed from the first letters of the Arabic script, the most widespread example of this kind. The term abugida is an Ethiopic word formed from the first four consonants and first four vowels in a traditional ordering.

24 Daniel (2001) explains that this classification of six writing systems is a solution to inadequacies with the traditional tripartite classification (logographs, syllabaries and alphabets), for "once abugidas are distinguished from syllabaries, a different historical sequence can be identified, which no longer privileges the alphabet teleologically" (p. 68).

25 Not to take Hansen's (1993) remarks completely out of context we should note that his sense of despair stems from what he sees as the limited application of a term that is used "to refer to a language-type with exactly one-and-a-half instances" (p. 376) referring to China and Japan, respectively. Still, one cannot help but feel that this comment is completely missing the point of the debate.

26 Sandler and Lillo-Martin (2001) and Steinberg, Nagata and Aline (2001) both provide recent, readable discussions of sign languages, acknowledging that sign languages are now recognized to be full languages in their expressive ability *to convey any and all thoughts*, and which can be acquired naturally. The reader is also referred to Steinberg, et al.'s interesting discussion of the written language bilingual approach for the deaf, where the authors comment that "writing differs from speech concerning the physical medium of information transmission: writing involves light, which speech involves sound" (p. 78).

Moreover, the fascinating development of Nicaraguan Sign Language (Kegl, Senghas & Coppola, 1999; Senghas, 1995) which evolved out of a "jumble of idiosyncratic homesigns/gestual systems" (Kegl, et al. p. 180) clearly refutes DeFrancis' (1989) assertion "that spoken language is indispensable to sign language must be acknowledged as a basic fact" (p. 19).

27 Although DeFrancis (1989) comments that "simple characters of pictographic origin … comprise only about one percent of the total number of Chinese characters. The remaining 99 percent … are compound characters whose main component is a phonetic element" (p. 100), the number 100% is used in the title of a figure (pp. 102-103).

As Sampson (1994) comments, "this claim confuses diachrony with synchrony. It may be correct that the creation of a script always involves phonetic considerations, but subsequent evolution of script and spoken language can remove the phonetic basis of a writing system. It is difficult to agree that modern Chinese writing is essentially phonetically based; and it is certain that phonetic motivation is not a necessary feature for a script" (p. 117).

28 Underscoring the mixed nature of all writing systems with some degree of both pleremic and cenemic writing, Sakai (1991) remarks that a "purely phonetic writing system is no doubt an irresponsible political fancy not only because no such writing system actually exists but also because such a system simply would not work. We must caution ourselves against the commonsensical notion that writing systems like the Japanese are abnormal; as a matter of fact, the Japanese system is perhaps a very accurate representation of the nature of writing itself" (p. 266).

29 See DeFrancis (1989), Sampson (1985) and Unger (1987, 1990), in particular, for discussion of the origins of this term and its history of misusage. Coulmas (1984) argues, "the misconception of "ideographic writing" as applied to Chinese characters, may have been induced, partly, by the fact that in some respects Chinese characters behave like a written language rather than a system for representing language" (p. 65).

30 In his glossary for the terminology of writing, Gelb (1952) makes the following entry. "*Logography* or *Word Writing.* A writing in which a sign normally stands for one or more words of the language" (p. 250). However, later writers have given more inclusive definitions. For example, Taylor and Taylor (1983) remark that "a writing system in which one grapheme represents primarily the meaning (and sometimes secondarily the sound) of one word or morpheme may be called a logography" (pp. 20-21). In their list of terminology of writing, Daniels and Bright (1996) provide the definition of logogram as "a character that denotes the meaning but not the pronunciation of a morpheme" (p. xlii).

We may also note that although many of the classifications reviewed have used the term logographic (Daniels, 1996, 2001; Halliday, 1985;

see also entry in Coulmas, 1996), they have also commented that the term morphographic is *more precise*, but if this fact is so widely recognized, then why do so many continue the practice of using the term logographic?

Chapter 3 Japanese word formation

31 Tamamura (1988) distinguishes between 語構成 *gokōsei* 'word structure' and 造語法 *zōgohō* 'word formation processes,' noting that the former is a more static, analytic concept whereas the latter is more dynamic in nature. He also observes that although the term word formation is more common in the English literature actually much of the discussion is concerned with word structure. This thesis is also guilty of this tendency, for although I have in the main used the term word formation, emphasis is primarily on word structure. Accordingly, little will be said about processes, such as abbreviation, which is very productive in Japanese word formation (Kaiser, 1993b).

32 Note that Tamamura's (1985) classification makes no provision for inflectional morphology. Under simple words, in addition to monomorphemic words like 手 *te* 'hand,' he also includes verb and adjective forms, such as 見る *miru* 'see' and 高い *takai* 'high,' which consist of a stem represented by a kanji and an inflectional ending represented by a hiragana character, and thus are polymorphemic. This oversight may simply reflect the fact that Tamamura's (1985) main concern is with compound words, rather than an intention to treat verbs and adjectives as monomorphemic items. As our concern is also with compound words, no attempt is made at correcting this situation by incorporating inflectional morphology within the classification.

33 This characterization is similar to Morioka's (1968) treatment, noted in Chapter 2, of on-readings and kun-readings of a kanji as being allomorphs, based on the shared semantic association to the morpheme represented by a given kanji.

34 Nomura (1988a) comments that although word bases can be divided into 自立語基 *jiritsu goki* free word bases, 発生語基 *hassei goki* derivative word bases, and 結合専用語基 *ketsugô senyô goki* bound word bases, because we are only concerned with compound words, this level of classification is not expanded upon.

35 Although the names for the principles used here are probably more consistent with Tamamura (1985) than with Nomura (1988a), these terms are intended to emphasize the syntactic and semantic relationships between the elements, especially, for the first five principles that are used in some of the experiments that will be reported in this thesis. As these experiments are concerned with the lexical retrieval and representation of two-kanji compound words in the Japanese mental lexicon, these contrasting patterns, including the reversed syntactic patterns of verb + complement and complement + verb, seemed to have interesting implications.

36 While we do not deny the phonological basis of 'phonetic borrowing' compound words, the idiosyncrasies in kanji usage for such compounds hardly constitutes a 'purely' cenemic writing system, for many jukujikun words are not completely without semantic reference (see Sproat (2000) for discussion of similar items in Chinese).

Chapter 4 Visual word recognition

37 Henderson (1982) also writes that, "often it happens that psychologists, impressed by some piece of linguistic analysis, have imported it without regard to its psychological plausibility. It is arguable that the notion of reading by grapheme-phoneme correspondences became incorporated into the psychological nest in this way" (p. 4). Given that grapheme-phoneme correspondence (GPC) rules are an integral part of the phonological route that dual-route models postulate (Coltheart, 1978; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001), presumably we may also question the 'psychological plausibility' of such models.

Although made a decade earlier, we may also pause to wonder about the relevance of Henderson's (1982) comments for the Universal Phonological Principle (UPP) proposed by Perfetti, Zhang, and Berent

(1992), which would appear to be the psychological counterpart of the 'language is speech' view (Perfetti, 1999), which we discussed in Chapter 2. As its central premise, the UPP principle states that across all writing systems encounters with most printed words (with the exceptions of a restricted list of sign-like words) automatically lead to phonological activation, beginning with phoneme constituents of the word and including the word's pronunciation. More generally, Perfetti, et al. (1992) claim that phonological processes are pervasive in reading, with respect to various reading processes (from comprehension to word identification), with respect to writing systems (from Chinese to English to Serbo-Croatian), and with respect to individuals (from children to hearing and deaf adults of high reading skill), and that the pervasiveness of phonological processes is less a function of writing systems, although they are important, than of the human capacity for language. All of this is, of course, compatible with the abstract entity view of language being suggested in this thesis, which does not deny the primacy of speech as a medium of expression for hearing persons; it only argues that functionally speech is not in a hierarchical relation to writing, or even sign. When printed words are encountered, the input is orthographic; this is linked through its functional role as a medium of expression to language and through the inter-translatability relations that exist between mediums of expression to speech. Based on this three-way relationship, we also assume that encounters with printed words, in addition to automatically leading to semantic activation, will automatically lead to phonological activation.

38 While we would replace the term logographic with morphographic, this in no way detracts from the basic point of Coulmas' (1984) astute observation that the "logographic status of a writing system does not depend on what its users do when they read. Whether the meaning of a morpheme represented by a character is grasped by mediation of the phonological form to which the character is related, or whether, conversely, the phonological form is arrived at via the meaning associated with the character, in no way changes the logographic status of the characters" (pp. 63-64).

39 The debate over the relative contributions to lexical retrieval of direct semantic activation based on orthographic information or phonological activation is perhaps most clearly contrast in the controversies regarding dual-route and single-route models (Balota, 1994; Coltheart, 1978; Coltheart, Daveleer, Jonasson, & Besner, 1977; Coltheart, Rastle, Perry, Langdon and Ziegler, 2001; Seidenberg, 1989, 1990, 1995; Seidenberg & McClelland, 1989).

40 While one may be tempted to discount the early remarks of Coltheart, Daveleer, Jonasson, and Besner (1977) commenting that they "think it is premature to claim that there is any evidence at all which demands acceptance of the view that proceeding from a printed word to the lexical entry for that word ever uses a phonological conversion of the printed characters," (p. 551) as being made before all the evidence is in, we should not so lightly dismiss their inclination "to think of the relationship between phonological and visual inputs to the lexicon as one of cooperation, rather than competition" (p. 550).

Touching on the issue of relative time courses for semantic and phonological activation which is a central aspect of this debate, Shen and Forster (1999) point out that "we agree that a reader is sensitive to the phonological properties of written words at some stage of the reading process, but this stage is probably so late that phonology plays little or no role in the identification of the word, but probably does play a role in postaccess comprehension" (p. 455).

In his discussion of lexical access, Seidenberg (1990) argues for a single mechanism, such as that underlying the Seidenberg and McClelland (1989) model, which we shall discuss below.

41 In particular, Krott, et al. (2001) question whether connectionist models, where the basic regulating mechanism is frequency of occurrence, can account for both the presence of a morphological family size, which is a type-frequency effect, and the absence of a cumulative morphological family effect, which is a token-frequency.

42 Balota (1994) also identifies the use of response latencies, the inclusion of top-down contextual influences, and the fact that the effect is robust and easy to replicate as other reasons why the paradigm matched the prevailing 'Zietgeist' when Meyer and Schvaneveldt's

(1971) study appeared (p. 337).

43 The basic constituent-morphemic priming paradigm may be seen as a version of what is sometimes referred to as (partial) repetition priming, which has also been used in studies of other polymorphemic words, such as derivatives. Napps (1989), for instance, used conditions which paired as prime and target words from sets of related words such as; *manage–manage*, *manages–manage*, and *manager–manage*; also *develop–develop*, *developed–develop*, and *development–develop*; as well as *enjoy–enjoy*, *enjoying–enjoy*, and *enjoyment–enjoy*. In that study, priming was obtained for regularly and irregularly suffixed morphemic relatives, leading Napps (1989) to suggest that morphemic relationships are represented explicitly in the lexicon. A study by Feldman (1992) on morphologically related English words has also used this technique, and claims that inflections and derivations produce a similar pattern of facilitation. Stolz and Feldman (1995) have also considered inflectional and derivational morphology in terms of semantic transparency. Investigating inflectional and derivational morphology in Italian and Dutch using an SOA technique to present partial primes, Jarvella, Job, Sandstrom, & Schreuder (1987) argue for existence of morphological entries in the mental lexicon.

44 Sandra (1990) discussed these results primarily in the context of Taft and Forster's (1975, 1976) notion of automatic morphological decomposition, which we shall discuss in Section 4.3.2.1, interpreting the absence of priming in the opaque compounds as evidence against this notion. Sandra (1990) suggested a modification to the model in which only semantically transparent compounds would lack whole-word representations and the morphemes of such words would be accessed once the system failed to match with a lexical representation.

45 Based on these results, Zwisterlood (1994) has proposed a lexical structure for Dutch compound words involving both morphological and semantic level representations. Claiming that all compounds are represented as morphologically complex, regardless of their semantic make-up, Zwisterlood (1994) essentially distinguishes between the different types of compound words in terms of inhibitory or facilitatory links. In this way, truly opaque compounds, as well as pseudo-compounds, are characterized as having semantic representations that are not linked to semantic representations of the component elements. In contrast, however, fully or partially transparent compounds are treated as having semantic representations that are linked with the representations of constituents.

46 Inevitably, the review must be highly selective. As we are concerned primarily with visual word recognition, mental lexicon models dealing with speech, such as the Cohort Model (Marslen-Wilson, 1987; Marslen-Wilson and Welsh, 1978) are not covered here (see also Levelt (1989) for discussion from speaker's perspective and Aitchison (1994) for English mental lexicon in terms of speech).

47 Dual-route models (Coltheart 1978; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) are not discussed here for two reasons. The first reason is that, as suggested earlier, it is not clear whether a strong distinction between phonological routes and direct visual routes is psychologically well motivated. The second reason is that the introduction of activation mechanism in a recent version of the Dual-Route Cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) blurs the distinction between this and other activation models. Coltheart, Rastle, Perry, Langdon, and Ziegler (2001) describe the DRC model as consisting of three routes, a lexical semantic route, a lexical non-semantic route and a GPC route, with each of these routes composed of a number of interacting layers.

Although the Logogen Model (Morton, 1970, 1979) has been extremely influential as a model of the mental lexicon, it is not discussed further because it does not contribute directly to our discussion of polymorphemic words. McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) acknowledge the strong similarity between their model and the logogen model, even suggesting that, "what we have implemented might be called a hierarchical, nonlinear, logogen model with feedback between levels and inhibitory interactions among logogens at the same level" (McClelland & Rumelhart, 1981; p. 388).

48 Sproat (2000) has criticized the Seidenberg and McClelland (1989) model, remarking that, "Seidenberg and McClelland have failed to provide convincing evidence that their model has learned the task that it is claimed to have learned; thus there is little reason to accept their conclusion that more traditional kinds of models have been superseded" (p. 179). This criticism would, however, seem to miss the fact that the model was, as Seidenberg (1989) points out, to a large extent, "a case study in connectionist modeling" providing a "basis for a exploration of the utility of this approach" (p. 25) rather than being the last word on the subject (p. 68).

49 Seidenberg and McClelland (1989) comment, "the notion of lexical access carries with it a concern with certain types of theoretical questions. The primary questions concern the number of lexicons, how they are organized and linked, and whether it is orthographic or phonological information that provides access to meaning. The primary processing mechanism is search through one or more ordered lists. In our model, the codes are distributed, they are computed on the basis of three orthogonal processes, and the primary processing mechanism is spread of activation. The primary theoretical questions concern the properties of these computations, which are determined by the properties of the writing system that are picked up by the learning algorithm on the basis of experience" (p. 560).

50 Jarema, Kehayia and Libben (1999) is a special issue of Brain and Language on the mental lexicon, presenting papers from the First International Conference on the Mental Lexicon held at the University of Alberta, Edmonton, Canada in September, 1999. The special issue consists of 52 articles on various aspects of the mental lexicon, but a clear indication of the importance of morphological issues is the fact that 19 (36%) of these explicitly include reference to morphemes or morphology in their titles. Jarema, et al. also mention the wide range of languages investigated in these papers, which includes Bulgarian, Chinese, Croatian, Dutch, English, Finnish, French, German, Greek, Italian, Japanese, Polish, Spanish and Turkish. The language-universal importance of morphology is also evident in the models that we review below, for the augmented addressed morphology model was proposed primarily to account for inflectional morphology in Italian, the multilevel interactive-activation framework mainly for derivational morphology in English, and the parallel dual-route model of morphological processing was for derivational morphology in Dutch.

51 As mentioned in Section 4.2.4, the papers by Baayen and Schreuder and their colleagues on morphological families (Baayen, Lieber, & Schreuder, 1997; Bertram, Baayen, & Schreuder, 2000; De Jong, Schreuder, & Baayen, 2000; Krott, Baayen, & Schreuder, 2001) as well as Schreuder, Neijt, van der Weide, & Baayen (1998) on plural forms have all been discussed in terms of this parallel dual-route model of morphological processing.

Chapter 5 The Japanese mental lexicon

52 The suggestion that the "early dichotomies" to which Kess and Miyamoto (1999) refer are in large measure the result of misleading notions concerning the classifications of writing systems should not surprise the reader greatly. The influence of such misconceptions can be discerned in some of the early studies on lateralization, aphasia and developmental dyslexia, and Stroop-effect experiments which Flores d'Arcais (1992) reviewed. From his review, however, Flores d'Arcais concludes that, "although there is some evidence from Stroop-type experiments and some, less clear, from clinical data and from lateralization experiments, indicating some differences in the processing of Chinese characters as compared to alphabetically printed words, the conclusion that Chinese characters are processed more "like pictures" than like words can hardly be maintained" (p. 50). See also Hatta (1992) for study of effects of kanji attributes on visual field differences.

53 Certainly, it seems highly unlikely that kanji orthography is more frequent for all six of the color words used in this study, which were 黒 *kuro* 'black,' 緑 *midori* 'green,' 茶色 *chairo* 'brown,' 灰色 *haiiro* 'gray' , 朱色 *shuiro* 'vermilion,' and 栗色 *kuruiro* 'chestnut.'

54 Although the distinction between word comprehension and word naming is extremely important when comparing pleremic writing systems and cenemic writing systems, the point often seems to be overlooked. This distinction would seem to be related to the importance of meaningfulness that Steinberg and Yamada (1978-9) have emphasized (see also Yamada, 1997). They report that three- and four-year-old Japanese children learned some kanji easier than kana, despite greater graphic complexity, because they were being associated with meaningful words rather than abstract sounds.

55 This is consistent with the results of Perfetti and Tan (1998) which over different SOA identified a sequence of facilitation from (a) graphic, (b) phonological, and (c) semantic.

56 These findings are similar to those of Van Orden (1987), mentioned in Chapter 4 and which Wydell, Patterson, and Humphreys (1993) discuss. As noted earlier, we must be careful not to misinterpret these results as strong evidence for phonological mediation, for as the results from Sakuma, Sasanuma, Tatsumi, and Masaki (1998) and also Saito, Masuda, and Kawakami (1998) suggest, lexical retrieval involves the activation of both phonological and orthographical information.

57 As it is concerned primarily with spoken word recognition, we shall not include here a review of Zhou and Marslen-Wilson's (1994, 1995) multi-level cluster representation model for Chinese disyllabic compound words which includes a syllable level, a morpheme level and a word level, making it similar to some of the other multilevel models we shall look at. More recently, however, Zhou and Marslen-Wilson (2000) have come to favor a different kind of model, where "morphemic and whole-word representations are semantic in nature, and are arranged at the same level of the system, with phonological representations (i.e., syllables) connecting to both of them" (p. 60).

58 As we shall see, although the metaphors may differ, there are clear points of similarity between Hirose's (1992, 1994, 1996) hypotheses and Forster's (1976, 1989) serial search model, for not only do Hirose's 'clusters' resemble Foster's 'bins,' but the notion of 'retrieval cue' is similar to that of 'access code.' The metaphor difference may be because Hirose (1992) presents his study as being concerned with the structure of semantic memory, drawing on concepts such as retrieval cues and activation diffusion in semantic activation (particularly Exp. 3), rather than explicitly as a model of visual word recognition.

59 This finding and the priming found in the present research make the absence of facilitation reported by Morton, Sasanuma, Patterson, and Sakuma (1992) somewhat difficult to explain. Morton, et al. only observed facilitation with the prior exposure of the identical word, but not with a different word, either a single kanji word or compound word sharing a character with the target word.

60 Although structurally similar to Seidenberg and McClelland's (1989) model, Ijuin, et al. (1999) note that while Seidenberg and McClelland modulated weightings using probabilities proportional to written frequency, they have employed the procedure used by Plaut, McClelland, Seidenberg, & Patterson (1996) who used frequency values to scale changes to weightings. This procedure means that during training there are greater changes in connection weightings associated with more frequent words.

61 The manipulations of character frequency and word frequency utilized in this experiment are similar to those used by Taft (1979), which we mentioned in Section 4.2.4. The techniques have also used in Taft, Huang and Zhu (1994), and in Japanese by Tamaoka and Hatsuzuka (1995), which we shall discuss later, as well as in van Jaarsveld and Rattink (1988) for Dutch compound words.

62 Examples of the co-ordinative words, that can clearly be associated with Japanese words, include 更改 *kôkai* 'renewal,' 監察 *kansatsu* 'inspection,' and 湖沼 *koshô* 'lakes and marshes.' Examples of the modifier words are 荒野 *kôya* 'wilderness,' and 牧童 *bokudô* 'shepherd boy.'

63 This claim is, of course, similar to the claims made by Taft and Forster (1975, 1976) in the prefix-stripping model discussed in Section 4.3.2.1.

64 Stimulus kanji characters were presented in pairs for 40 ms, then after a mask, and a cue indicating which radical to report first for the target character in a trial, participants were asked to write the kanji according to the indicated order.

65 In Japanese, of course, many of these words are often transcribed in katakana, such as 蚯蚓 *mimizu* 'earthworm,' and 珊瑚 *sango* 'coral,' although uses of these words, for instance, 珈琲 *kôhii* 'coffee' and even 麒麟 *kirin* 'giraffi,' are not unknown in advertising.

66 As an example of a bound morpheme which only occurs in the first position, they cite 殉 in Chinese words corresponding to the Japanese words 殉国 *junkoku* 'die for one's country,' and 殉職 *junshoku* 'die at one's post.'

67 In his discussion of the mental lexicon from the production perspective of speaking, Levelt (1989) uses the term lemma to refer to the abstract part of a lexical entry consisting of meaning and syntax, which is linked to a morpho-phonological form; the other part of the entry containing morphological and phonological information. Levelt (1989) comments that this "partitioning of the mental lexicon in two kinds of store is no more than a spatial metaphor acknowledging the existence of two kinds of internal organization in the mental lexicon: one according to the meaning of items and one according to their form properties" (pp. 187-188). This distinction is similar to that proposed by Kempen and Huijbers (1983), who used the terms L1-item to refer to "abstract pre-phonological (but syntactically specified) items" and L2-items of "concrete phonological shapes for abstract items." (p. 208).

68 The selection of these five principles was influenced by a number of factors. While the other principles of affixation, repetition, abbreviation, and phonetic borrowing would undoubtedly also provide data of interest, they are, for varying reasons, less suitable for the present experimental design.

As noted at the end of Chapter 3, Nomura (1988) includes two patterns of modification in his classification of compound words. The modification (1) pattern of verbal modification appears well motivated in contrast to the modification (2) pattern of nominal modification. The division of verbal patterns into those of modification and those of complements also seems justified as a natural consequence of the move. This grouping undeniably provides for a parsimonious linguistic account. However, it must be noted that this is at the expense of differentiating between the two syntactic patterns of verb + complement and complement + verb, for both are included under the complement pattern. Given the present concern with the organization of compound words in the mental lexicon, another factor in the selection of these five principles was the desire to include contrasting patterns, such as the reversed syntactic patterns of verb + complement and complement + verb, and thus, in this respect, these principles break somewhat with Nomura's (1988) classification.

69 Saito (1996) has commented that Hirose's (1992) hypothesis of serial processing from left to right is plausible given that compound words are written from left to right. However, although production order (in writing and speaking) may be a factor, it is unlikely to be the only factor in determining the organization of the mental lexicon, which, as we are arguing, must reflect the morphological relations between words.

70 Although Hirose (1994, 1996a, 1996b) has subsequently modified his hypothesis slightly by distinguishing between independent kanji and combination or bound kanji in terms of their effectiveness as retrieval cues, still the role of retrieval cue for compound words remains firmly with the first kanji.

71 This corpus of 1,000 two-kanji compound words was also surveyed for familiarity. As noted in Chapter 4, word frequency and word familiarity effects are among the most robust findings in experimental psychology which any experiment of visual word recognition must control for, but, unfortunately, at the time the constituent-morpheme priming experiments were being prepared, suitable word frequency counts were not readily available. As Wydell (1991) points out, although the National Language Research Institute has conducted two major word frequency surveys, one of magazines (1962-1964) and one of newspapers (1976), both of these were conducted before major script reforms, and in particular before the Jōyō kanji list was promulgated in 1981, and thus may be considered as being "out of date" (p. 100).

The corpus was randomized and divided into separate questionnaire forms, similar to those for the classification survey, which 63 native Japanese speakers (average age = 27.65, range 20-56, *SD* 7.0) rated (each respondent rating between 1 to 3 lists, with average number completed being 1.58, *SD* 0.58). Consistent with the fact that the majority of these words came from a survey of basic vocabulary for Japanese language education (National Language Research Institute, 1984), generally, there was a bias towards high-familiarity words, although the mean familiarity score for the associative pair compounds was lower, which is probably because more of these items were selected from an elementary school kanji dictionary (Ishii, 1996), which in itself might not be a reliable indication of frequency of use in daily life. Subsequent to completing this survey, Nozaki, Shimizu, and Yokoyama (1999) presented the results of a frequency count for two-kanji compound words based on the Asahi Newspaper corpus for 1993, from which they extracted 7,395,051 tokens for 34,934 types. The evaluations of familiarity obtained in this survey were found to be correlated to those newspaper frequency counts, r = 0.191, which is significant at the 0.01 level.

72 An interesting question for the lexical decision task using kanji is how can participants make the decision that a pair of kanji is not a word. In our brief discussion of findings from the lexical decision task in Chapter 4, we noted that responses to legal non-words in English, that is, words that conform to phonological rules, take longer to reject than illegal or unpronounceable non-words. However, phonological constraints are unlikely to be the only factor with non-words of kanji pairs. Given the high incidence of homophones in Japanese arising from many kanji having the same on-reading, some random combinations of kanji in forming non-words will inevitably result in pairs of kanji, the on-reading of which gives a pronunciation that is shared by a real word. For instance, one non-word kanji pair generated by randomizing the corpus of kanji was 制旨 *sei* 'regulation' + *shi* 'purport, gist,' but the pronunciation *seishi* is also shared by a number of real words, such as 生死 'life and death,' 制止 'control; restrain,' 静止 'stillness; repose,' and 製紙 'paper making.' The decision to reject a non-word must, therefore, be based on orthographic or semantic information, but studies by Saito and Kawakami (1992a, 1992b) and by Kawakami, Saito, and Yanase (2000) suggest that some non-word combinations of two kanji may be more word-like than other non-words.

Accordingly, a survey was conducted to control for the word-like-ness of the non-word items. Questionnaire forms for 500 generated non-words were prepared, similar to those of the other surveys, which 40 native Japanese speakers (average age = 25.9, range 22-43, *SD* 4.5) rated on a 7-point scale, with 1 for un-word-like and 7 for word-like.

73 Although it was possible with these procedures to ensure that no kanji appeared more than once in the compound words and more than once in the non-word pairs, unfortunately, because the non-words were generated from the corpus of compounds, 23 kanji did appear in both compounds and non-word pairs. In constructing the presentation lists, care was taken in assigning the non-word pairs to blocks, to ensure that no kanji appeared more than once as a prime in any list. For example, although the kanji 昼 *hiru* 'daytime' appeared in the compound word 昼寝 *hirune* 'nap' and in the non-word pair 昼書, it served as a prime to 昼寝 in one list, and as a prime to 昼書 in another list.

74 In experiments similar to the study by Zhang and Peng (1992) which focused on word frequency and the character frequencies of the constituent kanji in two-kanji compound words, Tamaoka and Hatsuzuka (1995) maintained a low word frequency over four conditions of high-high, high-low, low-high and low-low kanji character frequencies. Comparing the results from lexical decision and naming, they found that the pattern over the four conditions was different.

In the naming task, no difference was found due to frequency manipulations of the second element, but naming latency was faster when the first element was high frequency compared to when it was of low frequency. However, in the lexical decision task, the pattern was reversed, with no effect for the frequency of the first element, but faster reaction times when the frequency of the second element was high than when it was low. Tamaoka and Hatsuzuka (1995) explain this pattern of results in terms of serial, left-to-right, processing of the compound words. Their argument is that because naming can be initiated by phonological activation of the first kanji element, latencies will be faster when this is high frequency, but because the lexical decision task requires activation also of the second kanji element, reaction times will be faster when the second element is of high frequency. Note that this explanation is not denying an effect of frequency for the second element in the naming task or for the first element in the lexical decision task, but claims rather that their effects are hidden by the other kanji elements, which are more important in determining the reaction times for the respective tasks.

75 McRae & Boisvert's study of automatic semantic similarity priming was motivated by the claim made by Shelton and Martin (1992) that semantically related word pairs which are not associatively related do not automatically prime one another. The serious implication of the claim that priming is non-semantic in nature is that automatic semantic priming provides us with no information about how semantic knowledge is organized (Shelton & Martin, 1992, p. 1207). Postulating that the failure to find automatic priming for the semantically related words in Shelton and Martin (1992) was actually due to the low similarity of the stimulus items, McRae and Boisvert (1998) attempted in their experiment to show automatic priming for carefully selected word pairs which, although highly similar semantically, would not be linked in association norms. In this experiment, immediately after a fixation point for 250 ms, a prime was presented for 200 ms followed by a mask (&&&&&&& string) for 50 ms. The target was presented immediately after the mask, and displayed until the participants responded by pressing a key for the lexical decision task. The result of this experiment was that priming was found for highly similar prime-target pairs. Thus, McRae and Boisvert's (1998) findings dispel doubts about the usefulness of the automatic semantic priming paradigm as a tool for investigating the nature of the mental lexicon by demonstrating that automatic priming can be obtained for suitable stimulus items at an SOA of 250 ms.

76 Foster and Davies (1984) and Forster (1998) have argued that the masked priming paradigm minimizes the influence of any possible episodic trace in the lexical decision process. Masking is achieved by the combination of a very short SOA, typically around 60ms, and the presence of a masking pattern. Although, they have reported repetition priming using this procedure, is not clear whether semantic priming effects would be obtained using this masking procedure. Perfetti and Zhang (1995) working with Chinese characters have reported finding phonological interference in a semantic judgment task at 90ms, and semantic interference in a phonological judgment at 140ms. This result suggests that activation of semantic information from a prime might not be complete at intervals of less than 140ms. The masked priming paradigm was not adopted for this experiment into morpheme priming to compound word targets.

77 Separate subject and item analyses were also conducted for the 18 Experiment 1 stimulus items with the reaction time data obtained in Experiment 2. However, as the results of these were the same as the results for all 30 Experiment 2 stimulus items, only the more reliable analyses with the larger set of stimulus items are reported here.

78 While the compound words were matched closely for familiarity, there were inevitably slight differences in the mean familiarity evaluations across the word-formation conditions, and slight shifts in these over the two experiments, as shown in Table 1. The pattern of these familiarity differences closely match the pattern of differences in reaction times across the word-formation conditions, suggesting that this was the cause of the apparent effect of word-formation principle.

79 Tamaoka and Hatsuzuka (1998) found that opposite-concept-kanji compounds were processed more slowly than similar-concept-kanji compounds in a lexical decision task but not in a naming task, with both groups being slower than the control group in both tasks. However, inspection of their stimulus lists suggests that many of the control-group compound words are actually similar-concept-kanji compounds.

Chapter 6 The lemma unit model

80 As Joyce and Ohta (1999) point out although there are many highly familiar examples of this principle, overall, associative pairs were poorly represented in the National Language Research Institute's (1984) list of basic Japanese vocabulary, and in preparing the corpus of 1,000 two-kanji compound words, the set was supplemented with a number of compound words from a kanji dictionary for elementary school students (Ishii, 1996), but as the results of the familiarity survey suggested, the mean familiarity score was lower for the associative pair compounds.

81 Kello and Kawamoto (1998) developed and introduced Runword, a collection and acoustic analysis software package, which could overcome the voice key problem. Unfortunately, the program written for DOS and the original SoundBlaster 16 sound card does not run on newer computers, and is no longer being supported or distributed (C. T. Kello, personal communication, March 23, 2001).

82 These results offer a possible clue as to why Hirose (1992) obtained an advantage for the first constituent. In his Experiment 1, the participants responded to the prime kanji as well by naming it. The bias towards the activation of phonological information, which we might expect this to have evoked, may have been the cause of the first constituent advantage Hirose found, when the participants were processing the target compound words.

83 Shafiullah and Monsell (1999) claim that the cost of switching between kanji and kana was robust, although it was not numerically large, with a mean reaction time difference of 13.1 +/- 1.8 ms between switch and non-switch trials over all of their experiments (p. 595).

84 The numbers of kanji sharing on-readings and kun-readings mentioned here are the numbers of kanji and words provided for selection by the input conversion program IME 2000 (Microsoft, Corp.).

85 Although this pattern of facilitation matches that of Hirose's (1992) Experiment 1, we must remember that it is extremely difficult to explain the facilitation in the second constituent condition compared to the baseline unprimed condition with his notions of retrieval cues and clustered representations based on a shared first kanji.

86 For cross-modality priming studies, see Marslen-Wilson, Tyler, Waksler, and Older, (1994), for English derivative words, as well as Feldman, and Soltano (1999) and Feldman and Laralee (2001) for studies involving affixation, also Frost, Deutsch, Gilboa, Tannenbaum, and Marslen-Wilson, (2000) involving Hebrew. Marslen-Wilson, et al.

(1994) found evidence for the morphological decomposition of semantically transparent forms.

Zhou and Marlsen-Wilson (2000) also present results of cross-modality studies for English and Chinese compound words.

87 As there are no single sound files in the NTT database for the on-readings *bo*, *da*, *getsu*, *gyo*, *hoku*, *min*, and *zetsu*, these readings were edited from a number of longer sound files. Ten native speakers were asked to listen to the edited version and judge the most appropriate for the target on-reading.

88 The treatment of proper nouns is a difficult issue, particularly in the case of kanji. As shown in Table 6.9, while proper nouns account for 49% of all the two-kanji compound word types, they only represent 13% of the tokens. This is because, apart from a relatively small number of frequent proper nouns such as 東京 (average token count, 16,294) and 加藤 (average token count, 1,896), most proper nouns have very low token counts. Proper nouns also have special distribution characteristics in terms of component kanji. Although many place names and personal names consist of basic kanji, such as 山, 川, and 木, that are also constituents of many other words, there are many proper nouns that consist of kanji which are rarely, if at all, used in other words. For instance, 藤 is a component of 232 types (average token count, 10,506), of which all but 10 (average token count, 127) are proper nouns. Because of these considerations and the fact that proper nouns are not normally used in word recognition research, the constituent morpheme frequency data reported here is based on counts excluding proper nouns.

89 The data has not been corrected for the presence of jukujikun, monomorphemic words, represented orthographically with two-kanji, such as 葡萄, where the constituent kanji cannot be analyzed as separate morphemes. These are retained not because we wish to suggest that a morphemic analysis is possible, but simply because we know of no exhaustive list of monomorphemic words that would allow us to confidently eliminate all of these from our counts.

90 Of the 2,965 JIS level 1 kanji, 222 do not appear as constituents of

two-kanji compound words in six-year newspaper corpus used in this study.

Chapter 7 Conclusions

91 While we should be aware of the danger of reading too much into this kind of spatial metaphor, still, we must also acknowledge the power of such metaphors in influencing our understanding of the underlying issues. The similarities in terms of their linear hierarchical arrangements between the 'language is speech' view, also represented in Figure 2.5, and dual-route models of lexical access (see Coltheart, 1978; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) are also hardly matters of coincidence.

92 Page (2000) argues that, "although it is unlikely that anyone would deny that processing in the brain is carried out by many different processors in parallel (i.e., at the same time) and that such processing is necessarily distributed (i.e., in space), the logic that leads from a consequent commitment to the idea of *distributed processing*, to an equally strong commitment to the related, but distinct, notion of *distributed representation*, is more debatable" and claims that "the *thoroughgoing* use of distributed representation, and the learning algorithms associated with them, is very far from being mandated by a general commitment to parallel distributed processing" (p. 443). In addition to Rumelhart, McClelland, and the PDP Research Group (1986), see also Feldman and Ballard (1982) for arguments for parallel processing assumed in connectionist models.

Chapter 8 Future research

93 Although Kawakami (1997, 2000) has presented data relating to two-kanji compound words in order to control for 'orthographic neighborhoods,' which were discussed in Chapter 4, in line with the central claim that kanji are most accurately morphographic in nature, Joyce and Ohta (in press) have suggested that the notion of 'morphological family' is far more appropriate for kanji words.

In applying Coltheart, et al.'s (1977) simple definition of a neighbor—any word generated by changing just one letter of a given word while preserving letter positions—to two-kanji compound words, Kawakami (1997, 2000) would seem to be equating 'one letter' with 'one character.' This analogy is straightforward enough and, arguably, preserves a sense of the orthographic similarity that underlies neighborhoods in the degree of visual overlap or similarity that two-kanji compound word neighbors would possess by virtue of a shared character. However, the analogy completely overlooks the fact that orthographically letters and characters function at entirely different levels. This functional difference between alphabetic letters and kanji has important implications for models of the mental lexicon, particularly when considering the loci of word recognition effects. While Coltheart, et al.'s (1977) definition of neighbors does not explicitly limit itself to cenemic writing systems, the confounding of orthographic similarity and semantic relatedness that arises if the notion of orthographic neighbors is extended to a pleremic system like kanji is clearly problematic.

94 It should be noted that the survey data is only for two-kanji compound words, whereas by Krott, et al.'s (2001) definition, a complete constituent family would also include the frequencies of a morpheme as a word stem (i.e., in verbs, such as 化 in 化ける) and as constituents of longer compound words (i.e., 化 in 近代化).

95 As already noted, the desire to provide some account of varying degrees of semantic transparency was one of the factors motivating Taft, Liu and Zhu (1999) to incorporate lemma units into the multilevel interactive-activation framework.

96 The first experiment in this study required both American and Japanese undergraduate participants to match thirty unfamiliar, technical terms with their definitions or descriptions, mostly from botany, zoology, medical science, psychology, or linguistics. Thus, the materials were Latin- or Greek-derived English technical terms and 2-5 kanji compound Japanese translations, such as *limnology* 湖沼学 meaning the scientific study of physical, chemical, and biological conditions in lakes and ponds, and *piscivorous* 魚食性 meaning eating fish as a regular diet.

Aitchison, J. (1994). *Words in the mind: An introduction to the mental lexicon* (2nd ed.). Oxford, England: Basil Blackwell.

Akita, Kiyomi, & Hatano, Giyoo, (1999). Learning to read and write in Japanese. In Margaret Harris & Giyoo Hatano (Eds.). *Learning to read and write: A cross-linguistics perspective* (pp. 214-234). Cambridge, England: Cambridge University Press.

Amano, S., Kondō, T., & Kakehi, K. (1995). Modality dependency of familiarity ratings of Japanese words. *Perception & Psychophysics, 57,* 598-603.

Amano, S., & Kondō, T. (Eds.). (1999). *Nihongo no goitokusei* [Lexical properties of Japanese] Vols. 1-6, NTT database series. Tokyo: Sanseidō.

Amano, S., & Kondō, T. (Eds.). (2000). *Nihongo no goitokusei* [Lexical properties of Japanese] Vol. 7 (Frequency), NTT database series. Tokyo: Sanseidō.

Anderson, S. R. (1992) *A-morphous morphology.* Cambridge, England: Cambridge University Press.

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language, 25,* 726-740.

Andrews, S. (1992) Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234-254.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language, 37,* 94-117.

Baayen, R. H., Lieber, R., & Schreuder, R. (1997). The morphological complexity of simplex nouns. *Linguistics, 35,* 861-877.

Baayen, H., & Schreuder, R. (1999). War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model. *Brain and Language, 68,* 27-32.

Backhouse, A. E. (1984). Aspects of the graphological structure of Japanese. *Visible Language, 18,* 219-228.

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.). *Handbook of psycholinguistic* (pp. 303-358). San Diego, CA: Academic Press.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340-357.

Baron, J. (1973). Phonetic stage not necessary for reading. *Quarterly Journal of Experimental Psychology, 25,* 241-246.

Bauer, L. (1983) *English word formation.* Cambridge, England: Cambridge University Press.

Becker C. A (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 556-566.

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language, 42,* 390-405.

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 489-511.

Besner, D., & Hildebrandt, N. (1987). Orthographic and phonological codes in the oral reading of Japanese kana. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 335-343.

Bloomfield, L. (1933). *Language.* New York: Henry Holt & Co.

Boltz, W. G. (1996). Early Chinese writing. In P. T. Daniels & W. Bright (Eds.). *The world's writing systems* (pp. 191-199). New York: Oxford University Press.

Butterworth, B. (1983). Lexical representation. In B. Butterworth, (Ed.), *Language production: Volume 2 Development, writing and other language processes* (pp. 257-294). London, England: Academic Press.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 297-332.

Chen, H.-C. (Ed.). (1997). *Cognitive processing of Chinese and related Asian languages.* Sha Tin, N.T., Hong Kong: Chinese University Press.

Chen, H.-C. & Tzeng, O. J. L. (Eds.). (1992). *Language processing in Chinese.* Amsterdam: North-Holland.

Chen, H.-C., & Zhou, X. (1999). Processing East Asian languages: An introduction. *Language and Cognitive Processes, 14,* 425-428.

Chialant, D. & Caramazza, A. (1995). Where is morphology and how is it processed? The case of written word recognition. In L. B. Feldman, (Ed.), *Morphological aspects of language processing* (pp. 55-76). Hove, England: Lawrence Erlbaum Associates.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies in information processing* (pp. 151-216). London: Academic Press.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.). *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204-256.

Coulmas, F. (1984). Arbitrariness and double articulation in writing. In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics* (pp. 57-66). London: Erlbaum.

Coulmas, F. (1989). *The writing systems of the world.* Oxford, England: Basil Blackwell.

Coulmas, F. (1996). The Blackwell encyclopedia of writing systems. Oxford, England: Blackwell.

Crystal, D. (1987). *The Cambridge encyclopedia of language.* Cambridge, England: Cambridge University Press.

Daniels, P. T. (1996). The study of writing systems. In P. T. Daniels & W. Bright (Eds.). *The world's writing systems* (pp. 3-17). New York, NY: Oxford University Press.

Daniels, P. T. (2001). Writing systems. In M. Aronoff & J. Rees-Miller (Eds.).

The handbook of linguistics (pp. 43-80). Oxford, England: Blackwell.

Daniels, P. T., & Bright, W. (Eds.) (1996). *The world's writing systems.* New York, NY: Cambridge University Press.

De Groot, A. M., Thomassen, A. J, & Hudson, P. T. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition, 10,* 358-370.

De Jong, N. H., Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes, 15,* 329-365.

DeFrancis, J. (1984). *The Chinese language: Fact and fantasy.* Honolulu: University of Hawaii Press.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems.* Honolulu, HI: University of Hawaii Press.

DeFrancis, J., & Unger, J. M. (1994). Rejoiner to Geoffrey Sampson, "Chinese script and the diversity of writing systems". *Linguistics, 32,* 549-554.

Diringer, D. (1962). Writing. London: Thames and Hudson.

Drews, E. (1996). Morphological priming. *Language and Cognitive Processes, 11,* 629-634.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205-254.

Feldman, L. B. (1992). Morphological relationships revealed through the repetition priming task. In Downing, P., Lima, S. D., Noonan, M. (Eds.). *The linguistics of literacy* (pp. 239-254). Amsterdam: John Benjamins Publishing Co.

Feldman, L. B. (Ed.). (1995). *Morphological aspects of language processing.*

Hillsdale, NJ: Lawrence Erlbaum Associates.

Feldman, L. B., & Larabee, J. (2001). Morphological facilitation following prefixed but not suffixed primes: Lexical architecture or modality-specific processes? *Journal of Experimental Psychology: Human Perception and Performance, 27,* 680-691.

Feldman, L. B., & Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language, 68,* 33-39.

Feldman, L. B., & Turvey, M. T. (1980). Words written in kana are named faster than the same words written in kanji. *Language and Speech, 23,* 141-147.

Flores d'Arcais, G. B. (1992). Graphemic, phonological and semantic activation processes during the recognition of Chinese characters. In H.-C. Chen, & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 37-66). Amsterdam: North-Holland.

Flores d'Arcais, G. B., & Saito, H. (1993). Lexical decomposition of complex kanji characters in Japanese readers. *Psychological Research, 55,* 52-63.

Flores d'Arcais, G. B., Saito, H., & Kawakami, M. (1995). Phonological and semantic activation in reading kanji characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 34-42.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales, & E. Walker (Eds.). *New approaches to language mechanisms: A collection of psycholinguistic studies* (pp. 257-287). Amsterdam: North-Holland.

Forster, K. I. (1989). Basic issues in lexical processing. In W. Marslen-Wilson, (Ed.), *Lexical representation and process* (pp. 75-107). Cambridge, MA: MIT Press.

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research, 27,* 203-233.

Forster, K. I., & Davies, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 680-698.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., & Marslen-Wilson, W. (2000). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory & Cognition, 28,* 1277-1288.

Fujii, T. (1996). Goi [Vocabulary]. In Saji, K. & Sanada, S (Eds.), *Nihongo kyōshi yōsei shiriizu: Onsei, goi, moji hyôki* (pp. 85-104). Tokyo: Tokyo Houmei Shuppan.

Fushimi, T, Ijuin, M, Patterson, K., & Tatsumi, I. F. (1999). Consistency, frequency, and lexicality effects in naming Japanese kanji. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 382-407.

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23,* 71-87.

Garman, M. (1990) *Psycholinguistics.* Cambridge, England: Cambridge University Press.

Gaur, A (1992). *A history of writing* (Rev. Ed.). London, British Library.

Gelb, I. J. (1952). *A study of writing: The foundations of grammatology.* Chicago: University of Chicago Press.

Gernsbacher, M. A., (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256-281.

Gottlieb, N. (1993) Written Japanese and the word processor. *Japan Forum, 5,* 115-132.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228-244.

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language, 30,* 370-384.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception and Psychophysics, 45,* 189-195.

Haas, W. (1976). Writing: The basic options. In W. Haas (Ed.), *Writing without letters* (pp. 131-208). Manchester, England: Manchester University Press.

Haas, W. (1983). Determining the level of a script. In F. Coulmas, & K. Ehlich (Eds.), *Writing in focus* (pp. 15-29) Berlin, Germany: Mouton.

Habein, Y. S. (1984). *The history of the Japanese written language.* Tokyo: University of Tokyo Press.

Habein, Y. S., & Mathias, G. B. (1991). *The complete guide to everyday kanji.* Tokyo: Kodansha.

Halliday, M. A. K. (1985). *Spoken and written language.* Victoria, Australia: Deakin University, School of Education.

Halpern, J. (Ed.) (1990). *New Japanese-English character dictionary.* Tokyo: Kenkyusha.

Hampton, J. (1991). The combination of prototype concepts. In P. A., Schwannenflugel (Ed.), *The psychology of word meaning* (pp. 91-116). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Hansen, C. (1993). Chinese ideographs and western ideas. *The Journal of Asian Studies, 52,* 373-399.

Harley, T. A. (2001). *The psychology of language: From data to theory* (2nd ed.). Hove, England: Psychology Press.

Hatano, G. (1995). The psychology of Japanese literacy: Expanding 'the practice account'. In L. M. W. Martin, K. Nelson, & E. Tobach (Eds.). *Sociocultural psychology: Theory and practice of doing and knowing* (pp. 250-275). Cambridge, England: Cambridge University Press.

Hatano, G., Kuhara, K., & Akiyama, M., (1981). Kanji help readers of Japanese infer the meaning of unfamiliar words. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 3,* 30-33.

Hatta, T. (1992). The effects of kanji attributes on visual field differences: Examination with lexical decision, naming and semantic classification tasks. *Neuropsychologia, 30,* 361-371.

Hatta, T., & Kawakami, A. (1996). Lexical and naming processes of non-prototypical kanji: Evidence of the component parts activation. *Asia Pacific Journal of Speech, Language, and Hearing, 1,* 55-64.

Hatta, T., & Saito, H. (Eds.) (1999). Cognitive processing of the Japanese and Chinese languages I [Special issue]. *Psychologia, 42*(4).

Hatta, T., & Saito, H. (Eds.) (2000). Cognitive processing of the Japanese and

Chinese languages II [Special issue]. *Psychologia, 43*(1).

Hayes, E. B. (1988) Encoding strategies used by native and non-native readers of Chinese Mandarin. *The Modern Language Journal, 72,* 188-95.

Henderson, L. (1982). *Orthography and word recognition in reading.* London: Academic Press.

Henderson, L. (1984). Writing systems and reading processes. In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology and linguistics*, (pp. 11-24). Hove, England: Erlbaum.

Henshall, K. G. (1988). *A guide to remembering Japanese characters.* Tokyo: Tuttle & Co.

Hill, A. A. (1967). The typology of writing systems. In W. M. Austin. (Ed.). *Papers in linguistics in honor of Leon Dostert*, (pp 92-99). The Hague: Mouton.

Hirose, H. (1992). Jukugo no nichi katei ni kan suru kenkyū: Puraimingu hō ni yoru kentō [An investigation of the recognition process for jukugo by use of priming paradigms]. *Shinrigaku Kenkyū, 63,* 303-309.

Hirose, H (1994) Kanji no kioku kōzō ni kansuru kenkyū: 1ji kanji to jukugo no kanren ni tsuite [An investigation of the memory structure of Kanji: The relation of Kanji and Jukugo]. *The Japanese Journal of Psychonomic Science*, 12, 71-76

Hirose, H. (1996a). Functions of graphemic components of kanji on recognition of jukugo. *Hiroshima Forum for Psychology, 17,* 49-55.

Hirose, H. (1996b). Reply to Dr. Saito's Comments. *Hiroshima Forum for Psychology, 17,* 67-68.

Horodeck, R. A. (1987). *The role of sound in reading and writing kanji.* Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Ijuin, M., Fushimi, T., Patterson, K., & Tatsumi, I. (1999). A connectionist approach to Japanese kanji word naming. *Psychologia, 42,* 267-280.

Ishii, Shōji (Ed.) (1996). *Reinbō Shōgaku Kanji Jiten* [Rainbow elementary school kanji dictionary]. (Rev. Ed.). Tokyo: Gakken.

Jarema, G., Kehayia, E., & Libben, G. (Eds.). (1999) Mental lexicon [Special issue]. *Brain and Language, 68*(1/2).

Jarvella, R. J., Job, R., Sandstrom, G., & Schreuder, R. (1987). Morphological constraints on word recognition. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer, (Eds.). *Language perception and production: Relationships between listening, speaking, reading, and writing* (pp. 245-262). London: Academic Press.

Joyce, T. (1999). Lexical access and the mental lexicon for two-kanji compound words: A priming paradigm study. *Proceedings of the 2nd International Conference on Cognitive Sciences and 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference,* 27-30 July, Tokyo, Japan, 511-514.

Joyce, T. (2001). Kanji niji jukugo no puraimingu: Kōzō yōso jōhō kurosu suririputo no teiji to kurosu modariti no teiji [Two-kanji compound word priming: Cross-script and cross-modality presentation of constituent information] Nihon Shinrigaku Kai Dai 65 Kai Taikai Happyō Ronbunshū 218.

Joyce, T. (in press). Constituent-morpheme priming: Implications from the morphology of two-kanji compound words. *Japanese Psychological*

Research.

Joyce, T. & Ohta, N. (1999). The morphology of two-kanji compound words: Data from a word-formation classification survey. *Tsukuba Psychological Research, 22,* 45-61.

Joyce, T. & Ohta, N. (in press). Constituent morpheme frequency data for two-kanji compound words. *Tsukuba Psychological Research.*

Kabashima, T. (1977). Moji no taikei to kōzō [The systems and construction of letters]. In Iwanami Japanese language series. *Writing* (Vol. 8) (pp. 23-60). Tokyo: Iwanami Shoten.

Kadota, S. & Ishikawa, K. (1999). On the use of phonology in accessing the meaning of printed words. *Proceedings of the 2nd International Conference on Cognitive Sciences and 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference,* 27-30 July, Tokyo, Japan, 298-303.

Kageyama, T. (1982). Word formation in Japanese. *Lingua, 57,* 215-258.

Kaiho, H., & Nomura, Y. (1983) *Kanji jōhō shori no shinrigaku* [The psychology of kanji information processing]. Tokyo: Kyōiku Shuppan.

Kaiser, S. (1993a). *Japanese correspondence course for JET participants 1994-1995: Vol. III, Japanese language.* Tokyo: Council of Local Authorities for International Relations.

Kaiser, S. (1993b). Nihongo goi no kindaika: Gairaigo ukeire no hōhō ni tsuite [The modernization of the Japanese lexicon: On the acceptance of loan words]. In Committee to commemorate the 77 th. birthday of Prof. A. Matsumura (Ed.), *Kokugo Kenkyū* (pp. 470-489). Tokyo: Meiji Shoten.

Kaiser, S. (1995). Seikai no moji, chūgoku no moji, nihon no moji: Kanji no ichiduke saikō [The writing systems of the world, the Chinese writing system, the Japanese writing system: Reconsidering the status of kanji]. *Seikai Nihongo Kyōiku, 5,* 155-167.

Kao, H. S. R., van Galen, G. P., & Hoosain, R. (Eds.). (1986). *Graphonomics: Contemporary research in handwriting.* Amsterdam: North-Holland.

Kawakami, M. (1997). JIS isshu kanji 2965 ji o mochiite sakusei sareru kanji niji jukugo sūhyō: Macintosh ban iwanami kōjien daiyonban ni motozuku ruijigosū chōsa [Numerical data for two-kanji compound words formed from 2965 JIS level 1 kanji characters: Survey of synonyms based on Iwanami's Kojien dictionary (4th edition CD for Macintosh)]. *Bulletin of the School of Education, Nagoya University, 44,* 243-299.

Kawakami, M. (2000). JIS isshu kanji 2965 ji wo mochiita kanji niji jukugo no ruiseki ruijigo hindo hyō [A table of cumulative frequencies of Japanese neighbor-kanji-compound words] *Dokkai Kagaku, 44,* 150-159.

Kawakami, M., Saito, H., & Yanase, Y. (2000). Jukugorashisa no shinteki hyōka kikō to sono shinteki zaisei: Giji jukugo no sanshutsu to juyō [Evaluation mechanism of compound-wordlikeness and its psychological reality: Production and recognition of pseudo compound words] *Ninchi Kagaku, 7,* 303-318.

Keene, D. (1986). The difficulty I had learning Japanese. *Japan Quarterly, 3,* 2-7.

Kegl, J., Senghas, A., & Coppola, M. (1999). Creation through contact: Sing language emergence and sign language change in Nicaragua. In M.

DeGraff. (Ed.). *Language creation and language change: Creolization, diachrony, and development* (pp. 179-237). Cambridge, MA: The MIT Press.

Kehayia, E., Jarema, G., Tsapkini, K., Perlak, D., Ralli, A., & Kadzielawa, D. (1999). The role of morphological structure in the processing of compounds: The interface between linguistics and psycholinguistics. Brain and Language, 68, 370-377.

Keightley, D. N. (1989). The origins of writing in China: Scripts and cultural contexts. In W. Senner (Ed.), *The origins of writing* (pp. 171-202). Lincoln, NE: University of Nebraska Press.

Kello, C. T., & Kawamoto, A. H. (1998). Runword: An IBM-PC software package for the collection and acoustic analysis of speeded naming responses. *Behavior Research Methods, Instruments, & Computers, 30,* 371-383.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185-209.

Kess, J. F., & Miyamoto, T. (1999). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing.* Amsterdam: John Benjamins.

Koizumi, T. (1989). Linguistic revolutionaries of modern Japan. *Japan Echo, 16,* (Special ed.), 26-30.

Krott, A., Baayen, R. H., & Schreuder, R. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics, 39,* 51-93.

Kuhara-Kojima, K., & Hatano, G. (1992). Inferring meaning of unfamiliar kanji compound words. In B. Wilpert, H Motoaki, & J. Misumi (Eds.), *Proceedings of the 22nd international congress of applied psychology: Vol. 2 General psychology and environmental psychology* (p. 139). Hillsdale, NJ: Erlbaum.

Kuratani, N., Kobayashi, A., & Okunishi, S. (1982). *A new dictionary of kanji usage.* Tokyo: Gakken.

Laudanna, A., & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In L. B. Feldman, (Ed.), *Morphological aspects of language processing* (pp. 345-364). Hove, England: Lawrence Erlbaum Associates.

Leong, C. K., & Tamaoka, K. (Eds.). (1998). Cognitive processing of the Chinese and Japanese languages [Special issue]. *Reading and Writing, 10*(3-5).

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Liu, Y., & Peng, D.-L. (1997). Meaning access of Chinese compounds and its time course. In H.-C. Chen (Ed.), *Cognitive processing of Chinese and other related Asian languages* (pp. 219-232). Sha Tin, Hong Kong: The Chinese University Press.

Lyons, J. (1981). *Language and linguistics, An introduction.* Cambridge, England: Cambridge University Press.

Mair, V. H. (1991). Building the future of information processing in East Asia demands facing linguistic and technological reality (Preface). In V. H. Mair, & Y. Liu, (Eds.), *Characters and computers* (pp.). Amsterdam:

IOS Press.

Mair, V. H. (1996). Modern Chinese writing. In P. T. Daniels & W. Bright (Eds.). *The world's writing systems* (pp. 200-208). New York: Oxford University Press.

Makita, K. (1968). The rarity of reading disability in Japanese children. *American Journal of Orthopsychiatry, 38,* 599-614.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71-102.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101,* 2-33.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29-63.

Martin, S. F. (1972). Nonalphabetic writing systems: Some observations. In J. F. Kavanagh, & I. G. Mattingly, (Eds.), *Language by ear and by eye: The relationship between speech and reading* (pp. 81-102). Cambridge, MA: MIT Press.

Masuda, H., & Saito, H. (1997, December). *Effect of radical companion of ideographic characters in a cue report task.* Paper presented at the 8th International Conference on Cognitive Processing of Asian Languages & Symposium on Brain, Cognition, and Communication, Nagoya, Japan.

Matsunaga, S. (1996). The linguistic nature of kanji reexamined: Do kanji represent only meanings? *Journal of the Association of Teachers of*

Japanese, 30, 1-22.

Matthews, P. H. (1991) *Morphology* (2nd ed.). Cambridge, England: Cambridge University Press.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375-407.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology, Learning, memory, and Cognition, 24,* 558-572.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227-234.

Microsoft/ Shōgakukan, (1997) *Bukkusherufu bēshikku: Maruchimedia tōgō jiten* [Bookshelf basic: Multimedia comprehensive dictionary] CR-ROM.

Miller, R. A. (1967). *The Japanese language.* Chicago: University of Chicago Press.

Miller, R. A. (1986). *Nihongo: In defense of Japanese.* London: Athlone Press.

Mizuno, R. (1997). Kanji hyōkigo no onin shori jidōka kasetsu no kenshō [A test of a hypothesis of automatic phonological processing of kanji words]. *Shinrigaku Kenkyū, 68,* 1-8.

Monsell, S. (1985). Repetition and the lexicon. In A. W. Ellis (Ed.), *Progress in the Psychology of Language, Vol. 2*. London: Lawrence Erlbaum Associates.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.).

Models of human memory (pp. 203-254). New York: Academic Press.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. E. Wrolstrad, & H. Bouma, (Eds.), *Processing of visible language: Vol. 1* (pp. 259-268). New York: Plenum Press.

Morton, J. & Sasanuma, S. (1984). Lexical access in Japanese. In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology and linguistics*, (pp. 25-42). Hove, England: Erlbaum.

Morton, J., Sasanuma, S., Patterson, K., & Sakuma, N. (1992). The organization of the lexicon in Japanese: Single and compound kanji. *British Journal of Psychology, 83,* 517-531.

Nakanishi, A. (1980). *Writing systems of the world: Alphabets, syllabaries, pictograms*. Tokyo: Charles E. Tuttle.

Napps, S. E. (1989). Morphemic relationships in the lexicon: Are they distinct from semantic and formal relationships? *Memory and Cognition, 17,* 729-739.

National Language Research Institute, (1962-64) *Gendai zasshi kyūjū shu no yōgo yōji* [Vocabulary and characters in 90 current magazines] (Research reports 21, 22, and 25). Tokyo: Shuei Shuppan.

National Language Research Institute, (1976) *Gendai shinbun no kanji* [Kanji in current newspapers] (Research report 56). Tokyo: Shuei Shuppan.

National Language Research Institute, (1984). *Nihongo kyōiku no tame no kihon goi chōsa* [A study of the fundamental vocabulary for Japanese language teaching: General description and vocabulary tables].

(Research report 78). Tokyo: Shuei Shuppan.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 3, 226-254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys, (Eds.), *Basic processes in reading: Visual word recognition*, (pp. 264-336). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nelson, A. N. (1962). *The modern reader's Japanese-English character dictionary.* Tokyo: Tuttle & Co.

Nomura, M. (1988a). Niji kango no kôzô [The structure of 2 kanji Sino-Japanese words]. *Nihongogaku, 7, 5,* 44-55.

Nomura, M. (1988b). Kanji no zōgoryoku [Word formation productivity of kanji]. In Kiyoji Satō (Ed.), *Kanji kōza 1 Kanji to wa* (pp. 193-217). Tokyo, Japan: Meiji Shoin.

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition, 22,* 93-136.

Novik, N. (1974). Parallel processing in a word-nonword classification task. *Journal of Experimental Psychology, 102,* 1015-1020.

Nozaki, H., Shimizu, Y., & Yokoyama, S. (1999). Shinbun kiji CD-ROM o mochiite kanji jukugo no hindo tokusei ni kan suru chōsa kennkyū [The study of kanji word frequency on CD-ROM newspaper articles]. *Proceedings of the 2nd International Conference on Cognitive Sciences and 16th Annual Meeting of the Japanese Cognitive Science Society*

Joint Conference, 27-30 July, Tokyo, Japan, 752-755.

Okutsu, K. (1975). Fukugō meishi no seisei bunpō [Generative grammar of compound nouns]. *Kokugogaku, 101,* 19-34.

Ozaki, Y., Todome, H., Nishioka, H., Yamada, K., & Yamada, T. (Eds.). (1992) Daijigen [Large dictionary of etymology] Tokyo: Kadokawa Shoten.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23,* 443-512.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89,* 573-594.

Paradis, M., Hagiwara, H., & Hildebrandt, N. (1985). *Neurolinguistic aspects of the Japanese writing system.* Orlando, FL: Academic Press.

Perfetti, C. A. (1999). The cognitive science of word reading: What has been learned from comparisons across writing systems? *Proceedings of the 2nd International Conference on Cognitive Sciences and 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference,* 27-30 July, Tokyo, Japan, 37-42.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 101-118.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning Memory and Cognition, 21,* 24-33.

Perfetti, C. A., Zhang, S. & Berent, I. (1992). Reading in English and Chinese:

Evidence for a ‘universal’ phonological principle. In R. Frost & L. Katz (Eds). *Orthography, phonology, morphology, and meaning* (pp. 227-248). Amsterdam: North-Holland.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56-115.

Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of “neighborhood size” in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1142-1158.

Posner, M. I., & Snyder, C. R. R., (1975). Attention and Cognitive control. In Solso, R. L. (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rogers, H. (1995). Optimal orthographies. In I. Taylor, & D. R. Olson, (Eds.), Script and literacy: Reading and learning to read alphabets, syllaberies and characters, (pp. 31-43), Dordrecht, The Netherlands: Kluwer Academic Publishers.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970) Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9,* 487-494.

Rubin, G. S., & Becker, C. A., Freeman, R. H. (1979). Morphological structure and its effect on visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 18,* 757-767.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60-94.

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations*. Cambridge, MA: MIT Press.

Saito, H. (1996). The relation between radical functions within a single character kanji and roles of a single character within a compound word: A comment on Hirose's study on recognition of compound words. *Hiroshima Forum for Psychology, 17,* 57-65.

Saito, H. (1997). Shinteki jisho [Mental lexicon]. In Y. Matsumoto, T. Kageyama, M. Nagata, H. Saito, T. Tokunaga, *Iwanami Kōza Gengo no kagaku 3 Tango to jisho* (pp. 94-153). Tokyo: Iwanami Shoten.

Saito, H., & Kawakami, M. (1992a). Factors affecting lexical decision on pseudo-compound kanji characters. *International Journal of Psychology, 27,* 3-4, 110.

Saito, H., & Kawakami, M. (1992b). Rensō kioku ni okeru kensaku katei (3): Giji jukugo no jukugorashisa no hyōka ni okeru shinteki jisho no kanyo [Retrieval processes in associative memory (3): Evaluation of semantic saliency of pseudo compound words and mental lexicon]. *Kiyō,* Ser. B, College of General Education, Nagoya University, 36, 67-99.

Saito, H., Kawakami, M., Masuda, H. (1995). Kanji kōsei ni okeru buhin (bushu) no shitsugen hindo [Frequency of semantic and phonetic components of radical types in complex left-right kanji]. Jōhō Bunka Kenkyū (Bulletin of the School of Informatics and Science, Nagoya

University, Japan), 1, 113-134.

Saito, H., Masuda, H., & Kawakami, M. (1998). Form and sound similarity effects in kanji recognition. *Reading and Writing, 10,* 323-357.

Sakai, N. (1991). *Voices of the past: The status of language in eighteenth-century Japanese discourse.* Ithaca, NY: Cornell University Press.

Sakuma, N., & Fushimi, T., Tatsumi, I. F. (1997). Onseiha no shisatsu ni toru kana no ondoku senji no sokutei: Ondoku senji wa gotōon no chōonhō ni yori ōkiku kotonaru [Measurement of naming latency of kana characters and words based on the speech wave analysis: Manner of articulation of a word-initial phoneme considerably affects naming latency]. Shinkei Shinrigaku, 13, 126-136.

Sakuma, N., Sasanuma, S., Tatsumi, I. F., & Masaki, S. (1998). Orthography and phonology in reading Japanese kanji words: Evidence from the semantic decision task with homophones. *Memory & Cognition, 26,* 75-87.

Sampson, G. (1985). *Writing systems: A linguistic introduction.* Stanford, CA: Stanford University Press.

Sampson, G. (1994). Chinese script and the diversity of writing systems, *Linguistics, 32,* 117-132.

Sandler, W., & Lillo-Martin, D. (2001). Natural sign languages. In M. Aronoff & J. Rees-Miller (Eds.). *The handbook of linguistics* (pp. 533-562). Oxford, England: Blackwell.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly*

Journal of Experimental Psychology, 42A, 529-567.

Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes, 9,* 227-269.

Sandra, D., & Taft, M. (Eds.) (1994). Morphological structure, lexical representation and lexical access [Special issue]. *Language and Cognitive Processes, 9*(3).

Sansom, G. (1928). *A historical grammar of Japanese.* Oxford, England: Oxford University Press.

Satō, K. et al. (Eds.) (1996). *Kanji hyakka daijiten* [Encyclopedic dictionary of kanji]. Tokyo: Meiji Shoin.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In Laurie Beth Feldman, (Ed.), *Morphological aspects of language processing* (pp. 131-154). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language, 37,* 118-139.

Schreuder, R., Neijt, A., van der Weide, F., & Baayen, R. H. (1998). Regular plurals in Dutch compounds: Linking graphemes or morphemes? *Language and Cognitive Processes, 13,* 551-573.

Scribner, S., & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876-900.

Seeley, C. (1984). The Japanese script since 1900. *Visible Language, 18,* 267-301.

Seeley, C. (1991). *A history of writing in Japan.* Leiden: E. J. Brill.

Seidenberg, M. S. (1989). Visual word recognition and pronunciation: A computational model and its implications. In W. Marslen-Wilson, (Ed.). *Lexical representation and process* (pp. 25-74). Cambridge, MA: MIT Press.

Seidenberg, M. S. (1990) Lexical access: Another theoretical soupstone? In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner, Comprehension processes in Reading, (pp. 33-71). Hillsdale, NJ: Lawrence Erlbaum Associates.

Seidenberg, M. S. (1995). Visual word recognition: An overview. In J. Miller & P. D. Eimas (Eds.). *Speech, language, and communication* (pp. 137-179). San Diego, CA: Academic Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523-568.

Senghas, A. (1995). The development of Nicaraguan Sign Language via the language acquisition process. In D. MacLaughlin, & S. McEwen, (Eds.), *Proceedings of the 19th annual Boston university conference on language development (vol. 2)* (pp. 543-552). Somerville, MA: Cascadilla Press.

Senner, W. M. (1989). Theories and myths on the origins of writing: A historical overview. In W. M. Senner, (Ed.), *The origins of writing* (pp. 1-26). Lincoln, NE: University of Nebraska Press.

Shafiullah, M., & Monsell, S. (1999). The cost of switching between kanji and kana while reading Japanese. *Language and Cognitive Processes, 14,* 567-607.

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1191-1210.

Shen, D., & Forster, K. I. (1999). Masked phonological priming in reading Chinese words depends on the task. *Language and Cognitive Processes, 14,* 429-459.

Shibatani, M. (1990). *The languages of Japan.* Cambridge, England: Cambridge University Press.

Shimamura, A. P. (1987). Word comprehension and naming: An analysis of English and Japanese orthographies. *American Journal of Psychology, 100,* 15-40.

Shimomura, M. & Yokosawa, K. (1991). Processing kanji and kana characters within Japanese words. *Perception and Psychophysics, 50,* 19-27.

Shimomura, M. & Yokosawa, K. (1992). Identification of kanji compounds in Japanese words. In B. Wilpert, H Motoaki, & J. Misumi, (Eds.), *Proceedings of the 22nd international congress of applied psychology: Vol. 2. General psychology and environmental psychology* (pp. 138-139). Hillsdale, NJ: Erlbaum.

Shimomura, M. & Yokosawa, K. (1995). The processing of constituent characters in kanji words in proofreading Japanese sentences. *Psychological Research, 58,* 51-60.

Singleton, D. (2000). Language and the lexicon: An introduction. London:

Arnold.

Smith, J. S. (1996). Japanese writing. In P. T. Daniels & W. Bright (Eds.). *The world's writing systems* (pp. 209-217). New York: Oxford University Press.

Sproat, R. (2000). *A computational theory of writing systems.* Stanford, CA: Cambridge University Press.

Stanners, R. F. & Forbach, G. B. (1973). Analysis of letter strings in word recognition*. Journal of Experimental Psychology, 98,* 31-35.

Stanners, R. F., Forbach, G. B., Headley, D. B. (1971). Decision and search processes in word-nonword classification. *Journal of Experimental Psychology, 90,* 45-50.

Steinberg, D. D., Nagata, H., & Aline, D. P. (2001). *Psycholinguistics: Language, mind and world* (2nd ed.). Harlow, England: Longman.

Steinberg, D. D., & Yamada, J. (1978-9). "Are whole word kanji easier to learn than syllable kana?" *Reading Research Quarterly, 14,* 88-99.

Stevenson, H. W., Lee, S. -Y., Stigler, J., Kitamura, S., Kimura, S., & Kato, T. (1986) Learning to read Japanese. In H. Stevenson, M. Azuma, & K Hakuta, (Eds.), *Child development and education in Japan* (pp. 217-235). New York: Freeman.

Stolz, J. A., & Feldman, L. B. (1995). The role of orthographic and semantic transparency of the base morpheme in morphological processing. In L. B. Feldman, (Ed.). *Morphological aspects of language processing* (pp. 109-129). Hove, England: Lawrence Erlbaum Associates.

Sugishita M., Otomo K., Kabe S., & Yunoki K. (1992). A critical appraisal of neuropsychological correlates of Japanese ideogram (kanji) and phonogram (kana) reading. *Brain, 115,* 1563-1585.

Suzuki, T. (1975). On the twofold phonetic realization of basic concepts: In defense of Chinese characters in Japanese. In F. C. C. Peng (Ed.), *Language in Japanese society* (pp. 175-192). Tokyo: University of Tokyo Press.

Suzuki, T. (1977). Writing is not language, or is it? *Journal of Pragmatics, 1,* 407-420.

Suzuki, T. (1978). Kango wa gairaigo ka [Are Sino-Japanese words foreign loan words?] *Gengo, 7,* 2-8.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7,* 263-272.

Taft, M. (1991). *Reading and the mental lexicon.* Hillsdale, NJ: Erlbaum.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9,* 271-294.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behaviour, 14,* 638-647.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior, 15,* 607-620.

Taft, M., Huang, J., & Zhu, X. (1994). The influence of character frequency on word recognition responses in Chinese. In H.-W. Chang, J. T. Huang, C. W. Hue, & O. Tzeng (Eds.), *Advances in the Study of Chinese Language Processing: Volume 1* (pp. 59-73). Taipei, Taiwan: Department of Psychology, National Taiwan University.

Taft, M., Liu, Y., & Zhu, X. (1999). Morphemic processing in reading Chinese. In J. Wang, A. Inhoff, & H.-C. Chen, (Eds.), *Reading Chinese script: A cognitive analysis* (pp 91-113). Hillsdale, NJ: Lawrence Erlbaum Associates.

Taft, M., & Zhu, X. (1995). The representation of bound morphemes in the lexicon: A Chinese study. In L. B. Feldman, (Ed.), *Morphological aspects of language processing* (pp. 293-316). Hove, England: Lawrence Erlbaum Associates.

Taft, M., & Zhu, X. (1997a). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23,* 761-775.

Taft, M., & Zhu, X. (1997b). Using masked priming to examine lexical storage of Chinese compound words. In H.-C. Chen (Ed.), *Cognitive Processing of Chinese and Related Asian Languages* (pp. 233-241). Shatin, Hong Kong: The Chinese University Press.

Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. Journal of Memory and Language, 40, 498-519.

Tagashira, Y., & Hoff, J. (1986) *Handbook of Japanese compound verbs.* Tokyo: Hokuseido Press.

Tamamura, F. (1985). Go no kōsei to zōgohō [The structure of words and principles of word formation] In Kokuritsu Kokugo Kenkyûjo. *Go no kenkyū to kyōiku (shita) Nihongo kyōiku shidō sankōsho 13* (pp. 1-78). Tokyo, Japan: Ōkurashō Insatsukyoku.

Tamamura, F. (1988). Fukugōgo no imi [The meaning of compound words].

Nihongogaku, 7, 5, 23-32.

Tamaoka, K. (1991). Psycholinguistic nature of Japanese orthography. *Studies in Language and Literature, 11,* 49-82.

Tamaoka, K. (1994). A review of psychological studies of kana and kanji: A single phonological route to a multiple interactive activation. *Studies in Language and Literature, 13,* 1-18.

Tamaoka, K., & Hatsuzuka, M. (1995). Kanji niji jukugo no shori ni okeru kanji shiyō hindo no eikyō [The effects of kanji printed-frequency on processing Japanese two-morpheme compound words]. *Dokkai Kagaku, 39,* 121-137.

Tamaoka, K., & Hatsuzuka, M. (1998). The effects of morphological semantics on the processing of Japanese two-kanji compound words. *Reading and Writing, 10,* 293-322.

Tan, L. H. & Perfetti, C. A. (1997). Visual Chinese character recognition: Does phonological information mediate access to meaning? *Journal of Memory and Language, 37,* 41-57.

Tan, L-H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. *Reading and Writing, 10,* 165-200.

Taylor, I. (1883). *The alphabet: An account of the origin and development of letters*, 2 vols. London: Kegan, Paul, Trench.

Taylor, I., & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean and Japanese.* Amsterdam: John Benjamins.

Tsuzuki, T. (1998). Ninchi shinrigaku ni okeru kioku hyōshō to gengo shori ni kansuru nettowāku moderu: Imi nettowāku, bun no nettowāku, heiretsu bunsan shori. [Cognitive psychology network models of memory representation and language processing: Semantic networks, sentence model representations, parallel distributed processing]. *Ōyō Shakaigaku* Kenkyū *(Rikkyō Daigaku), 40,* 27-44.

Tsuzuki, T., Kawamoto, A. H., & Yukihiro, R. (1999). Goiteki tagisei no shori ni kansuru heiretsu bunsan shori moderu: Bunmyaku to tomo ni teiji sareta tagigo no ninchi ni kansuru jikken dēta no rironteki tōgō [A parallel distributed processing model of lexical ambiguity resolution: The theoretical integration of empirical data for processing ambiguous Japanese words presented with contexts]. *Ninchi Kagaku, 6,* 91-104.

Twine, N. (1991). *Language and the modern state: The reform of written Japanese.* London: Routledge.

Unger, J. M. (1984). Japanese orthography in the computer age. *Visible Languages, 18,* 238-253.

Unger, J. M. (1987). *The fifth generation fallacy: Why Japan is betting its future on artificial intelligence.* Oxford, England: Oxford University Press.

Unger, J. M. (1990). The very idea: The notion of ideography in China and Japan. *Monumenta Nipponica, 45,* 391-412.

Unger, J. M., & DeFrancis, J. (1995). Logographic and semasiographic writing systems: A critique of Sampson's classification. In I. Taylor, & D. R. Olson, (Eds.), *Script and literacy: Reading and learning to read alphabets, syllabaries and characters*, (pp. 45-58), Dordrecht, The Netherlands: Kluwer Academic Publishers.

Van Jaarsveld, H. J., & Rattink, G. E. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of*

Psycholinguistic Research, 17, 447-473.

Van Orden, G. C. (1987). A rows is a rose: Spelling, sound, and reading. *Memory & Cognition, 15,* 181-198.

Watt, W. C. (Ed.) (1994). *Writing systems and cognition: Perspectives from psychology, physiology, linguistics, and semiotics.* Dordrecht, The Netherlands, Kluwer Academic Publishers.

Wisniewski, E. J., & Gentner, D. (1991). On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning. In G. B. Simpson, (Ed.), *Understanding word and sentence* (pp. 241-284). Amsterdam: North-Holland.

Wydell, T., (1991). *Processing in the reading of Japanese: Comparative studies between English and Japanese orthographies.* Unpublished doctoral dissertation, London University, England.

Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for kanji: Is a *rows* still a *rose* in Japanese kanji? *Journal of Experimental Psychology Learning, Memory, and Cognition, 19,* 491-514.

Yamada, J. (1992a). Asymmetries of reading and writing Kanji by Japanese children. *Journal of Psycholinguistic Research, 21,* 563-580.

Yamada, J. (1992b). Why are kana words named faster than kanji words? *Brain and Language, 43,* 682-693.

Yamada, J. (1994). Latencies for naming bound morphemes and words in kanji. *Language and Speech, 37,* 251-258.

Yamada, J. (1997). *Learning and information processing of Kanji and Kana.* Hiroshima, Japan: Keisuisha.

Yamada, J. (1998). The time course of semantic and phonological access in naming kanji and kana words. *Reading and Writing, 10,* 425-437.

Yamada, J. & Kayamoto, Y. (1998). Valency, secondary frequency, and lexical access: A Japanese study. *Applied Psycholinguistics, 19,* 87-97.

Yamada, J., Mitarai, Y., & Yoshida, T. (1991). Kanji words are easier to identify than katakana words. *Psychological Research, 53,* 136-141.

Yokosawa, K. & Umeda, M. (1988). Processes in human kanji-word recognition. *Proceedings of the 1988 IEEE international conference on systems, man, and cybernetics* (pp. 377-380). August 8-12, 1988, Beijing and Shenyang, China.

Yokoyama, S., Sasahara, H., Nozaki, K., & Long. E., (1998). *Shinbun denshi media no kanji: Asahi Shinbun CD-ROM ni yoru kanji hindohyō* [Electronic newspaper media kanji: Kanji frequency lists based on Asahi Newspaper CD-ROM]. Tokyo: Sanseido.

Yokoyama, S., Sasahara, H., Long, E., & Tanimoto, S. (2001). Shinbun kanji chōsa no genzai to shōrai [The present and future of newspaper kanji surveys] Nihongo Kagaku, 9, 33-42.

Zhang, B. & Peng, D.-L. (1992) Decomposed storage in the Chinese lexicon. In H.-C. Chen & O. J. L. Tzeng (Eds.). *Language processing in Chinese* (pp. 131-149). Amsterdam: North-Holland.

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes, 9*, 393-422.

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes, 10,* 545-600.

Zhou, X., & Marslen-Wilson, W. (1999a). The nature of sublexical processing in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 819-837.

Zhou, X., & Marslen-Wilson, W. (1999b). Phonology, orthography, and semantic activation in reading Chinese. *Journal of Memory and Language, 41,* 579-606.

Zhou, X., & Marslen-Wilson, W. (2000). Lexical representation of compound words: Cross-linguistic evidence. *Psychologia, 43,* 47-66.

Zwisterlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes, 9,* 341-368.

	CS	FS		CS	FS		CS	FS
Modifier + modified								
旅館	6.2	6.5	美術	6.0	6.4	今回	6.2	6.6
夕飯	6.7	6.7	住所	6.3	6.9	豚肉	6.4	6.4
多数	6.6	6.4	漢字	6.2	6.6	敬語	5.9	6.4
国道	6.8	6.4	市民	6.2	6.5	洋画	6.6	6.2
名札	6.2	6.4	古本	6.7	6.6	休日	5.9	6.8
恋人	6.0	6.6	体重	6.3	6.6	弱点	6.6	6.0
Verb + complement								
育児	6.6	6.0	駐車	6.4	6.4	投球	6.8	5.7
犯罪	6.4	6.6	禁煙	6.4	6.4	喫茶	6.5	5.9
預金	6.6	6.6	飲酒	7.0	6.5	評価	6.2	6.4
免税	6.9	5.9	充電	6.7	5.9	開店	7.0	5.9
提案	6.7	6.5	就職	6.2	6.5	作文	6.4	6.6
断念	6.4	6.2	録音	6.4	6.2	伝言	6.4	6.1
Complement + verb								
外食	6.5	6.3	夜勤	6.9	6.2	公認	6.8	6.1
逆転	6.5	6.3	水泳	6.0	6.4	予習	6.7	6.4
特定	6.7	5.9	別居	6.6	6.1	自信	6.7	6.7
仮設	6.4	5.8	歓迎	6.2	5.8	未婚	6.5	6.0
広告	6.4	5.9	共感	6.4	5.9	昼寝	6.3	6.2
再建	6.8	5.8	専用	6.2	6.5	軽視	6.5	6.0

(table continues)

	CS	FS		CS	FS		CS	FS
Associative pairs								
心身	6.8	5.9	父母	6.7	6.1	親子	6.9	6.7
男女	6.9	6.4	難易	6.9	6.0	優劣	6.9	6.4
生死	6.9	5.9	手足	6.6	6.1	出欠	6.9	6.6
進退	6.9	6.1	左右	6.9	6.9	前後	6.9	6.3
上下	6.9	6.8	大小	6.9	5.9	欧米	6.7	6.5
明暗	6.8	6.6	勝負	6.9	6.2	往復	6.9	5.9
Synonymous pairs								
歴史	5.9	6.4	詳細	5.9	6.2	睡眠	6.0	6.6
停止	6.4	6.1	利益	6.5	6.5	苦痛	6.0	6.3
破壊	6.4	6.3	行為	6.1	6.3	労働	5.8	6.6
中央	5.8	6.7	変化	6.2	6.7	年齢	6.8	6.7
貯蓄	6.8	6.7	終了	6.6	6.1	募集	6.2	6.7
省略	6.6	6.3	収穫	6.0	6.4	乾燥	6.1	6.5

Note. CS = classification score. FS = familiarity score. These scores are from Joyce and Ohta (1999).

実験の教示

この実験は、漢字の認知に関するものです。

実験では、最初に「＋」という記号がコンピューターの画面の真中に呈示され、その後同じ位置に漢字 1 字が呈示されます。この漢字 1 字は見ていて下さい。

次に、「＋＋」という記号が真中に呈示された後、今度は漢字の対が呈示されます。この実験でして頂きたいことは、なるべく速く正確に、この漢字の対が意味をもつ、本当の単語であるかどうかを判断して、レスポンスボックスの正しいキーを押すことです。

もし、この漢字の対が本当の単語であると思えば、「緑」のキーを押して下さい。また、もし、この漢字の対が本当の単語ではないと思えば、「赤」のキーを押して下さい。

レスポンスボックスのキーを押したら、次の項目の集合が現われるので、前に行った判断を気にせずに、次の項目の集合を見て下さい。

本実験に入る前に、漢字について判断し、レスポンスボックスのキーを押すことを慣れて頂くために、練習セッションを 10 回行います。

その後、本実験に入ります。

この教示に対してご質問がありましたら、実験に入る前に、気軽に聞いて下さい。

	CS	FS		CS	FS		CS	FS
Modifier + modified								
旅館	6.2	6.5	灰皿	6.2	5.7	美術	6.0	6.4
今回	6.2	6.6	夕飯	6.7	6.7	白菜	6.0	6.3
住所	6.3	6.9	豚肉	6.4	6.4	制服	6.0	6.1
北緯	6.1	5.7	多数	6.6	6.4	同時	6.3	5.8
農村	6.0	5.7	漢字	6.2	6.6	表面	6.3	5.9
敬語	5.9	6.4	国道	6.8	6.4	市民	6.2	6.5
半額	6.4	5.6	洋画	6.6	6.2	牛乳	6.1	5.9
政党	6.0	5.8	名札	6.2	6.4	古本	6.7	6.6
休日	5.9	6.8	体重	6.3	6.6	恋人	6.0	6.6
黒板	6.3	5.8	弱点	6.6	6.0	愛犬	6.2	5.9
Verb + complement								
育児	6.6	6.0	駐車	6.4	6.4	被害	6.7	5.6
投球	6.8	5.7	延期	6.6	5.5	製薬	6.4	5.5
犯罪	6.4	6.6	禁煙	6.4	6.4	喫茶	6.5	5.9
脱毛	6.6	5.6	預金	6.6	6.6	飲酒	7.0	6.5
評価	6.2	6.4	看病	6.3	5.8	失業	6.4	5.6
免税	6.9	5.9	充電	6.7	5.9	接客	6.7	5.7
開店	7.0	5.9	保温	6.7	5.6	提案	6.7	6.5
就職	6.2	6.5	作文	6.4	6.6	帰宅	6.7	6.0
占領	6.5	5.6	断念	6.4	6.2	編曲	6.5	6.0
入社	6.8	5.8	録音	6.4	6.2	伝言	6.4	6.1

(table continues)

	CS	FS		CS	FS		CS	FS
Complement + verb								
潜在	6.6	5.5	密輸	6.3	6.0	外食	6.5	6.3
改築	6.6	5.7	夜勤	6.9	6.2	誤算	6.2	6.1
公認	6.8	6.1	逆転	6.5	6.3	水泳	6.0	6.4
予習	6.7	6.4	特定	6.7	5.9	伴奏	6.2	5.7
別居	6.6	6.1	自信	6.7	6.7	仮設	6.4	5.8
歓迎	6.2	5.8	口答	6.0	5.6	未婚	6.5	6.0
先着	6.7	5.8	私立	6.1	5.6	広告	6.4	5.9
共感	6.4	5.9	昼寝	6.3	6.2	傍観	6.8	5.6
再建	6.8	5.8	独学	6.6	5.6	等分	6.8	5.7
専用	6.2	6.5	軽視	6.5	6.0	毒殺	6.5	5.8
Associative pairs								
心身	6.8	5.9	父母	6.7	6.1	朝晩	6.9	5.5
兄弟	6.6	5.6	親子	6.9	6.7	男女	6.9	6.4
夫婦	6.7	5.8	難易	6.9	6.0	優劣	6.9	6.4
高低	6.9	5.9	生死	6.9	5.9	東西	6.9	5.8
天地	6.9	5.6	手足	6.6	6.1	増減	6.9	5.8
出欠	6.9	6.6	進退	6.9	6.1	左右	6.9	6.9
前後	6.9	6.3	遠近	6.9	5.7	伸縮	6.9	5.5
上下	6.9	6.8	緩急	6.9	5.6	胃腸	6.5	5.7
大小	6.9	5.9	欧米	6.7	6.5	和英	6.6	5.7
明暗	6.8	6.6	勝負	6.9	6.2	往復	6.9	5.9

(table continues)

	CS	FS		CS	FS		CS	FS
Synonymous pairs								
歴史	5.9	6.4	詳細	5.9	6.2	永久	6.4	6.1
柔軟	6.6	6.1	睡眠	6.0	6.6	思考	5.8	5.9
停止	6.4	6.1	貧乏	6.4	6.0	利益	6.5	6.5
苦痛	6.0	6.3	破壊	6.4	6.3	行為	6.1	6.3
労働	5.8	6.6	中央	5.8	6.7	変化	6.2	6.7
豊富	6.3	5.8	削除	6.0	6.0	年齢	6.8	6.7
河川	6.8	6.3	意志	5.8	6.3	貯蓄	6.8	6.7
繁盛	6.5	5.9	終了	6.6	6.1	超過	6.4	5.7
募集	6.2	6.7	戦争	6.6	5.6	援助	6.3	5.8
省略	6.6	6.3	収穫	6.0	6.4	乾燥	6.1	6.5

Note. CS = classification score. FS = familiarity score. These scores are from Joyce and Ohta (1999).

実験の教示

この実験は、漢字の認知に関するものです。

実験では、最初にコンピューターの画面の真中に「＋」という記号、次に漢字 1 字、最後に「※」という記号が順番に瞬間的に呈示されますので、それらを見ていて下さい。その後すぐ、真中に 2 つの漢字が呈示されます。この実験でして頂きたいことは、なるべく速く正確に、この漢字の対が意味をもつ、本当の単語であるかどうかを判断して、レスポンスボックスの正しいキーを押すことです。この判断は、前に見た漢字 1 字が、その漢字対の 1 字であったのかどうかということは考えないで行って下さい。

もし、この漢字の対が本当の単語であると思えば、「緑」のキーを押して下さい。また、もし、この漢字の対が本当の単語ではないと思えば、「赤」のキーを押して下さい。

レスポンスボックスのキーを押したら、次の項目の集合が現われるので、前に行った判断を気にせずに、次の項目の集合を見て下さい。

本実験に入る前に、漢字について判断し、レスポンスボックスのキーを押すことを慣れて頂くために、練習セッションを 10 回行います。その後、本実験に入ります。

この教示に対してご質問がありましたら、実験に入る前に、気軽に聞いて下さい。

	CS	FS		CS	FS		CS	FS
Modifier + modified								
円形	6.6	5.50	怪獣	6.8	5.59	重病	6.9	5.66
短歌	6.9	5.69	良心	6.1	5.78	赤飯	6.6	5.94
町内	6.5	6.00	表紙	6.7	6.09	歩道	6.8	6.22
住所	6.3	6.44	聴力	6.8	5.53	旧友	7.0	5.63
西暦	6.6	5.66	美味	6.9	5.69	童話	6.6	5.84
草原	6.5	5.94	白鳥	7.0	6.03	薬品	5.5	6.13
空港	6.4	6.34	来年	6.7	6.50	羊毛	6.7	5.53
定時	6.8	5.63	借家	6.3	5.69	祭日	6.6	5.78
花粉	6.4	5.94	次男	6.9	5.97	多数	6.6	6.06
海岸	6.4	6.22	今月	6.6	6.34	牛乳	6.1	6.56
Verb + complement								
採血	6.8	5.50	敬老	6.9	5.53	起床	6.6	5.72
変色	6.6	5.81	投球	6.8	5.91	乗車	6.9	6.03
飲酒	7.0	6.09	挑戦	6.8	6.09	失恋	6.8	6.13
登場	6.3	6.19	産卵	6.7	5.50	防寒	6.9	5.53
入館	6.9	5.75	改善	5.5	5.84	握手	6.4	5.97
進学	6.2	6.03	化石	5.5	6.09	犯罪	6.4	6.09
消火	6.8	6.16	預金	6.6	6.28	避暑	6.5	5.50
落馬	6.2	5.59	保温	6.7	5.78	洗顔	7.0	5.88
競走	5.6	6.03	撮影	6.7	6.06	帰国	6.7	6.06
開店	7.0	6.13	有効	6.2	6.16	求人	6.9	6.34

(table continues)

	CS	FS		CS	FS		CS	FS
Complement + verb								
再選	6.9	5.50	陰謀	6.4	5.59	強打	7.0	5.69
音響	5.6	5.75	清掃	5.9	5.84	公認	6.8	5.97
歓迎	6.2	6.03	誤解	5.8	6.13	早退	7.0	6.22
独立	6.1	6.28	軽視	7.0	5.50	予備	6.8	5.66
激励	6.5	5.69	下記	6.3	5.75	全勝	6.5	5.91
夜勤	6.9	5.97	暗殺	6.4	6.06	外食	6.5	6.16
半分	6.1	6.25	広告	6.4	6.38	木造	6.0	5.50
未知	6.8	5.66	民営	6.9	5.69	自炊	6.4	5.81
先着	6.7	5.94	右折	6.9	6.03	冬眠	6.8	6.06
共通	6.0	6.22	水泳	6.0	6.28	旅行	6.7	6.47
Synonymous pairs								
衣装	6.4	5.63	彫刻	6.2	5.66	取得	6.4	5.72
切断	6.3	5.75	減少	6.5	5.81	忍耐	6.1	5.84
救助	6.8	5.94	優秀	6.5	6.13	移動	5.8	6.25
使用	6.3	6.34	終止	5.9	5.63	出現	6.3	5.69
庭園	6.1	5.72	崩壊	6.3	5.75	悲惨	5.7	5.81
困難	6.3	5.88	攻撃	5.8	6.06	建設	6.1	6.16
募集	6.2	6.28	疑問	5.7	6.38	衰弱	6.2	5.63
運送	6.3	5.72	願望	6.4	5.75	削除	6.0	5.78
思考	5.8	5.84	貯蓄	6.7	5.88	森林	5.5	6.09
永久	6.8	6.22	新鮮	6.6	6.31	死亡	6.6	6.38

Note. CS = classification score, which from Joyce and Ohta (1999). FS = familiarity score, which is from the NTT database (Amano & Kondō, 1999).

実験の教示

- ✧ この実験は、単語の認知に関するものです。
- ✧ 実験では、瞬間的にコンピューター画面が変わっていきます。その順番は、まず、最初に画面の真中に「＋」という記号が出てきます。次に、漢字 1 字が出てくる場合と、白い画面が出てくる場合があります。そして、最後に「※」という記号が呈示されますので、それらを見ていて下さい。その後すぐ、画面の真中に漢字 2 字熟語が呈示されます。
- ✧ この実験で行って頂きたいことは、なるべく速く正確に、その漢字 2 字熟語をマイクの方に向かって、声を出して読むことです。
- ✧ この漢字 2 字熟語を読んでいただいた後、レスポンスボックスのキーを押して下さい。キーは、どれでもかまいません。キーが押されると、次の試行が開始されます。前に読んだ漢字 2 字熟語は気にせずに、次の試行に進んで下さい。
- ✧ 本実験に入る前に、実験の流れ（画面の変化やレスポンスボックスのキーを押すことなど）に慣れて頂くために、練習セッションを 12 回行います。その後、本実験に入ります。
- ✧ この教示に対してご質問がありましたら、気軽に聞いて下さい。

	CS	FS		CS	FS		CS	FS
Modifier + modified								
異色	6.2	5.53	屋上	6.2	6.03	温泉	6.4	6.59
花粉	6.4	5.94	怪獣	6.8	5.59	旧友	7.0	5.62
空港	6.4	6.34	軽傷	7.0	5.66	今月	6.6	6.34
祭日	6.6	5.78	次男	6.9	5.97	住民	6.8	6.28
瞬間	6.6	6.22	商品	6.7	6.5	西暦	6.6	5.66
赤飯	6.6	5.94	鮮魚	6.7	5.56	前例	6.6	5.50
早朝	6.7	6.16	体臭	6.6	5.62	太鼓	6.4	5.50
定時	6.8	5.62	同様	6.7	5.87	白鳥	7.0	6.03
歩道	6.8	6.22	宝石	6.5	6.19	毛布	6.4	6.12
楽園	6.7	5.97	良心	6.1	5.78	老眼	6.7	5.84
Verb + complement								
握手	6.4	5.97	飲酒	7.0	6.09	加速	6.8	5.91
改装	6.8	5.78	開店	7.0	6.12	看病	6.9	5.72
起床	6.6	5.72	帰国	6.7	6.09	休学	7.0	5.69
献血	6.8	6.19	殺人	6.7	6.25	撮影	6.7	6.06
失恋	6.8	6.12	乗馬	7.0	5.75	整形	6.9	5.69
絶望	6.5	5.97	洗顔	7.0	5.87	耐震	6.2	5.53
駐車	6.7	6.00	投球	6.8	5.91	入場	6.8	6.28
犯罪	6.4	6.09	噴火	6.3	5.72	閉館	6.9	5.84
返答	6.3	5.62	変身	6.6	5.84	保健	6.6	6.06
補欠	6.7	5.84	防災	6.9	5.91	預金	6.4	5.59

(table continues)

	CS	FS		CS	FS		CS	FS
Complement + verb								
陰謀	6.4	5.59	右折	6.9	6.03	横断	6.5	5.97
下記	6.3	5.75	仮説	6.4	5.59	外出	6.8	6.16
歓迎	6.2	6.03	偽造	6.6	5.66	急増	7.0	5.62
共通	6.0	6.22	強打	7.0	5.69	激励	6.5	5.69
厳守	7.0	5.84	広告	6.4	6.37	公認	6.8	5.97
再会	6.9	6.06	私立	6.1	6.06	自炊	6.4	5.81
重視	7.0	5.69	水泳	6.0	6.28	専用	6.2	6.19
全勝	6.5	5.90	直進	6.9	5.91	冬眠	6.8	6.06
独走	6.9	5.81	半熟	6.2	5.50	必要	6.3	6.50
未知	6.8	5.66	夜勤	6.9	5.97	予想	6.4	6.28
Synonymous pairs								
運送	6.3	5.72	永久	6.8	6.22	回転	6.6	6.03
獲得	6.4	5.78	祈願	6.4	5.50	恐怖	6.3	6.12
救助	6.8	5.94	苦痛	6.00	5.69	建設	6.1	6.16
攻撃	5.8	6.06	困難	6.3	5.87	削除	6.0	5.78
死亡	6.6	6.37	飼育	6.1	5.94	児童	6.7	6.00
消滅	6.2	5.62	衰弱	6.2	5.62	清潔	6.3	5.84
戦争	6.6	6.56	尊敬	6.6	6.06	端末	6.4	5.94
貯蓄	6.7	5.87	超過	6.4	5.53	彫刻	6.2	5.66
到着	6.4	5.94	燃焼	6.8	5.78	破壊	6.4	6.03
分裂	6.4	5.87	募集	6.2	6.28	優秀	6.5	6.12

Note. CS = classification score, which from Joyce and Ohta (1999). FS = familiarity score, which is from the NTT database (Amano & Kondō, 1999).

実験の教示

✧ この実験は、単語の認知に関するものです。

✧ 実験では、今から説明するように、瞬間的にコンピューター画面が変わっていきます。その順番は、まず、最初に画面の真中に「＋＋」という記号が出てきます。次に、仮名文字が出てくる場合と、白い画面が出てくる場合があります。そして、最後に「※※※」という記号が呈示されますので、それらは見ていて下さい。その後すぐ、真中に二つの漢字が呈示されます。この実験でして頂きたいことは、ここで呈示された漢字二字が意味をもつ、本当の単語であるかどうかを判断して、なるべく速く正確にレスポンスボックスのキーを押すことです。

✧ もし、この漢字二字が本当の単語であると思えば、「緑」のキーを押して下さい。また、もし、この漢字二字が本当の単語ではないと思えば、「赤」のキーを押して下さい。レスポンスボックスのキーを押したら、次の項目が現われるので、前に行った判断は気にせずに、次の項目に進んで下さい。

✧ 本実験に入る前に、漢字について判断し、レスポンスボックスのキーを押すことに慣れて頂くために、練習セッションを 10 回行います。その後、本実験に入ります。また、実験中には、休憩を約 1 分間はさみます。つまり、実験ー休憩ー実験という流れです。

✧ この教示に対してご質問がありましたら、気軽に聞いて下さい。

	CS	FS		CS	FS		CS	FS
Modifier + modified								
運河	6.6	5.78	花粉	6.4	5.94	怪獣	6.8	5.59
牛乳	6.1	6.56	強力	6.9	6.00	空港	6.4	6.34
軽油	6.8	5.66	今月	6.6	6.34	祭日	6.6	5.78
次男	6.9	5.97	借家	6.3	5.69	住所	6.3	6.44
重病	6.9	5.66	商品	6.7	6.50	新米	6.5	5.66
西暦	6.6	5.66	草原	6.5	5.94	多数	6.6	6.06
中年	6.4	5.90	町民	6.7	5.59	定時	6.8	5.62
同様	6.7	5.87	童話	6.6	5.84	白鳥	7.0	6.03
美味	6.9	5.69	表紙	6.7	6.09	歩道	6.8	6.22
旅館	6.5	6.12	良心	6.1	5.78	隣人	6.8	5.62
Verb + complement								
握手	6.4	5.97	飲酒	7.0	6.09	化石	5.5	6.09
加速	6.9	5.50	改装	6.8	5.78	開店	7.0	6.12
帰国	6.7	6.09	起床	6.6	5.72	休学	7.0	6.89
救命	6.6	5.53	求人	6.9	6.34	競走	5.6	6.03
撮影	6.7	6.06	産卵	6.7	5.50	失恋	6.8	6.12
消火	6.8	6.16	乗車	6.9	6.03	洗顔	7.0	5.87
挑戦	6.8	6.09	登場	6.3	6.19	投球	6.8	5.90
入館	6.9	5.75	犯罪	6.4	6.09	被災	6.9	5.90
避暑	6.5	5.50	変色	6.6	5.81	防水	7.0	5.81
有効	6.2	6.16	預金	6.6	6.28	落馬	6.2	5.59

(table continues)

	CS	FS		CS	FS		CS	FS
Complement + verb								
暗殺	6.4	6.06	陰謀	6.4	5.59	音響	5.6	5.75
下記	6.3	5.75	外食	6.5	6.16	歓迎	7.0	6.03
軽視	7.0	5.50	激励	6.5	5.69	公認	6.8	5.97
広告	6.4	6.37	再選	6.9	5.5	私有	6.8	5.66
自炊	6.4	5.81	手話	5.8	5.87	宿泊	5.5	5.90
水泳	6.0	6.28	正解	5.6	6.31	清掃	5.9	5.84
先着	6.7	5.94	専用	6.2	6.19	前進	6.9	5.94
早退	7.0	6.22	冬眠	6.8	6.06	独立	6.1	6.28
半分	6.1	6.25	未知	6.8	5.66	民営	6.9	5.69
木造	6.0	5.50	夜勤	6.9	5.97	予想	6.4	6.28
Synonymous pairs								
疑問	5.7	6.37	使用	6.3	6.34	募集	6.2	6.28
移動	5.8	6.25	学習	5.8	6.22	競争	6.0	6.19
建設	6.1	6.16	悲惨	5.7	5.81	到着	6.4	5.94
救助	6.8	5.94	忍耐	6.1	5.84	思考	5.8	5.84
庭園	6.1	5.72	取得	6.4	5.72	運送	6.3	5.72
保守	5.7	5.66	彫刻	6.2	5.66	衣装	6.4	5.62
衰弱	6.2	5.62	祈願	6.4	5.50	新鮮	6.6	6.31
攻撃	5.8	6.06	飼育	6.1	5.94	貯蓄	6.7	5.87
切断	6.3	5.75	永久	6.8	6.22	変化	6.2	6.19
行進	6.2	5.87	削除	6.0	5.78	損失	6.3	5.50

Note. CS = classification score, which from Joyce and Ohta (1999). FS = familiarity score, which is from the NTT database (Amano & Kondō, 1999).

実験の教示

- ✧ この実験は、単語の認知に関するものです。
- ✧ 実験では、今から説明するように、瞬間的にコンピューター画面が変わっていきます。その順番は、まず、最初に画面の真中に「++」という記号が出てきます。次に、仮名文字が出てくる場合と、白い画面が出てくる場合があります。そして、最後に「※※※※※」という記号が呈示されますので、それらは見ていて下さい。その後すぐ、真中に二つの漢字が呈示されます。この実験でして頂きたいことは、ここで呈示された漢字二字が意味をもつ、本当の単語であるかどうかを判断して、なるべく速く正確にレスポンスボックスのキーを押すことです。
- ✧ もし、この漢字二字が本当の単語であると思えば、「緑」のキーを押して下さい。また、もし、この漢字二字が本当の単語ではないと思えば、「赤」のキーを押して下さい。レスポンスボックスのキーを押したら、次の項目が現われるので、前に行った判断は気にせずに、次の項目に進んで下さい。
- ✧ 本実験に入る前に、漢字について判断し、レスポンスボックスのキーを押すことに慣れて頂くために、練習セッションを 10 回行います。その後、本実験に入ります。また、実験中には、休憩を約 1 分間はさみます。つまり、実験ー休憩ー実験という流れです。
- ✧ この教示に対してご質問がありましたら、気軽に聞いて下さい

実験の教示

- ✧ この実験は、単語の認知に関するものです。
- ✧ 実験では、最初にコンピューター画面の真中に「++」という記号が出てきます。少し後れて、ヘッドホンから音が再生されます。「++」の記号を見ながら、その音を聞いてください。ただし、音は再生されない場合もあります。「++」の記号が消えると、画面の真中に二つの漢字が呈示されます。この実験でして頂きたいことは、ここで呈示された漢字二字が意味をもつ、本当の単語であるかどうかを判断して、なるべく速く正確にレスポンスボックスのキーを押すことです。
- ✧ もし、この漢字二字が本当の単語であると思えば、「緑」のキーを押して下さい。また、もし、この漢字二字が本当の単語ではないと思えば、「赤」のキーを押して下さい。レスポンスボックスのキーを押したら、次の項目が現われるので、前に行った判断は気にせずに、次の項目に進んで下さい。
- ✧ 本実験に入る前に、漢字について判断し、レスポンスボックスのキーを押すことに慣れて頂くために、練習セッションを 10 回行います。その後、本実験に入ります。また、実験中には、休憩を約 1 分間はさみます。つまり、実験ー休憩ー実験という流れです。
- ✧ この教示に対してご質問がありましたら、気軽に聞いて下さい。

実験の教示

- ✧ この実験は、単語の認知に関するものです。
- ✧ 実験では、最初にコンピュータ画面の真中に「＋＋」という記号が出てきます。少し後れて、ヘッドホンから単語が聞こえます。「＋＋」の記号を見ながら、それを聞いてください。ただし、単語は聞こえない場合もあります。「＋＋」の記号が消えると、画面の真中に二つの漢字が呈示されます。この実験でして頂きたいことは、ここで呈示された漢字二字が意味をもつ、本当の単語（熟語）であるかどうかを判断して、なるべく速く正確にレスポンスボックスのキーを押すことです。
- ✧ もし、この漢字二字が本当の単語であると思えば、「緑」のキーを押して下さい。また、もし、この漢字二字が本当の単語ではないと思えば、「赤」のキーを押して下さい。レスポンスボックスのキーを押したら、次の項目が現われるので、前に行った判断は気にせずに、次の項目に進んで下さい。
- ✧ 本実験に入る前に、漢字について判断し、レスポンスボックスのキーを押すことに慣れて頂くために、練習セッションを 10 回行います。その後、本実験に入ります。また、実験中には、休憩を約 1 分間はさみます。つまり、実験－休憩－実験という流れです。
- ✧ この教示に対してご質問がありましたら、気軽に聞いて下さい。

	CS	FS		CS	FS		CS	FS
LPR-Verb + Complement								
送信	6.6	5.63	求愛	6.7	5.69	換気	6.5	5.47
化石	5.5	6.09	刻印	6.4	5.00	迎春	6.8	5.50
定義	6.2	5.78	習字	6.2	5.97	止血	7.0	5.50
降雪	5.9	5.19	造園	6.5	5.28	録音	6.4	6.06
点火	6.8	5.78	還元	5.7	5.56	待機	6.1	5.41
断食	6.6	5.59	始業	6.3	5.41	鎖国	6.5	5.50
HPR-Verb + Complement								
提案	6.7	5.88	免罪	6.6	5.13	登校	6.2	5.94
改札	5.8	5.94	預金	6.6	6.28	殉職	6.5	5.03
創刊	6.4	5.66	洗髪	6.7	5.53	救命	6.6	5.53
処罰	6.6	5.44	遭難	6.1	5.53	移籍	6.6	5.53
停学	6.2	5.59	耐震	6.2	5.53	被災	6.9	5.50
返答	6.3	5.63	排水	6.5	5.66	克己	6.6	5.11
Complement + LPR-Verb								
墓参	6.1	5.25	再建	6.8	5.53	市販	6.3	5.41
皆無	5.5	5.50	特異	6.0	5.34	木製	6.4	5.63
日記	6.3	6.28	豆腐	5.5	6.34	車検	6.1	6.13
暴飲	6.7	5.28	順延	6.1	5.09	空輸	6.3	5.34
急増	7.0	5.63	密閉	5.9	5.53	病欠	6.1	5.59
完投	6.7	5.56	複写	5.6	5.53	脈拍	5.7	5.41
Complement + HPR-Verb								
激励	6.3	5.69	予備	6.8	5.66	暗算	6.5	5.69
外泊	6.8	5.84	合唱	6.6	5.81	満載	5.8	5.44
共済	5.6	5.22	既婚	6.4	5.63	反響	5.5	5.66
銃撃	6.8	5.34	自滅	6.9	5.75	和訳	6.2	5.25
主催	6.3	5.72	永眠	6.3	5.69	伴奏	6.2	5.47
専務	6.5	5.66	全壊	6.8	5.69	公用	5.8	5.19

実験の教示

✧ この実験は、単語の認知に関するものです。

✧ 実験では、瞬間的にコンピューター画面が変わっていきます。その順番は、まず、最初に画面の真中に「＋＋」という記号が出てきます。次に、漢字 2 字熟語が出てくる場合と、白い画面が出てくる場合があります。そして、最後に「※※」という記号が呈示されますので、それらを見ていて下さい。その後すぐ、画面の真中に漢字 1 字か漢字仮名まじりの単語が呈示されます。

✧ この実験で行って頂きたいことは、なるべく速く正確に、その漢字 1 字か漢字仮名まじりの単語をマイクの方に向かって、声を出して訓読みで読むことです。

✧ この単語を読んでいただいた後、レスポンスボックスのキーを押して下さい。キーは、どれでもかまいません。キーが押されると、次の試行が開始されます。前に読んだ単語は気にせずに、次の試行に進んで下さい。

✧ 本実験に入る前に、実験の流れ（画面の変化やレスポンスボックスのキーを押すことなど）に慣れて頂くために、練習セッションを 12 回行います。その後、本実験に入ります。

✧ この教示に対してご質問がありましたら、気軽に聞いて下さい。