The mental representations of Hanja: Exploring cross-script semantic cohorts in Korean



Yoolim Kim

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2019

저를 키워주신 안양 할머니에게 이 논문을 바칩니다.
보고 싶습니다.

Acknowledgements

This truly took a village. Every single person who came by this village, whether for just 3 minutes or for 3 years, I thank each and every one of you tremendously.

I would like to first thank my supervisors, Aditi Lahiri and Sandra Kotzor, for their unreserved support and patience. At numerous times during this journey, I had the thought that *this* moment would be the one to break the camel's back, but each time, both remained unwavering in their support, making such a thought a fleeting one. I would like to express additional gratitude to Aditi, who, during the MPhil, found me in the CLP Common Room, in dire straits, reading Ladefoged, trying to figure out how to measure VOT. Had it not been for her commitment and dedication at the eleventh hour, all those years ago, I would not be writing these acknowledgements now. I owe a great deal of my academic success at Oxford to the opportunities that have been made possible through her support, from answering every last question and concern to always finding the time to meet with me, even if it meant at 6pm on a Friday. I thank Aditi for always seeing the bigger picture and for throwing a lifeline when it was most needed, and Sandra, for everything in between. I would also like to thank my examiners, Nina Kazanina and Wolfgang de Melo, for taking the time to read and help improve this thesis.

I have also had the great fortune of meeting a number of people along the way, whose friendships have made what would have otherwise been a very bleak journey, a colorful one: Alicia Cleary-Venables, Jamie Findlay, Kim Fuellenbach, Rosy Hall, Henry Hawthorne, Stephen Jones, Eleanor Ridge, Kate Sim, Ana Werkmann Horvat, and Cong Zhang.

And of course, a very heartfelt thanks to Madeline LaRue, who has been reading my writing since school. May this be the last dissertation of mine you ever have to read.

I would also like to thank all members, past and present, of the Language & Brain Laboratory: Colin Brooks, Emily Darley, Zoë Fletcher, Holly Kennard, Emily Lindsay-Smith, Yaxuan Meng, Ben Mollineaux, Adam Roberts, Swetlana Schuster, Jelena Sučević, Johanneke Sytsema, Allison Wetterlin, and Hilary Wynne, Rebecca Zeng, Beinan Zhou, all of whom I feel very privileged to call colleagues and friends.

And most importantly, I thank all of the participants without whom quite literally none of this would have been possible. While it might have been a hour or so out of your day, this research was the livelihood of my existence at Oxford; the experiences I gained, the lessons I learned, all of it is the result of you sharing a small part of your day with me. In particular, I thank the students of Sookmyung Women's University, Hansung University, and Kyung Hee University. I would also like to thank my father, who used up most of his sabbatical to support me during my fieldwork, and by support, I mean, he patiently stood outside the experiment room, quietly shushing students passing by in the corridors to ensure no disruptions to my research. The members of Hanmaum Seon, who were always willing to help in any way possible, also deserve my deep gratitude.

I would also like to thank Hanna Koo, who so kindly let me spend a day traipsing around her office, taking volunteers from her colleagues to participate in my study. Hanna has been making my research possible since the early days of Wellesley, where her ability to recruit participants was first noticed and remain unrivalled to this day. I started with fifteen Wellesley friends then, who graciously gave up their time, and here we are now, many, many participants later. Thank you all.

It was also at Wellesley where I first started to study Linguistics, and where I met my first mentor, Angela Carpenter, who remains one of my most active cheerleaders to this day. I am indebted to her support and belief in me, which enabled the move across the pond in the first place. I am forever grateful that she signed off on my form where I declared Linguistics as my major.

And last but not least I thank the members of the Oxford church, who remain some of the kindest people I have ever met, not only for taking time to partake in my research, but also for taking the time to get to know me, not just as a researcher, but as a person. I especially thank William Do, Grace Jeon, and the Jeon family for their immeasurable kindness. I also thank Jaehoo Park for showing interest in my research and for welcoming me into the Oxford Korean Academic Society.

Finally, this research would not have been possible without the generous support of the Clarendon Scholarship, the Basant Kumar and Sarala Birla Graduate Studentship, and the Scatcherd European Scholarship, which funded an invaluable research visit to the Max Planck Institute for Psycholinguistics. This visit could not have been better timed, and as a result of it, I was able to return to Oxford with a renewed sense of self and independence. I thank my hosts, Antje Mejer and Falk Huettig, as well as the members of the Psychology of Language Department,

for their time and support, particularly, Saoradh Favier and Eirini Zormpa, who were so incredibly kind and welcoming. I also thank Laurel Brehm for her statistical support and for teaching me the importance of paying it forward.

In many ways, the actual thesis is the least important outcome of this whole journey. (Let's just say a convenient but necessary outcome.) I feel most grateful that as a result of this thesis, I was able to forge the relationships I did, and in the process, surround myself with a village of people who always saw the light at the end of the tunnel when I could not. In particular, I thank Louise Mycock, Chi Lun Pang, Kerstin Hoge, and Wolfgang de Melo for their unmatched support and generosity. I have been the privileged recipient of their guidance and mentorship. They make writing these acknowledgements all the more gratifying.

늘 부족한 저에게 항상 밝고 아름다운 순간들을 보여 주려 하셨고, 힘들 때마다 따뜻한 격려의 말 아끼지 않으셨던, 저에게는 너무도 특별하고 소중한 부모님께 감사의 마음을 전합니다.

이제 시작입니다. 앞으로 더 많은 아름다운 날들을 보여드리겠습니다.

Abstract

More than half of the Korean lexicon comprises words of Chinese origin. Sino-Korean words, unlike native Korean words, can be written two different ways, using either the native alphabetic script, Hangul, or logographic Chinese characters (called Hanja in Korean). While all Korean words, whether native or Sino Korean, can be written using Hangul, only the latter can additionally be written using Hanja. This means Sino-Korean words have two separate orthographic representations which are crucially equivalent to one other. Previous studies have shown that the mental representations of words take into account such lexical information as the morphology and the phonology as well as the orthography. An open question concerns languages, such as Korean, that have two structurally distinct scripts comprising the writing system, and how the mental lexicon represents the contributions of different scripts. A further complication arises with Korean, as it is often the case that there is more than one Hanja orthographic representation which is equivalent to the Hangul representation. The extensive use of Hanja subsequently resulted in numerous loans into the Korean lexicon, effectively creating two sub-lexicons, one which is native to the language (native Korean), and one which is not (Sino-Korean). This thesis tested whether native Korean speakers' mental representations of Sino-Korean words encoded the semantics of different Hanja characters, and whether the processing of Sino-Korean was modulated by the size of the semantic cohort. Through a series of four behavioral studies, the status of Hanja characters in the Korean mental lexicon was investigated, and the findings confirmed that lexical processing in Sino-Korean involved the active processing of the semantic contributions of Hanja characters. Furthermore, the data demonstrated that Sino-Korean processing involves semantic access to multiple meanings, as encoded by different Hanja characters. This shows that semantic access is possible across two different scripts within one writing system, and that for Sino-Korean specifically, such semantic access is mediated by the shared representation in Hangul. For each individual Sino-Korean morpheme, a cross-script cohort is generated, comprising the different meanings derived from the semantic contributions of all of the possible Hanja characters that share that particular Hangul representation. In auditory processing of Sino-Korean in particular, lexical access was significantly modulated by the size of this cross-script semantic cohort, with faster response times associated with larger cohort sizes. While these cohorts theoretically exist for all Sino-Korean morphemes, the collective findings across all four experiments seemed to suggest that access to all available meanings was not uniform for all morphemes, highlighting the

potentially variable status of individual Sino-Korean morphemes in the Korean mental lexicon. As a result of this empirical investigation, a revised model for the structure of the Korean mental lexicon is proposed, one which accounts for cross-script semantic cohorts, allowing for a more conceptual representation of the contributions of Hanja characters in the Sino-Korean lexicon.

Contents

List of Figures **xi**

List of Tables **xvi**

1 Introduction **1**
- 1.1 Contribution to the field . . . 3
- 1.2 Outline of thesis . . . 4

2 Writing systems, the innovation of Hangul, and the Korean lexicon **5**
- 2.1 Scripts and writing systems . . . 6
- 2.2 The Korean writing system . . . 9
  - 2.2.1 The use and adaptation of Hanja in Korean . . . 10
  - 2.2.2 The history of Hangul and the sounds of Korean . . . 12
  - 2.2.3 The geometry of Hangul . . . 16
- 2.3 The Korean lexicon . . . 23
  - 2.3.1 The Sino-Korean lexicon . . . 24
  - 2.3.2 Encoding multiple meanings in Sino-Korean . . . 27

3 Non-Roman script processing, cohorts, and the bi-scriptal mental lexicon **30**
- 3.1 Cohorts in lexical processing . . . 31
- 3.2 Semantic processing . . . 37
  - 3.2.1 Effects of semantic transparency on processing . . . 37
  - 3.2.2 Spreading activation and semantic network theory . . . 40
- 3.3 Multiple semantic access . . . 44
  - 3.3.1 Initial meaning access . . . 45
  - 3.3.2 Initial meaning selection . . . 47
  - 3.3.3 Semantic competition . . . 48
- 3.4 Processing in a bi-scriptal lexicon . . . 50
- 3.5 Processing of non-Roman scripts: Chinese, Japanese, and Korean . . . 52
  - 3.5.1 Semantic processing of Chinese characters . . . 52
  - 3.5.2 Script processing in Japanese . . . 59
  - 3.5.3 Processing Hanja and Hangul . . . 68
- 3.6 Current models of the Korean mental lexicon . . . 72
- 3.7 Main research objectives . . . 75

4 Status of Hanja in the mental lexicon 76
4.1 Introduction . . . 76
4.2 Cross-script semantic representations . . . 77
4.2.1 A cohort across scripts: Sino-Korean . . . 82
4.2.2 Experimental paradigm and priming . . . 87
4.3 Experiment 1: Visual Hanja fragment priming . . . 92
4.3.1 Methods . . . 92
4.3.1.1 Materials and design . . . 92
4.3.1.2 Participants . . . 93
4.3.1.3 Procedure . . . 94
4.3.2 Results . . . 95
4.3.3 Discussion . . . 95
4.3.4 Methodological considerations . . . 96
4.4 Experiment 2: Visual Sino-Korean full-word priming . . . 98
4.4.1 The effects of varying cross-script semantic cohort size . . . 99
4.4.2 Methods . . . 100
4.4.2.1 Materials and design . . . 100
4.4.2.2 Participants . . . 102
4.4.2.3 Procedure . . . 103
4.4.3 Results . . . 103
4.4.4 Discussion . . . 106

5 Conceptual encoding of Hanja 109
5.1 Introduction . . . 109
5.2 A cohort of cross-script meanings . . . 110
5.3 Experiment 3: Auditory Hanja fragment priming . . . 119
5.3.1 Methods . . . 120
5.3.1.1 Materials and design . . . 120
5.3.1.2 Participants . . . 121
5.3.1.3 Procedure . . . 121
5.3.2 Results . . . 122
5.3.3 Discussion . . . 123
5.4 Experiment 4: Auditory Sino-Korean full-word priming . . . 124
5.4.1 Methods . . . 125
5.4.1.1 Materials and design . . . 125
5.4.1.2 Participants . . . 126
5.4.1.3 Procedure . . . 127
5.4.2 Results . . . 127

5.4.3 Discussion 132
5.4.3.1 Effects of modality on cross-script cohorts 132
5.4.3.2 Beyond phonological cohorts 135
5.4.3.3 The variable status of Hanja in Sino-Korean 138

6 The state of the Korean mental lexicon 140
6.1 Introduction 140
6.2 Investigating by-item variance 141
6.2.1 Visual Sino-Korean full-word priming 143
6.2.2 Auditory Sino-Korean full-word priming 146
6.2.3 Effects of cohort with no semantic priming 148
6.2.4 Point of saturation in cohort size 151
6.2.4.1 Semantic cohort size of ten meanings or more . . . 151
6.2.4.2 Semantic cohort size of ten meanings or fewer . . . 151
6.2.4.3 Semantic cohort size of five meanings or fewer . . . 153
6.3 Cross-script semantic cohorts in the Korean mental lexicon 156

7 Conclusion 167
7.1 Summary of the thesis 168
7.2 Methodological considerations 170
7.3 Cohort activation and asymmetric representations 171
7.4 Expanding the revised model of the Korean mental lexicon 176

Appendices
A Pre-experiment information-gap activity 181
B Stimulus set for Experiment 1 (Chapter 4) 182
C Stimulus set for Experiment 2 (Chapter 4) 183
D Subset of targets primed in Experiment 2 184
E Stimulus set for Experiment 4 (Chapter 5) 185
F Subset of targets primed in Experiment 4 186

Bibliography 187

List of Figures

1.1 The punchline is provided in Korean Hangul, along with the corresponding Chinese characters. 2

2.1 The main vowels based on three main elements: *heaven*, *earth*, and *man*. 16
2.2 The fourteen simple consonants and their phonetic transcriptions (Sohn, 1994). The nine visually prototypic consonants are within the blue box. 16
2.3 The additional horizontal stroke represents the phonological feature of aspiration, creating two phonologically contrastive characters, ㄱ /k/ and ㅋ /k^h/. 17
2.4 The phonological encoding of aspiration via an additional stroke can be observed in other contrastive pairs. 17
2.5 The shape of the mouth and the lips and the specific movements involved in the production of ㅁ /m/, ㅍ /p^h/, and ㅂ /p/ (Park, 2015: 57). 18
2.6 The shape of the mouth and the lips and the specific movements involved in the production of ㄴ /n/, ㅌ /t^h/, and ㄷ /t/ (Park, 2015: 58). 19
2.7 The shape of the mouth and the lips and the specific movements involved in the production of ㄱ /k/ and ㅋ /k^h/ (Park, 2015: 59). . 19
2.8 The shape of the mouth and the lips and the specific movements involved in the production of ㅅ /s/, ㅈ /ts/, and ㅊ /ts^h/ (Park, 2015: 60). 20
2.9 The shape of the mouth and the lips and the specific movements involved in the production of ㅇ /ŋ/ and ㅎ /h/ (Park, 2015: 61). . . 20
2.10 The ten simple vowels and their phonetic transcriptions (Sohn 1994). 21
2.11 Semi-vowels and diphthongs, and their phonetic transcriptions (Sohn, 1994) 21
2.12 Korean is composed by way of non-linear assembly of consonant and vowel symbols, as demonstrated by 눈. 22
2.13 The six main types of syllabic blocks in Korean and their relative frequencies; (a) and (b) show the vowel ㅏ[a], with (a) and without (d) a final C ; (b) and (e) the vowel ㅗ [o], with (b) and without (e) a final C; (c) and (f) demonstrate use of both vowels ㅏ and ㅗ, with (c) and without (f) a final C. 22

2.14 The two syllable types in the Korean lexicon and the structural composition of individual syllables. In both native (A) and non-native (i.e., Sino-Korean) (B) syllables, the Hangul characters are assembled non-linearly into blocks, but Sino-Korean syllables can also be rendered one step further using Hanja characters. 23
2.15 The same phrase rendered in both Hanja and Hangul 24
2.16 The semantic mapping of [hak] from morpheme to full word. Identical homographs in Korean map onto different Hanja characters; however, although identical, each homograph of [hak] encodes a different meaning, and each lexical instantiation is semantically related to one of corresponding Hanja characters presented on the left (by color). . 28

3.1 For each time point during which a new phoneme becomes available, a new shortlist (gray box) is created. Lexical candidates that overlap with one another at any position compete with one another. For *sun*, this would mean that all candidates inhibit one another (Weber and Scharenborg, 2012). 34
3.2 A collection of concepts operationalized to capture a word meaning and corresponding links for the word *bird* (Traxler, 2012: 83). . . . 41
3.3 The architecture proposed by Taft et al. (1999: 95) and Zhu and Taft (1994) to account for morphemic processing in Chinese (e.g., 現代, meaning 'modern'), involving multiple pathways which interact with one another at the orthograpahic, phonological and the semantic levels. 55
3.4 Revised model of the interactive-activation model which includes a level for lemmas, eliminating orthographic and phonological representations of complete compound words (Taft, Liu, & Zhu, 1999: 107). 57
3.5 A model of the Japanese mental lexicon, based on connections mediated by lemma-unit representations (Joyce & Ohta, 1999; Joyce, 1999; 2002: 89), which shows the representation of two words which are synonymous for 'mount climbing', the Sino-Japanese 登山 [to.zan] and the native Japanese 山登り [jama.no.borji]. 65
3.6 A model of the Korean mental lexicon for the representation and processing of Sino-Korean words, as proposed by Yi (2009: 407). . . 73

4.1 The meanings of the different Hanja characters, all of which approximate the same phonological sound (discounting the tonal value), represented as 안 in Hangul, which can partner with other syllables to create lexical items that relate semantically to one of the meanings encoded by the relevant Hanja logogram. 82

4.2 Semantic access across scripts, as mediated by the shared representation in Hangul, activates a starting cohort, which also activates an extended cohort containing all Sino-Korean words that contain that particular homograph, and thus the specific semantics of the appropriate Hanja character. 84
4.3 Two branches of the schema representing the many meanings of 안. 84
4.4 Potential semantic access across scripts, as mediated by the shared representation in Hangul. 86
4.5 Overview of test stimuli. 93
4.6 A road sign in Seoul, which displays the street name (*Itaewon*) three different ways – in Romanized Korean (*Itaewon-ro 19-gil*), Hangul (*이태원로 19 길*, offset in yellow), and Hanja (underlined in red). . . 97
4.7 A cover of a book titled *April 23rd*, with the Hanja characters used to express *month* and *day* (underlined in red). 97
4.8 Much of Korean television is 'annotated' with additional commentary, presented in Hangul (offset in yellow), and very often featuring Hanja characters (underlined in red; n.b., this text is not to be mistaken for closed captioning. Unlike closed captioning, in which speech is simultaneously presented onscreen in written form, this particular form of captioning, which has been described as *graphic overtitles* (Park, 2009), does not reiterate what is being said onscreen. Rather, such captioning tends to express unspoken voices, such as the internal thoughts of a character, comments by the producer, or a predicted response from the audience. Importantly, these unspoken voices tend to contrast with the spoken language and/or imagery that is being heard or observed onscreen). 97
4.9 The representations of [mi] and [hak] in Hangul can map onto a multitude of different Hanja characters, thereby encoding a number of different meanings, the relative number varying between [hak], which has fewer, and [mi], which has more. 99
4.10 Experimental conditions. Depending on the condition, the prime-target pair involved a direct or indirect semantic relationship that required mediated semantic access via shared homograph (e.g., *hak*). 102
4.11 Distribution of reaction times (RTs) for all three experimental conditions. 105

5.1 The set of possible meanings of *mi*. The representation of *mi* in Hangul can map onto a multitude of different Hanja characters, thereby encoding the number of different meanings. 112

5.2 The semantic mapping of *mi* from morpheme to full word. Identical homographs in Korean map onto different Hanja characters; however, although identical, each homograph of *mi* encodes a different meaning, and each lexical instantiation is semantically related to the corresponding Hanja character (associations by color). 112
5.3 The set of possible meanings of *hak*. 113
5.4 Semantic access mediated by shared homographs *hak* or *mi*. The shared homograph mediates semantic access to different semantic connections, as contributed by the different Hanja characters. Homographs – *hak* or *mi* – vary in the number of semantic connections, which may influence the viability of lexical access to semantic associates via mediated routes. 115
5.5 Overview of test stimuli. 120
5.6 The distribution of reaction times in both the Related and Unrelated conditions. 123
5.7 Distribution of reaction times (RTs) for the Related condition for Experiment 3, the blue line representing a fitted linear regression, and the shaded grey showing SE. An additional meaning or member that is added to the cohort does not affect reaction times. 124
5.8 Experimental conditions. Depending on the condition, the prime-target pair involved a direct or indirect semantic relationship that required mediated semantic access via shared homograph (e.g., *hak*). 126
5.9 Distribution of reaction times (RTs) for the Indirectly and Directly Related conditions, the blue line representing a fitted linear regression, and the shaded grey showing SE, for Experiment 4 (audio-to-visual cross-modal priming). Only in the Indirectly Related condition did targets with larger cohorts elicit significantly faster reaction times than those with smaller, with each additional member significantly decreasing response times. 129
5.10 Distribution of reaction times (RTs) for the Indirectly and Directly Related conditions in Experiment 4 (cross-modal), according to semantic salience of the presented meaning, and the relative number of other meanings with which the presented meaning is competing. The blue line represents a fitted linear regression, and the shaded grey shows SE. 130
5.11 Distribution of reaction times (RTs) for the Indirectly and Directly Related conditions, the blue line representing a fitted linear regression, and the shaded grey showing SE, for Experiment 2 (visual intra-modal). Based on our model, neither condition elicited significant effects of cohort. 133

6.1 Depending on the condition, the prime-target pair involved a direct or indirect semantic relationship that required mediated semantic access via shared homograph (e.g., *hak*). 142
6.2 Yi's (2009: 407) model does not make clear how items which do not overlap in form are activated, especially as his model specifically posits a morphemic level, and thus it remains unclear how semantically related lexical items would be accounted for, e.g., *cooking*, in his model. 159
6.3 A revised model of the Korean mental lexicon for the representation and processing of Sino-Korean words, as proposed by Yi (2009). . . 160
6.4 Proposed structure of the Korean mental lexicon based on the varying role of the semantic cohort in the Sino-Korean sub-lexicon. 163
6.5 The "swinging" within the semantic network between the morphemic lemma level and the conceptual level causes resonance at the conceptual level, the effects of which reverberate into the lexicon and are observed in priming effects between semantic associates. 165

7.1 Two different Sino-Korean morphemes with different cohort sizes on both the semantic and the featural level. 175

List of Tables

2.1 The different ways in which characters are phonologically rendered across Sino-Korean and Chinese. To note, all of the Sino-Korean Hanja characters presented in this table can also be rendered in Hangul (which is provided in the second column). 12
2.2 The consonantal inventory of Korean (Kim, 2015). 14
2.3 The vowel inventory of Korean (Sohn, 1999). To note, Sohn (1999) uses phonemic values to represent individual sounds. 15
2.4 The duality of the lexicon, with parallel forms in Native and Sino-Korean; the corresponding Hanja characters are provided alongside the Sino-Korean words. 26

4.1 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 104

5.1 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 122
5.2 Results of linea mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 128
5.3 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 132

6.1 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 144
6.2 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 144
6.3 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 147
6.4 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 148
6.5 Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level. 149
6.6 Results of linear mixed effects analysis for the intra-modal priming experiment (Experiment 2), containing only primes with a semantic cohort size of ten meanings or fewer. The asterisk indicates significant results at the 95% significance level. 152

6.7 Results of linear mixed effects analysis for the cross-modal fragment priming experiment (Experiment 3), containing only primes with a semantic cohort size of ten meanings or fewer. The asterisk indicates significant results at the 95% significance level. 153
6.8 Results of linear mixed effects analysis for the intra-modal priming experiment (Experiment 2), containing only primes with a semantic cohort size of ten meanings or fewer. The asterisk indicates significant results at the 95% significance level. 154
6.9 Results of linear mixed effects analysis for the cross-modal fragment priming experiment (Experiment 3), containing only primes with a semantic cohort size of five meanings or fewer. The asterisk indicates significant results at the 95% significance level. 154
6.10 Results of linear mixed effects analysis for the cross-modal full-word priming experiment (Experiment 4), containing only primes with a semantic cohort size of five meanings or fewer. The asterisk indicates significant results at the 95% significance level. 155

1 Introduction

Contents

1.1 Contribution to the field **3**
1.2 Outline of thesis **4**

No words seem better suited to begin a thesis about the Korean writing system, Hangul, at the University of Oxford, in the Faculty of Linguistics, Philology & Phonetics, than the following:

> "In the mid 1990s, the Linguistics, Philology & Phonetics Department of the University of Oxford, which is renowned for being at the forefront of linguistics research, evaluated 30 writing systems in terms of their rational, scientific, and unique characteristics. This study ranked Hangul first" (The Korea Foundation, 2010: 14).

Hangul is indeed a highly celebrated innovation; every year in South Korea, October 9th marks *Hangul Day*, a national holiday to commemorate the achievement of King Sejong and his team of linguists, who were responsible for the creation of the Korean alphabet.

The creation of Hangul represents an important step in Korean history, not just in the innovation of Hangul itself, but also in the history leading up to its development – the history of using Chinese characters in lieu of a native script. Spanning the course of many hundreds of years, written Korean was expressed using borrowed Chinese characters, using the sounds and meanings of a non-native script to fit the needs of a native grammar. Today, although Hangul is the primary script,

the influence of Chinese remains pervasive, its use still reflected both in print and in the media at large. In many ways, the Hangul script represents only half of the Korean writing system, the other half being Hanja, the borrowed Chinese characters that still remain in use. The focus of this thesis is thus on the holistic Korean writing system which comprises both scripts, the native Hangul and the non-native Hanja, both of which are used to write Korean. The main aim of this research is to examine the relationship between the two in psycholinguistic processing, to understand how a language's writing system can affect semantic parsing.

I was reminded of my interest in Hanja characters very recently, while I was watching a Korean television show about a man who helps an orphaned boy find his family.[1] I am grateful that I came across this show, because it helpfully reminded me why I asked these questions in the first place and what motivated me to embark on this project. In one particular scene, the orphaned boy makes a joke to the man, asking him who in Korean history is the poorest king. The man is visibly confused, and the boy quickly supplies the punchline, "Minimum wage". It is not immediately clear how the punchline relates to the joke, nor it is obviously funny, until one begins to think about the relationship between Hangul and Hanja. Figure 1.1 provides an explanation of the joke:

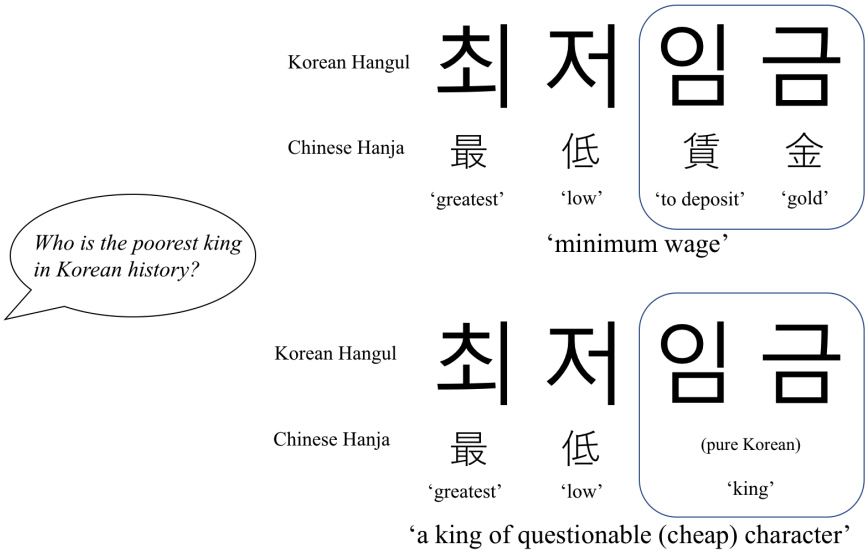


Figure 1.1: The punchline is provided in Korean Hangul, along with the corresponding Chinese characters.

[1] It actually turned out that the man was his biological father.

The punchline 최저임금 is a borrowed phrase, in that each of the syllables derives its meaning from a traditional Chinese character, which itself encodes a meaning.

Taking the composite meaning of the four syllables, the phrase comes to mean 'the lowest amount of money', or 'minimum wage'; however, in pure Korean (i.e., Korean which is not derived semantically from Chinese), the last two syllables can also be used to mean 'king'. The combination of the two types of Korean (one which is native, and one which is derived in part from the relationship between the two scripts) creates a phrase meaning 'a cheap king' (i.e., a king of cheap or questionable character). When the boy asks the man "Who is the poorest king in Korean history?", the response is a play on words based on the semantic relationship that is borne out between the two scripts of the Korean writing system. The answer is 'a king of cheap character', but the same written form can equally be used to express the meaning of 'minimum wage'.

1.1 Contribution to the field

While playful, this joke raises a number of interesting questions, especially with regards to how semantic meaning in a bi-scriptal lexicon like Korean is represented in the mental lexicon. The joke is clever because it makes use of the semantic contributions of the Hanja characters that are in many ways 'hidden' behind a bi-scriptal relationship, ultimately creating two different readings of the same written expression. In Korean, does semantic processing actively rely on the mental representations of scripts and the semantic contributions of scripts that are, in fact, hidden behind a cross-script dimension? Does processing of Sino-Korean involve semantic access to multiple meanings, as encoded by different Hanja characters?

More broadly, one of the key objectives of this thesis is to address the limited coverage of East Asian languages in psycholinguistic literature. While there are accounts of the Chinese and the Japanese mental lexicons, there is limited research in the way of an exhaustive account of the Korean mental lexicon. Existing accounts make little room for the processing role that Hanja characters play in terms of their semantics, and jokes like the one presented in Figure 1.1 provide important

anecdotal evidence for the importance of Hanja characters in semantic processing. The main aims of this thesis are thus to address these questions, and in doing so, lay the foundation needed to theorize the structure of the Korean mental lexicon, and importantly, one which takes into account the representation of Hanja characters.

1.2 Outline of thesis

The thesis proceeds as follows: Chapter 2 introduces Korean and its two scripts, highlighting some of the important history surrounding the innovation of Hangul, as well as a description of the current composition of the Korean lexicon, as a result of the extensive Chinese influence. This is then followed by Chapter 3, which provides a brief survey of different models of lexical processing in auditory and visual word recognition, along with existing accounts of semantic processing where more than one meaning is available. Chapter 3 will also provide an overview of the available research on script processing, both in Korean as well as other East Asian languages, such as Chinese and Japanese, as well as introduce some of the existing models of the mental lexicon for Chinese, Japanese, and Korean.

Chapters 4 and 5 present the findings of the four behavioral studies conducted in this thesis to investigate the status of Hanja in the mental lexicon. This is followed by Chapter 6, where a revised model of the structure of the Korean mental lexicon is proposed. This revised model is based on the experimental findings presented in this thesis and builds on the currently available model of the Korean mental lexicon, as first put forth by Yi (2009). The proposal described in this thesis offers a new way to account for the semantic contributions of the different Hanja characters, by introducing semantic cohorts based on cross-script relations between Hanja and Hangul. These cross-script semantic cohorts offer a more thorough account of the Korean mental lexicon, by positing a model of the Sino-Korean sub-lexicon that reflects the mental representation of the semantics of Hanja characters. Finally, Chapter 7 offers some concluding remarks as well as suggestions for future research.

2

Writing systems, the innovation of Hangul, and the Korean lexicon

Contents

2.1 Scripts and writing systems 6
2.2 The Korean writing system 9
2.2.1 The use and adaptation of Hanja in Korean 10
2.2.2 The history of Hangul and the sounds of Korean 12
2.2.3 The geometry of Hangul 16
2.3 The Korean lexicon 23
2.3.1 The Sino-Korean lexicon 24
2.3.2 Encoding multiple meanings in Sino-Korean 27

In the present chapter, I introduce the Korean writing system. I begin the chapter by explaining key terminology relevant to the discussion of writing systems and scripts. The chapter then moves on to discuss Hangul and Hanja, including a brief description of the history surrounding the use of Chinese characters, which bears important consequences on the composition of the Korean lexicon. As part of the description of Korean, I also include some explanation of the phonology of Korean, so as to better contextualize how the different scripts in Korean represent the sounds of the language. Finally, I conclude the chapter by illustrating how the structure of a writing system plays an important role in lexical processing, bringing to light some theoretical considerations for how we then might conceptualize the structure of the mental lexicon for languages like Korean that use two different scripts to express semantic meaning.

2.1 Scripts and writing systems

Before further discussion, the definition of two recurring terms – *script* and *writing system* – is presented. A script is commonly defined as the visual appearance of a writing system, while a writing system refers to the principles that govern how symbols are mapped to the sounds of the language, including spelling rules (Coulmas, 1996; Hansen, 2008; Katz & Frost, 1992; Perfetti & Dunlap, 2008). Most fundamentally, the orthography of a language represents a form of standardization of the writing system, forming a collection of rules that allow language to be visually represented. Orthographic systems can be contrasted with pictographic systems; in the latter, pictures and signs represent meaning directly, characterized by non-arbitrary relations between the graphic sign and the transcribed meaning (Frost, 2005). The visual shape of a graphic sign encodes in some part its meaning, and this meaning can be recovered without explicit prior instruction. On the other hand, for written words, the relationship between graphic signs and meaning are marked by a degree of arbitrariness. This arbitrariness is the result of graphic signs conveying only the phonological units systematically, while the mapping of phonological units onto semantic meaning remains arbitrary.

The way in which different orthographic systems represent the spoken language depends largely on the specific characteristics of each language. For example, the individual scripts of different languages differ categorically in what they represent (Taylor & Olson, 1995). That is, writing systems can vary in terms of the size of the linguistic unit that the orthographic unit transcribes. Alphabetic scripts like Spanish and English represent the phonemes, while syllabic scripts, such as those used in Japanese, represent syllables (or moras). Such differences in the unit of representation may be attributable to reasons of efficiency pertaining to memory constraints or processing limitations (Frost, 2005). In the Japanese syllabic system, kana, for example, the graphemes represent a possible 116 permissible syllables, including those with a final nasal, in the Japanese language, which is notably small given that Japanese has fourteen consonants and five vowels, and the only syllabic structures permitted in the language are V and CV combinations. Given these

parameters, the most efficient representational system in Japanese is one which provides a distinct graphemic representation for each of these possible syllables.

Learning 116 graphic signs by rote memory in Japanese is less demanding given the sharp contrast English provides, where the syllabic structure is not only far more complex (Akmaijian, Demers, Farmer, & Hamish, 2001), but also more numerous. English has twenty-four consonants and fifteen vowels, which can combine a number of different ways to create a multitude of different syllabic combinations (e.g., CV, VC, CVC, CCVC). This means that English would require approximately 15,000 distinct representations in order to accommodate all possible structures, which would also have serious consequences for the ability of the English native speaker to retrieve individual syllables. To complicate matters, syllable boundaries are not always clear in English words, with issues of ambisyllabicity arising (Kahn, 1976). Thus, unlike Japanese, for the aforementioned reasons, a syllabic orthography would be poorly suited to English.

While both Japanese and English exemplify alphabetic systems of writing, the sub-linguistic unit that is represented by the orthographic unit is different for each language (i.e., the phoneme for English, and the syllable for Japanese). In addition to alphabetic systems, languages can also employ writing systems with a graphemic structure that represents meaning, rather than sound. For example, logographic systems, such as Chinese, represent meaningful morphemes at the lexical level, rather than phonological units at the sub-lexical level.[1] Each logographic grapheme in Chinese, in representing entire concepts, encodes meaning. Thus, different writing systems represent different units of the spoken language, whether it is the phoneme, syllable or morphophoneme.

Based on the different ways in which a writing system maps sound to form, the extent to which such mappings are transparent or opaque will have important consequences on processing. This particular property – the level of consistency that characterizes the relationship between phoneme and grapheme – is called

[1]Frost (2005) notes that even in Chinese 90% of the characters are phonetic compounds, though not always consistent, and only 10% are semantic determiners that differentiate between the many homophones present in Chinese (DeFrancis, 1989).

orthographic depth. Differences in orthographic depth point to structural differences between different writing systems, which may also influence whether during lexical processing, more sub-lexical or lexical routes are recruited. The notion of orthographic depth can be conceptualized as a continuum, with languages falling at various points depending on the relative degree of transparency with which their writing systems encode phonology. An orthography that represents its phonology consistently according to simple phoneme-to-grapheme correspondences is considered *shallow*, while an orthography where the relationship is less transparent is known as *deep*. Before addressing Hangul and Hanja more specifically, I provide a brief theoretical discussion of whether a writing system of a particular orthographic depth is preferable. Differences in orthographic depth have also been shown by a number of studies to affect processing, namely, that languages of shallow orthographic depth, such as Persian (Baluch & Besner, 1991) and Italian (Tabossi & Laghi, 1992) tend to recruit more sub-lexical information for word recognition than do those reading a deeper orthography, such as English.

A shallow system may seem preferable for ease of mapping between sound and form, but there are potential inconsistencies worth considering which arise at the morphemic level. For example, in Dutch, the word for 'hand' is *hand*, which is pronounced /hant/, while the plural for 'hand' is *handen*, but is pronounced /handən/, rather than /hantən/. By not marking syllable-final devoicing in /handən/, this suggests that the orthographic system and spelling conventions of Dutch prefer morpheme consistency (***hand*** and ***hand****en*).[2] In other cases of Dutch, particularly for fricatives, the final devoicing is marked, as observed in the singular and the plural for 'grape', *druif* and *druiven*.[3] Dutch exemplifies a case where morphemic inconsistencies arise despite the structural advantages of a shallow orthographic system.

Another possible issue with a shallow system is that it requires a particular dialect to be considered the standard. Italian, for example, draws a distinction between open /e/ and closed /e/; however, different dialects of Italian will distribute these

[2]In earlier Dutch, it should be noted that the singular of 'hand' was *hant*.

[3]Both examples in Dutch are provided by W. de Melo (personal communication, 20 October 2017).

phonemes differently, which means that their orthographic representations may not be consistent across different distributions (W. de Melo, personal communication, 20 October 2017). Orthographically, both phonemes are expressed as *e*, which creates a degree of underspecification that makes the system less prescriptive from a dialectal perspective. With all this in tow, the following section discusses some key aspects of Korean, as it relates to its writing system and phonology, and the history surrounding the development of Hangul.

2.2 The Korean writing system

The Korean writing system comprises two distinct scripts: Hangul and Hanja. Although confined to one writing system, these two scripts are structurally different. Hangul is fundamentally an alphabetic system, much like English, where the letters of the system each represent a phoneme available in the language. However, unlike English, this relationship maintains much greater phoneme-to-grapheme consistency in Korean (Sohn, 1994), which creates highly transparent mappings between the spoken and the written forms. As a result, this places Hangul at the shallow end in terms of orthographic depth.

Hangul is made further unique by the fact that, although it is an alphabetic system, it is written like a syllabary where two or more alphabetic letters, or graphemes, are grouped together to create syllables. This characteristic has led researchers to describe the system as an alphabetic syllabary (Taylor, 1980). In addition to the native script Hangul, Korean also uses Chinese characters, or Hanja, which are logographs and, unlike alphabetic letters, represent meaning, rather than a particular sound. Thus, the granularity of each of the scripts within the Korean writing system is different. Given the use of both, this has important implications for how semantic information within the Korean mental lexicon is stored and accessed across the two different scripts. The following subsection will discuss in greater detail the specific complexities of each script.

2.2.1 The use and adaptation of Hanja in Korean

The Hanja script represents the spoken language of Korean through a logographic representation of morphemes (Kim-Renaud, 2009). Before the 15th century, there was no system in place that fully enabled Korean speakers to express the characteristics of their native language, and subsequently, they had to use Chinese characters, or Hanja, the Korean name for Chinese characters. Hanja characters are borrowed Chinese characters, each one directly encoding and representing meaning-bearing morphemes. Before the creation of a native script, the Korean language was written by way of using literary Chinese or modified Chinese to transcribe Korean, a process that often involved borrowing the meaning or the pronunciation of the Chinese characters, and in many cases in the early days, both. Due to this ideographic nature, Hanja characters were used to convey meanings as well as sounds in Korean (Chang, 1996).

Furthermore, the ideographic nature of Hanja characters allows for ease of creation of new words by way of compounding one with another; for example, consider 화 (話), 'speech', it can be used as a root, a prefix, or a suffix, creating such forms as 수**화** ([su.hoa]), meaning 'sign language', **화**용 (hoa.joŋ]), meaning 'pragmatics', and **화**행 ([hoa.hɛŋ]), meaning 'speech act' (Sohn, 1994; Chang, 1996). Just as well, since each morpheme has a sound, the Hanja character also represents the sound of the morpheme. As logographs, Hanja characters maintain almost no transparency in the relationship between orthography and phonology. It is important to note that Hanja characters are traditional Chinese characters, and not simplified Chinese characters.[4] The pronunciation of Sino-Korean words represent approximations of the original Chinese sounds, but crucially, the original Chinese lexical tones are no longer retained in Sino-Korean. For example, the word for *Korea*, which is written as 韓國, in Hanja characters, is pronounced [han.kuk] in Korean, and [hán.k'wó] in modern Mandarin (Taylor, 1980).

Within Sino-Korean, only a relatively small number of Sino-Korean characters have maintained the same sounds as those of the original Mandarin Chinese, with the majority of Sino-Korean words having modified the sounds to be different

[4]To note, simplified characters are still in use in both Hong Kong and Taiwan.

from their original counterparts. Thus, in some ways, the sounds of Sino-Korean form their own autonomous system, though it remains open to further question the precise origins of Sino-Korean sounds. The research, however, does seem to converge on the understanding that pronunciations of Chinese characters used in the northern part of China during the Sui and Tang dynasties around the seventh and eighth centuries constitute the basis of Sino-Korean sounds and pronunciations (Sohn, 1999). Therefore, broadly, it seems that contemporary Sino-Korean sounds are similar to those of Middle Chinese, although the phonological adaptation does not include a few independent vowel and consonant changes, nor does it involve retaining the original tones. Also to note, Middle Chinese is the ancestor of all Sinitic languages other than the Min branch.

To demonstrate some of the phonological differences Sino-Korean exemplifies from the original Middle Chinese, highlighting the ways in which they were adapted phonologically, Table 2.1 presents Sino-Korean pronunciations alongside their corresponding sounds in Early Middle Chinese and Contemporary Mandarin (Pulleyblank, 1991).

Table 2.1: The different ways in which characters are phonologically rendered across Sino-Korean and Chinese. To note, all of the Sino-Korean Hanja characters presented in this table can also be rendered in Hangul (which is provided in the second column).

Hanja	Hangul	Sino-Korean	Middle Chinese	Mandarin
一 'one'	일	[il]	[ʔjit]	[jī]
百 'hundred'	백	[pɛk]	[paɨjk]	[p'ǎj]
國 'country'	국	[kuk]	[kwək	[k'wó]
七 'seven'	칠	[ts^{h}il]	[tshit]	[tsī]
兵 'soldier'	병	[pjʌŋ]	[piajŋ]	[p'īŋ]
不 'not'	불	[pul]	[put]	[p'ù]
冊 'book'	책	[ts^{h}ɛk]	[tshɛːjk]	[ts^{h}ə̀]
答 'answer'	답	[tap]	[tap]	[t'ā]
立 'stand'	립	[ɾip]	[lip]	[lì̀p]
葉 'leaf'	엽	[jʌp]	[jiap]	[jè]
店 'shop'	점	[tsʌm]	[tɛmh]	[t'jàn]
見 'see'	견	[kjʌn]	[kɛnh]	[tʃ'jàn]
江 'river'	강	[kaŋ]	[kaɨwŋ]	[tʃ'jāŋ]

As Table 2.1 shows, morpheme-final stops in Middle Chinese are also reflected in Sino-Korean, but ultimately deleted in Contemporary Mandarin. Furthermore, the morpheme-final [t] in Middle Chinese has changed to [l] in Korean. The syllable-final /p/, /t/, /k/, and /m/ still remain in some southern dialects of Chinese, as they do in Korean, and Cantonese, for example, preserves all final stops. However, using a non-native script to express Korean involved obvious challenges, least of all, the issue of exclusivity, given that literacy in Chinese characters was often limited to the upper class. Thus, as it became increasingly unfit for use, King Sejong was motivated to develop a native script better fit to express written Korean.

2.2.2 The history of Hangul and the sounds of Korean

King Sejong, the fourth monarch of the Joseon Dynasty (1392 – 1897), commissioned *Hangul*[5], the Korean system of writing, in 1444 (Lee, 1983; Lee, 1990). Utilizing

[5]The script was originally known as *Hunmin jeongeum*, which refers to the document describing the script itself. Hangul is the name by which the Korean alphabet is known today.

both the pronunciations and the meanings of the Chinese characters, the ancient Korean people implemented the following archaic systems of writing: *Hyangchal* and *Idu*, both of which involved some, if not all, Korean words being represented using Hanja. The Hyangchal system is best known for its use in *Hyangga*, a form of poetry developed during the Silla era, and involved all Korean words being rendered in Chinese characters; the Idu script, on the other hand, used Chinese characters in their proper meaning, to represent content words, such as nouns, verbs and adjectives, while relatives and adverbs were represented using Korean-phoneticized Chinese characters, bearing no relation to their proper meaning (H.-G. Kim, 1983: 122).

The ancient Korean people devised yet another system called *Gugyeol* during the Joseon era. Under this system, the original text remained in classical Chinese, and instead additional markers were placed to demarcate phrases, the Korean reader thereby reading parts of the Chinese sentence out of sequence to approximate the Korean (SOV), rather than the Chinese (SVO) order. Thus, the innovation of the Hangul alphabet was grounded in three main motivations (Park, 2015: 28; Lee, 1990):

1. Due to the sounds of Korean words being different to those of the sounds found in Chinese words, it proved increasingly difficult for Koreans to communicate with one another by means of Chinese characters.

2. The need for Chinese-character literacy for communication required privilege of education, which was inaccessible to those who were not in the elite class; much of the lower class proved unable to express their thoughts in written form.

3. The twenty-eight characters (of Hangul) were designed to be learned easily by the ordinary people, so as to increase literacy.

Unlike any other system of writing, which typically evolves from abstract shapes or prototypic forms, Hangul exemplifies deliberate invention, an innovation informed by King Sejong's acute awareness of his subjects' needs. The King was also motivated

by his general dissatisfaction with the inability to communicate with all of his subjects, since learning of classical Chinese was confined to the upper class.

Above all, King Sejong valued the notions of convenience and accessibility for daily use, explaining that "a wise man can acquaint himself with them before the morning is over; a stupid man can learn them in the space of ten days" (Ledyard, 1998: 258; Park, 2015). As a result, a native alphabet well-suited to write Korean was created, and no longer were the Korean people confined by the limitations of using a logographic system phonemically to transcribe Korean. As such, there was no further need to learn the Chinese characters of the region to express native Korean sounds. Hangul was henceforth promulgated as the official Korean script. Since the time of promulgation, the system has undergone a number of phonological changes, with some phonemes now obsolete, but the system that is known today contains twenty-four graphemes, each mapping onto a specific phoneme. The nineteen consonant phonemes of Korean are presented in Table 2.2.

Table 2.2: The consonantal inventory of Korean (Kim, 2015).

		Place			
Manner		Bilabial	Alveolar	Velar	Glottal
Stop	Lax	p	t	k	
	Tense	p*	t*	k*	
	Aspirated	p^h	t^h	k^h	
Affricate	Lax		ts		
	Tense		ts*		
	Aspirated		ts^h		
Fricative	Lax		s		h
	Tense		s*		
Nasal		m	n	ŋ	
Liquid			l		

To briefly note, there are more phonemes than there are graphemes because the tense consonants are represented by doubling graphemes (e.g., the grapheme ㅂ represents the phoneme /p/, but is doubled (ㅃ) to express tensed /p/, which is represented by the asterisk (*) in Table 2.2.)

The most notable of features in the phonology of Korean is the three-way distinction that is found in stops and affricates, along the dimensions of tenseness

and aspiration (Ahn, 2009). The lax segments are characterized by a breathy quality, while the aspirated segments are strongly aspirated. Tense segments are marked by greater glottal tension compared to the other segments found in Korean (Cho, Jun, & Ladefoged, 2002; Lee & Ramsey, 2000). This three-way distinction is phonemic, such that /tal/, /t*al/, and /t^{h}al/, by difference of a single phoneme, all have different meanings (e.g., 'moon', 'daughter', and 'mask', respectively). As Table 2.2 also shows, Korean has bilabial, alveolar, and velar stops (Ahn, 2009; Lee & Ramsey, 2000).

The lax series of voice segments can be realized as their voiced allophones; for example, the Korean word for *wave* is represented phonemically as /p^{h}a.to/, but it is phonetically realized as [p^{h}a.do] (Kim, 2015; Lee & Ramsey, 2000). The Korean alveolar fricatives maintain a two-way phonemic distinction according to the degree of tenseness. The words for *rice* and *flesh* in Korean exemplify this contrast, as the words are realized as /s*al/ and /sal/, respectively. The lax fricative, /s/, is not voiced, but can be palatalized before high front vowels and slackened intervocalically (Ahn, 2009). The fricative, /h/, is the only glottal sound in Korean, and it is often dropped between voiced segments in natural speech (Lee & Ramsey, 2000). The remaining consonants are three nasals and one liquid. The liquid is realized as [l] either word-finally or when in a sequence of two liquids at a syllable boundary, but is realized as [ɾ] word initially or intervocalically. In addition to the consonants, Table 2.3 provides the inventory of vowels in Korean.

Table 2.3: The vowel inventory of Korean (Sohn, 1999). To note, Sohn (1999) uses phonemic values to represent individual sounds.

place / lips / tongue	front		back	
	unround	round	unround	round
high	i	y	ɨ	u
mid	e	ø	ə	o
low	ɛ		a	

The ten vowel phonemes in Table 2.3 are categorized in terms of the place of articulation, lip shape, and tongue height (Sohn, 1999). All ten vowels are orthographically distinguished across all dialects of Korean, although some dialects may exhibit less pronounced contrasts between them (Sohn, 1999).

To express the sounds of Korean, symbols were composed from a combination of lines and circles, which, in various combinations, produced approximately 12,000 phoneme combinations, or syllables (Park, 2015). The three main vowels (a dot, a vertical line, and a horizontal line) represented the roundness of the heaven, the flatness of the earth, and the upright posture of man, depicted in Figure 2.1 (Lee, 1990: 208; Kim-Renaud, 2009: 9).

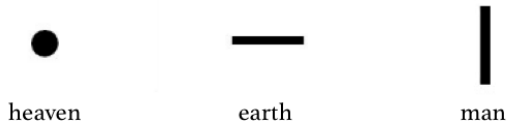


Figure 2.1: The main vowels based on three main elements: *heaven*, *earth*, and *man*.

Although the dot was eventually replaced by a short line for ease of calligraphy done by brush, the sound that the dot encoded was lost altogether from the language, and therefore the symbol is no longer in use today (Kim-Renaud, 2009: 9; Ramsey, 1997).

2.2.3 The geometry of Hangul

The twenty-four graphemes of the Korean alphabet are presented in Figures 2.2 and 2.10. The consonants are presented in Figure 2.2, with the nine visually prototypic consonants offset in blue.

ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ
[k]	[n]	[t]	[l]	[m]	[p]	[s^h]
ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
[ŋ]	[ts]	[ts^h]	[k^h]	[t^h]	[p^h]	[h]

Figure 2.2: The fourteen simple consonants and their phonetic transcriptions (Sohn, 1994). The nine visually prototypic consonants are within the blue box.

Building from the prototypic forms given in blue, the remaining consonants can be created by simple addition of vertical or horizontal lines. This can be observed in the character ㅋ /k^h/, which is the result of adding a secondary horizontal line to the character ㄱ /k/ (Sohn, 1994; Chang, 1996). This is interesting in its

own right as it highlights the "featural" design of Hangul, where different strokes represent specific phonologically distinctive features, such as place and manner of articulation, e.g., aspiration and tenseness (Sampson, 1985). The additional horizontal line used in creating the character ㅋ encodes the phonetic feature of aspiration. This is visualized in Figure 2.3.

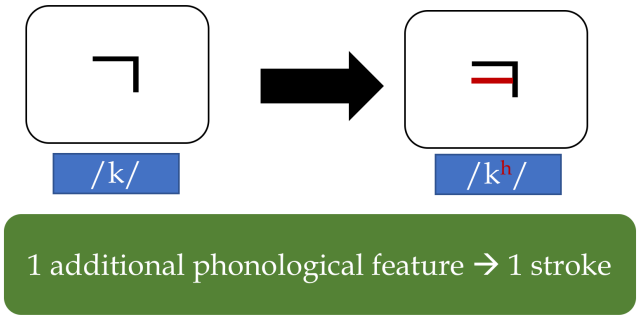


Figure 2.3: The additional horizontal stroke represents the phonological feature of aspiration, creating two phonologically contrastive characters, ㄱ /k/ and ㅋ /kʰ/.

The same pattern of letter formation can be seen for other Hangul characters in Figure 2.4.

	Nasals	Stop and affricate	
		Unaspirated	Aspirated
Velar		ㄱ /k/	ㅋ /kʰ/
Alveolar	ㄴ	ㄷ /t/	ㅌ /tʰ/
Alveolo-palatal/ palatal		ㅈ /ts/	ㅊ /tsʰ/
Bilabial	ㅁ	ㅂ /p/	ㅍ /pʰ/
Glottal	ㅇ		

Manner of articulation
+ Phonology
+ Visual

Figure 2.4: The phonological encoding of aspiration via an additional stroke can be observed in other contrastive pairs.

Along the dimension of manner of articulation, for each of the unaspirated and aspirated pairs, their visual shape encodes the phonetic feature of aspiration through an additional stroke. This phonological contrast is directly manifest in the visual representation of the Hangul character. Thus, the previous illustrations show that

Hangul is a system built based on the principles of sound, and closer examination reveals that even individual characters reflect the shape of the lips and the mouth when producing that particular sound (Park, 2015).

For example, bilabial sounds are those that are made when the mouth is closed and then opened. The Hangul character ㅁ /m/ is a depiction of this movement. The character ㅂ /p/ is used to depict the shape of the lips parting when the mouth opens, and the character ㅍ /p^h/ reflects the lips spreading sideways. Figure 2.5 illustrates the ways in which the movements of the mouth are reflected in the visual shape of Hangul characters for bilabial sounds.

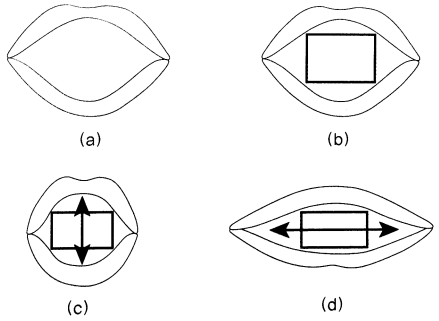


Figure 2.5: The shape of the mouth and the lips and the specific movements involved in the production of ㅁ /m/, ㅍ /p^h/, and ㅂ /p/ (Park, 2015: 57).

The following depicts the movements of the mouth that are involved in the production of aveolar sounds in Korean, and how the corresponding Hangul forms reflect such movements.

For example, the shape of the tongue when it touches the aveolar ridge (upper palate) and comes down is reflected in the character ㄴ /n/. The characters ㄷ /t/ and ㅌ /t^h/ reflect the shape of the speech organ and the position of the tongue at the time of articulation.

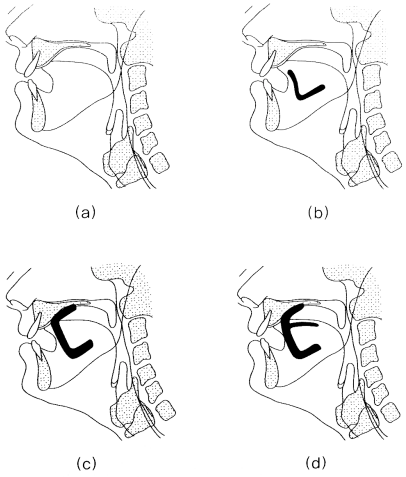


Figure 2.6: The shape of the mouth and the lips and the specific movements involved in the production of ㄴ /n/, ㅌ /t^h/, and ㄷ /t/ (Park, 2015: 58).

As Figure 2.7 shows, the characters ㄱ /k/ and ㅋ /k^h/ depict the movement of the root of the tongue involved during the production of velar sounds.

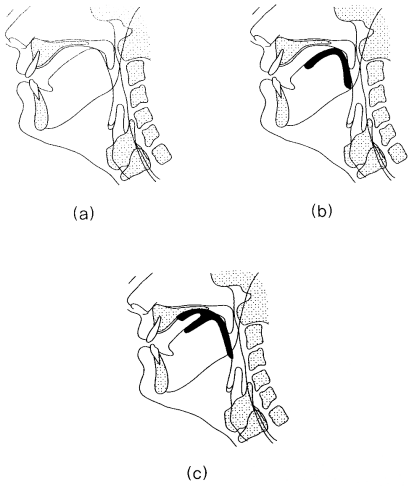


Figure 2.7: The shape of the mouth and the lips and the specific movements involved in the production of ㄱ /k/ and ㅋ /k^h/ (Park, 2015: 59).

The characters ㅅ /s/, ㅈ /ts/, and ㅊ /ts^h/ reflect the movement of the tongue between the upper and lower teeth or towards the back of the upper teeth.

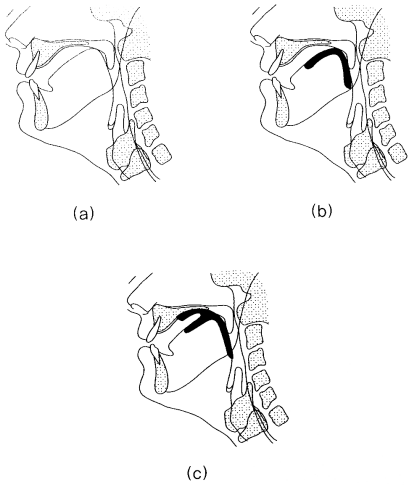


Figure 2.8: The shape of the mouth and the lips and the specific movements involved in the production of ㅅ /s/, ㅈ /ts/, and ㅊ /ts^{h}/ (Park, 2015: 60).

Finally, the characters ㅇ /ŋ/ and ㅎ /h/ visualize the shape of the mouth for sounds made from the throat.

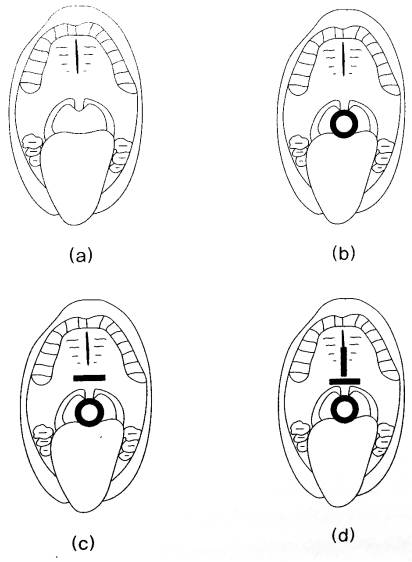


Figure 2.9: The shape of the mouth and the lips and the specific movements involved in the production of ㅇ /ŋ/ and ㅎ /h/ (Park, 2015: 61).

Korean also has ten simple vowels, which are presented in Figure 2.10. Compared to consonants, vowels are less "featural" in their visual design, simply because vowels require no constriction in the vocal tract for production.

ㅏ	ㅓ	ㅗ	ㅜ	ㅣ
[a]	[ʌ]	[o]	[u]	[i]
ㅑ	ㅕ	ㅛ	ㅠ	ㅡ
[ja]	[jʌ]	[jo]	[ju]	[ɨ]

Figure 2.10: The ten simple vowels and their phonetic transcriptions (Sohn 1994).

Different combinations vowels yield different diphthongs and semi-vowels, which are given in Figure 2.11.

ㅐ	ㅒ	ㅔ	ㅖ	ㅢ	ㅘ	ㅙ	ㅚ	ㅝ	ㅞ	ㅟ
[ɛ]	[jɛ]	[e]	[je]	[ɨi]	[oa]	[oɛ]	[ø]	[uʌ]	[ue]	[ui]

Figure 2.11: Semi-vowels and diphthongs, and their phonetic transcriptions (Sohn, 1994)

With both consonants and vowels discussed, the following describes the syllabic structure in Korean and the different visual configurations of phonemes to create syllables.

Unlike most other alphabetic systems, which use a linear arrangement of characters (often Roman), Hangul is arranged non-linearly; that is, the arrangement involves both left-to-right and top-to-bottom directions, creating a block-like assembly of characters (Park, 2009; Shin, Kiaer, & Cha, 2013). There are two possible structures a syllabic block can follow: either it is of C1 + V composition, or it is built according to a C1 + V + C2 structure, where C1 refers to the initial consonant character, V the medial vowel character, and C2 the final consonant character (Sohn, 1994; 1999). There are nineteen initial, twenty-one medial, and twenty-seven final consonants, totalling a possible 11,172 Hangul blocks, though this is not to say that all are in frequent use (Park, 2009). Kim and Kang (1997) reported that Koreans use, at most, 2,305 of the possible blocks, but in actuality, only 808 are used with frequency.

Fundamentally a phonetic script, a syllabic block can be decomposed into its constituent phonemes, making pronunciation of any one block relatively transparent. This is visualized in Figure 2.12.

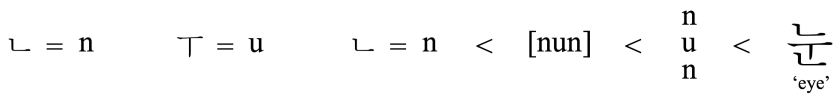


Figure 2.12: Korean is composed by way of non-linear assembly of consonant and vowel symbols, as demonstrated by 눈.

In addition to the formation presented in Figure 2.12, Figure 2.13 illustrates the other commonly used types of syllabic blocks, and their relative usage in the language (Kim & Kang, 1997; Park, 2009).

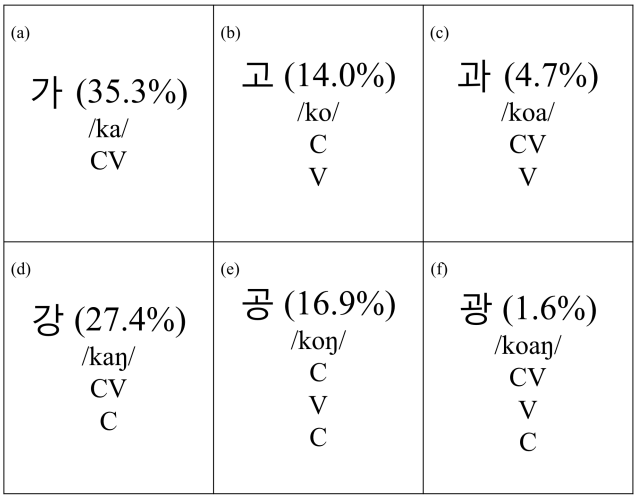


Figure 2.13: The six main types of syllabic blocks in Korean and their relative frequencies; (a) and (b) show the vowel ㅏ[a], with (a) and without (d) a final C ; (b) and (e) the vowel ㅗ [o], with (b) and without (e) a final C; (c) and (f) demonstrate use of both vowels ㅏ and ㅗ, with (c) and without (f) a final C.

This non-linear arrangement underscores the visual saliency of the syllable in Korean, so that the syllable can be easily identified in text. Individual syllables represent morphemes, and depending on its status as bound or unbound, the syllabic block can constitute a word on its own. This transparent mapping of syllable to morpheme is particularly high in the case of words of Chinese origin, i.e., Sino-Korean, relative to native Korean words or Western loans. Given the duality of script in the Korean writing system, I now discuss how the current composition of the Korean lexicon reflects the extensive history of Chinese influence.

2.3 The Korean lexicon

The influence of Chinese can be directly observed in the current composition of the Korean lexicon, which consists primarily of Sino-Korean words. As previous sections have shown, the use of Chinese Hanja characters represents an important part of Korean history, the consequences of which can be observed in the composition of the Korean lexicon that is seen today. There are three sub-lexicons comprising the larger Korean lexicon: native (or pure) Korean, Sino-Korean, and European (i.e. loanwords) (Taylor, 1997), although this thesis will focus primarily on the native and the Sino-Korean sub-lexicons.[6] Morphemes belonging to the native Korean sub-lexicon make use of all of the sounds and syllables available in Korean. Native Korean words are always written in Hangul and cannot be expressed using Hanja characters.[7]

While all words can be written in Hangul, not all can be written in Hanja. Sino-Korean can be written in both Hangul and Hanja, but native Korean can only be written in Hangul, as Figure 2.14 shows both the Sino-Korean, and the pure Korean word for 'drawing'. The Sino-Korean word for 'drawing' can be written in Hanja, while the pure Korean word for 'drawing' can only be written in Hangul.

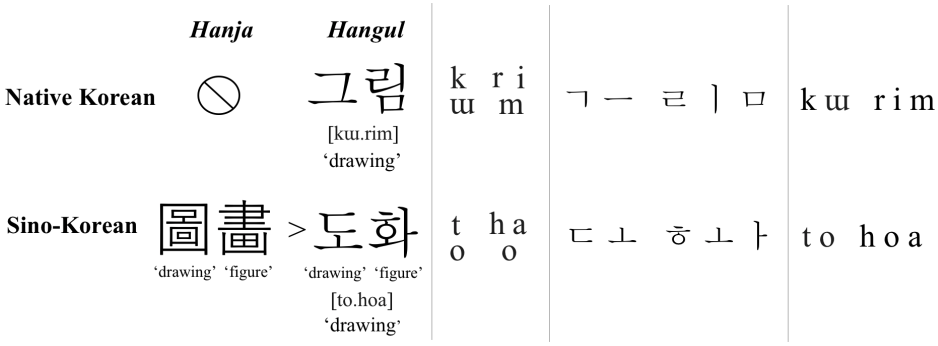


Figure 2.14: The two syllable types in the Korean lexicon and the structural composition of individual syllables. In both native (A) and non-native (i.e., Sino-Korean) (B) syllables, the Hangul characters are assembled non-linearly into blocks, but Sino-Korean syllables can also be rendered one step further using Hanja characters.

Equally, Figure 2.15 shows how the same Sino-Korean phrase can be expressed in both Hangul and Hanja. The top line represents the original Hanja characters

[6] European loanwords consist mostly of English loanwords, such as *computer* and *coffee*, which when written in Korean, can be done so in Hangul (e.g., 컴퓨터, 커피, respectively).

[7] It is important to clarify that the distinction between Sino-Korean and native Korean is a matter of type, rather than of token, as native Korean words are generally more frequent that Sino-Korean words, despite the composition of the Korean lexicon being heavily Sino-Korean dominant.

from which the Hangul is derived, both in meaning and pronunciation. That is, the meaning of 'heaven' that is attributed to the Hangul 천 comes from the original Hanja; likewise, the Hanja is pronounced [tʃʌn] in Chinese, and as such the Hangul characters ㅊ[tʃ], ㅓ[ʌ], and ㄴ[n] are assembled to form 천 to approximate the same sound.

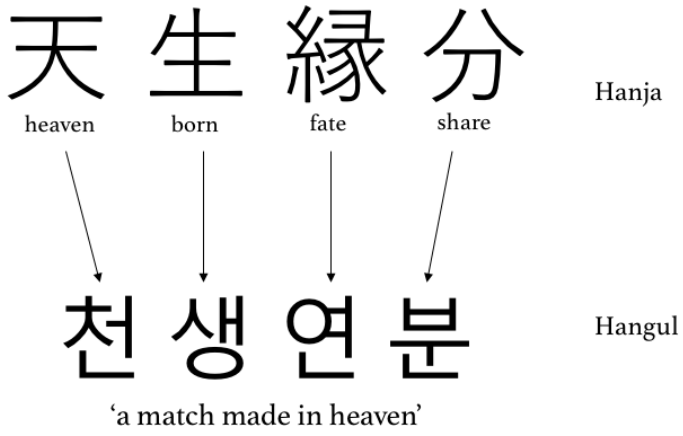


Figure 2.15: The same phrase rendered in both Hanja and Hangul

The two sub-lexicons differ in origin (native Korean versus Sino-Korean), but share a common script (Hangul).

2.3.1 The Sino-Korean lexicon

The Sino-Korean sub-lexicon comprises Chinese words that have been borrowed into Korean over the span of a thousand years until the early 20th century (Taylor, 1997). As Taylor describes, "They are not only numerous but also deeply entrenched in the Korean vocabulary" (1997: 300). Sohn (1999) explains that while it is not clear exactly how and when Chinese characters were first introduced into Korea, the earliest beginnings of mass borrowings can be dated starting from the Goryeo Dynasty (935-1392), continuing on into the subsequent Joseon Dynasty (1392-1897).[8] There continued to be transmission of Chinese characters and loanwords even after the development of Hangul, further underscoring the extent to which the influence of

[8]Sohn (1999) explains that because Wiman from Yen in China founded a primitive Korean state in the north-western part of the Korean peninsula in 194 BC, practical knowledge of the Chinese script dates as early as the second century BC, predating even the first set of loans entering Korean.

Chinese was entrenched in the Korean language. Laws, regulations, administrative decrees, official documents, petitions to government agencies, scholarly works continued to be written exclusively in Chinese Hanja characters, even with Hangul as a working script, until 1894, at which point official reforms were enacted to ensure Hangul was used as the primary script for all administrative purposes, including the passing of laws and royal decrees. Although the status of Hangul has clearly been promoted to the dominant script in Korea since then, Hanja characters have never been wholly abandoned.

The most important feature of the Sino-Korean lexicon, and the reason for the phenomenon motivating the research in this thesis, is that Sino-Korean words use only about 440 of the 2,000[9] most frequently used syllables in Korean, which means that it is very often the case that morphemes and words are in fact homophones (Sohn, 1999; Taylor, 1997). For example, the morphemes for 'hand' and 'water' share the syllable 수 [su]. These homophones can be disambiguated when written by their corresponding Hanja characters, e.g., 手, 'hand', and 水, 'water'. These Chinese characters, each having different meanings, are written differently. Although they sound different in Chinese because of different tones, the absence of tones in Korean means that when they are adapted, they end up sounding the same, and then within Hangul, are also spelled the same. This repeated use of the same syllable to represent more than one meaning is a key characteristic of the Sino-Korean sub-lexicon that forms the basis of the main focus of this thesis, namely, how such Sino-Korean words are represented, and to what extent semantic processing of the different Hanja characters play a role in Sino-Korean processing.

Sino-Korean words tend to refer to abstract concepts, technical terms, and institutions, such as the word 심리학 [sim.ri.hak], meaning 'psychology', which comprises three morphemes, each of which refers to a concept based in meaning of a Hanja character: 心 ('mind') 理 ('logic') 學 ('study'). The constituent meanings of the individual characters combine to create the meaning of 'psychology'. Approximately

[9]There is a possible total of 12,800 syllable configurations that can be generated based on the number of phonemes in Korean, but only about 2,000 syllables are used (Taylor, 1997). It should be noted that not all 12,800 may be permissible syllables given the phonotactic constraints of Korean. Park (2009) calculates a possible 11,172 syllables in Hangul and reports that only about 2,305 are permissible, of which, only about 808 are, in fact, in use.

57% of the Korean lexicon is Sino-Korean (Shin, Kiaer, & Cha, 2013; Chang, 1996).[10] It is important to clarify that the 57% of the Korean lexicon referring to Sino-Korean words are types (i.e., dictionary entries), rather than tokens, as grammatical function words are frequent and of native Korean origin. In fact, of the 140,464 listed words, compiled and published by the Korean Language Society in *Keun Sajun*, or *The Great Dictionary*, 81,362 words are of Chinese origin (Woong, 1983). The influence of Chinese and the lasting consequences on the Korean language are in part similar to the effects of the influence Romance languages have had on English. However, the extent to which such influences have been integrated fully into the native language is varied, as C.-H. Kim explains:

> "But there is a great difference between the two cases. In the case of English and French, both belong to the same language family, and the marriage of the two met little difficulty and produced no mutation. In the case of Korean and Chinese, however, the matrimony involved languages belonging to two entirely different stocks. This created a situation of strange bedfellow, and to this day, consummation has not been achieved" (1983: 25).

The two sub-lexicons of pure and Sino-Korean have resulted in two sets of numerals and certain common words and noun phrase constructions. Table 2.4 lists exemplars of this duality that characterizes almost the entire structure of Korean (C.-H. Kim, 1983).

Table 2.4: The duality of the lexicon, with parallel forms in Native and Sino-Korean; the corresponding Hanja characters are provided alongside the Sino-Korean words.

Native Korean	Sino-Korean	semantic meaning
목숨 [mok.sum]	생명 [sɛŋ.mjʌŋ] (生命)	'life'
물결 [mul.kjʌl]	파도 [pa.to] (波濤)	'waves'
사람 [sa.ɾam]	인간 [in.kan] (人間)	'person'

Today, in South Korea, only 1,800 Hanja characters (out of a possible 50,000) are used day-to-day, in educational and common reading materials.[11]

[10] In informal speech and non-scholarly writing, native words potentially outnumber Sino-Korean words (Taylor, 1997).

[11] Most Korean names can also be written in Hanja characters, which would bring the total of 1,800 to 2,854 Hanja characters that are in use.

As is done in Table 2.4, Hanja characters can be provided in parentheses to provide semantic clarity, especially for idioms or abbreviations, but equally as well for common words and phrases that are of Chinese origin. Recall that grammatical morphemes cannot be written in Hanja characters. While Hanja is not the primary script in printed media, the proportion of Hanja in a text varying greatly, from none to over 50%, depending on the subject matter, author, and the level of proficiency of the reader, having Hanja in text may improve ease of reading, by marking the critical words in the text (Taylor, 1997). While in many cases Sino-Korean roots have become so nativized that even those without any proficiency of Hanja can understand the Sino-Korean lexicon (Kim-Renaud, 2009), more complex forms may require the presentation of Hanja characters next to the Hangul for added semantic clarity.[12]

2.3.2 Encoding multiple meanings in Sino-Korean

Kim-Renaud (2009) explains that although some Sino-Korean words do not require explicit knowledge of Hanja characters to understand the meaning of the word, more complex Sino-Korean words require some awareness of Hanja characters.[13] This is especially the case given the repeated use of syllables in the Sino-Korean sub-lexicon, where multiple different Hanja characters and their semantic contributions can map onto a single Sino-Korean form. As such, despite ongoing government policies to decrease the use of Hanja characters, they continue to be taught as part of the

[12]Hittite cuneiform presents an interesting parallel to the Korean writing system in terms of a heterographic lexicon, particularly from a typological standpoint (W. de Melo, personal communication, 20 October 2017). Hittite, which is written in a native syllabic script, also makes use of a number of logograms from Sumerian (*Sumerograms*) and Akkadian (*Akkadograms*, written syllabically), thus creating a complex writing system with graphic components of three different non-related languages (Sumerian, Akkadian, and Hittite). Akkadograms and Summerograms often alternate with syllabic Hittite spellings in duplicate texts, which suggests that these logograms were used as a form of rebus writing, showing graphic variants of the Hittite words actually pronounced. The same can be seen by the common practice of following a Sumerogram with a phonetic complement functioning to indicate grammatical endings. For example, the Sumerogram DINGIR, meaning 'god', may be followed by the Akkadian phonetic complement *LIM*, conventionally transliterated superscript DINGIR^{LIM}*-na-aš*, to write the genitive singular of the word for *god* in Hittite, *šiunaš*. Another example of Hittite's use of Sumerian was the borrowed practice of determinatives prefixed to words and names for purposes of classification by semantic category, often useful in disambiguating homophones. They, too, were conventionally transliterated superscript and conveyed categories like male person (^{m} or ^{i}), female person (^{f}), god/goddess (^{D} which was an abbreviation for ^{DINGER}), city (^{URU}), as well as many others. It is important to note that these determinatives were not pronounced. Such examples in Hittite illustrate that the use of additional characters not native to the language for purposes of semantic disambiguation extends far beyond the remit of Korean and Japanese, having typological parallels in historical Indo-European languages.

[13]For example, no Korean speaker necessarily needs to know that the Sino-Korean word 시간 [si.kan] derives its meaning from the Hanja characters 時 ('occasion') and 間 ('between') to understand that 시간 [si.kan] means 'time' (Kim-Renaud, 2009: 70).

educational curriculum (almost uniformly), starting from as early as the third grade in elementary school (Yoon, 2017; Chang, 1996).[14] Thus, given the composition of the Korean lexicon and the extent to which Chinese influence is entrenched, the question of how the two scripts interact in lexical processing remains unanswered, particularly given the multiplicity of meanings that potentially arise for syllables in the Sino-Korean lexicon. Consider Figure 2.16, which shows the Sino-Korean mono-syllable morpheme [hak] mapping onto two different meanings ('learning', 'harm'), based on two different Hanja characters.

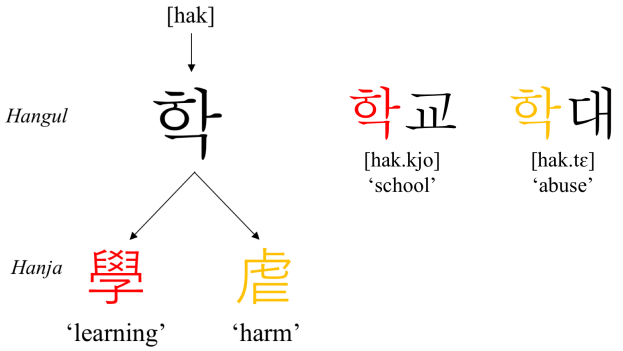


Figure 2.16: The semantic mapping of [hak] from morpheme to full word. Identical homographs in Korean map onto different Hanja characters; however, although identical, each homograph of [hak] encodes a different meaning, and each lexical instantiation is semantically related to one of corresponding Hanja characters presented on the left (by color).

In both instantiations of [hak], the form and the phonetic shape are identical, but they are attributed different meanings. The Hanja characters, unlike the shared syllable representation in Hangul, are different in form. Figure 2.16 also demonstrates that when Sino-Korean morphemes like [hak] are lexically instantiated, the full word bears a meaning semantically related to the constituent meaning (right).

How does the mental lexicon represent meaning in a bi-scriptal language? Korean offers an interesting lens through which to investigate such a question, as one of its two scripts is native to the language, and the other is not. This duality of scripts within one writing system arises from Korean's early use of Chinese logography (known as Hanja) before the innovation of its own native script (Hangul). As

[14]Other than English, Japanese and Chinese (beyond literacy in Hanja characters) is also often taught as a foreign language in schools in Korean. Depending on the school, however, European languages like French, Spanish, Italian and German may also be taught, but this is considered a special curriculum.

logograms, Hanja characters directly encode a particular concept or meaning. Use of such characters resulted in numerous Chinese loans into Korean, which spanned over a thousand years until the early twentieth century, when Hangul was officially promulgated (Lee, 1983; H.-G. Kim, 1983; Kim-Renaud, 1997). This means that not only has the number of borrowed morphemes been vast, but the influence of Chinese loans has also become deeply entrenched in the Korean lexicon. Yet, it remains unclear how the two written sub-systems interact during processing, and whether lexical processing of Sino-Korean morphemes, like [hak], as shown in Figure 2.16, is affected by the semantic multiplicity due to the different Hanja characters. Does the representation of [hak] encode both meanings, 'harm' and 'learning', and does lexical processing of words which instantiate [hak], like [hak.gjo], meaning 'school', necessarily probe both possible meanings of the morpheme [hak]? Before such questions can be properly addressed, it is important to revisit some of the key literature available on lexical processing and multiple semantic access. Thus, the following chapter will provide an overview of different models of lexical access and survey the literature on non-Roman script processing, focusing on Chinese, Japanese, and Korean.

3

Non-Roman script processing, cohorts, and the bi-scriptal mental lexicon

Contents

3.1 Cohorts in lexical processing 31
3.2 Semantic processing 37
3.2.1 Effects of semantic transparency on processing 37
3.2.2 Spreading activation and semantic network theory . . . 40
3.3 Multiple semantic access 44
3.3.1 Initial meaning access 45
3.3.2 Initial meaning selection 47
3.3.3 Semantic competition 48
3.4 Processing in a bi-scriptal lexicon 50
3.5 Processing of non-Roman scripts: Chinese, Japanese, and Korean 52
3.5.1 Semantic processing of Chinese characters 52
3.5.2 Script processing in Japanese 59
3.5.3 Processing Hanja and Hangul 68
3.6 Current models of the Korean mental lexicon 72
3.7 Main research objectives 75

A central issue in speech recognition is how information is stored in the mental lexicon. Of particular interest is how words are represented in relation to other words, a topic that has received much attention within the field of psycholinguistics, but a particular aspect of this which remains under-explored is how words that are connected via the writing system of a language are represented. Through Korean, this research focuses on how semantic information encoded across two different scripts is stored, and how such cross-script semantic relations affect the recognition

process. Cross-script semantic processing has yet to be properly investigated in psycholinguistic literature, largely due to the fact that languages commonly studied in the field do not have this particular feature. As a result, there is limited research which directly contextualizes the present work. However, the literature on multiple semantic access and the different models of lexical processing will prove useful and provide an important theoretical backdrop for the questions addressed in this thesis.

The following chapter draws on the existing literature on script processing in Korean, Japanese and Chinese. The discussion on Korean will introduce some of the current models of the mental lexicon as well as the available literature on the processing of Hanja characters, which, although within the psycholinguistic domain, tends to focus more on literacy and effects of reading proficiency, and has less to do with a proper treatment of how the different contributions of the two scripts are connected in the structure of the mental lexicon. The models of the Japanese and Chinese mental lexicons are more advanced than those available for Korean, so some discussion of these models will be offered, as Japanese shares a similar history with Chinese characters, and understanding how Chinese characters are processed will give important insight into how Hanja characters are processed in relation to Sino-Korean. Finally, this chapter will also provide a description of some of the relevant models of lexical processing, with a special focus on semantic processing, while also bringing into the fold some of the literature on bi-scriptal lexicons.

3.1 Cohorts in lexical processing

The way in which words are retrieved from the mental lexicon is an issue that has long been the subject of focus in the literature on lexical access and processing. Questions relating to the recognition of words in the lexicon have been addressed in both the visual (e.g., Coltheart, Davelaar, Jonassen, & Besner, 1977; Forster, 1976; Henderson, 1982; Taft, 1985) and the auditory domains (Cole & Jakimik, 1980; Marslen-Wilson, 1984; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; Tyler, 1984). Given the multiple meanings that can be encoded by a single Sino-Korean morpheme, of most relevance here is the literature on the Cohort Model

(Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980) of on-line spoken word recognition. Although the notion of cohort as first presented in the Cohort Model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980) is rooted in phonological processing, the idea of multiple activation to create a cohort of candidates lends itself well to the present research. As Chapter 2 demonstrated, due to the repeated use of the same sequence of sounds to represent many meanings, Sino-Korean one-syllable morphemes can map onto multiple different meanings, each of which is based on different Hanja characters. This effectively means that for any given Sino-Korean morpheme, due to the various semantic contributions of the different Hanja characters, a cohort of possible meanings could potentially be activated during lexical processing of Sino-Korean words. Before applying the notion of cohort to the present research, I briefly introduce here the Cohort Model, as well as subsequent models which also incorporate the notion of cohort, such as Shortlist and Shortlist B.

Auditory stimuli can only be presented to the listener in a strictly linear fashion. This means that, unlike in visual word recognition where incoming sensory information is presented holistically, presentation of disambiguating information unfolds incrementally. In auditory word recognition, listeners are inherently constrained by the transient nature of the incoming signal. However, although access routes may vary depending on modality of processing, it is important that for purposes of parsimony both converge on the same representations which ultimately contain the same information about the lexical entry. The Cohort Model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980) suggests that a word is recognized via a successive process of elimination, whereby the number of possible lexical candidates is reduced as each new phoneme is perceived. This idea rests on the notion that the onset of phonemes simultaneously activates a group – or *cohort* – of words matching that initial sequence.

When only one candidate remains in the cohort of possible contenders, the word is recognized; in other words, the mental lexicon activates multiple lexical candidates, from which we resolve, based on context and the incoming acoustic

signal, which is the intended output. Candidates are eliminated as soon as they cease to match the incoming speech signal, regardless of the contextual fit. Much of the early empirical investigations into auditory word recognition provided evidence supporting activation of a set of cohort-compatible candidates during sentential processing (Cole & Jakimik, 1980; Marslen-Wilson, 1984; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; Tyler, 1984). Zwitserlood (1989: 36), for example, found that even highly predictive sentences such as *With dampened spirits the men stood around the grave. They mourned the loss of their cap...* activated both the predicted word *captain*, as well as other phonologically matching candidates, such as *capital*, despite being contextually unsupported.

Previous research by Burton, Baum, & Blumstein (1989) has challenged the role of top-down effects, indicating that contextual effects in speech perception are primarily observed in perceptually degraded stimuli. Thus, the Cohort Model in its revised implementation takes into account the effects of context, while still maintaining a strict linearity in the recognition process whereby possible candidates are activated and appropriately eliminated as more perceptual information becomes incrementally available (Burton, Baum & Blumstein, 1989; Burton & Blumstein, 1995). Lexical candidates compete for selection until the only remaining activated lexical candidate matches the signal, the point which is known as the uniqueness point (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1984; 1990; Tyler, Marslen-Wilson, Rentoul, & Hanney, 1988; Schriefers, Zwisterlood, & Roelofs, 1991; for more recent discussion cf. McQueen & Cutler, 2001).

Crucially, the relative size of a cohort, whether phonological, morphological or semantic, that is activated during word recognition will depend on the number of competing lexical candidates and therefore entries. Because Korean encodes semantic information across two different scripts, from the borrowed Chinese characters to the native Hangul script, the semantic cohort of activated candidates encodes, to some part, a bi-scriptal relationship; the cohort is composed of not only the semantics, but also of important cross-script relations. We are interested in the notion of semantic representations as it relates to scriptal information, and whether

co-activated semantically and scriptally related candidates share a representation, or are stored under separate entries that partake in a process similar to phonological cohort competition (as observed in Zwitserlood 1989).

More recent conceptualizations of the Cohort Model assume a more probabilistic approach to the process of selection and elimination of lexical candidates, rather than an all-or-none approach (cf. Weber & Scharenborg, 2012 for review of models of spoken word recognition). Combining aspects of feed-forward modelling, Shortlist (Norris, 1994) posits a two-stage model in which the process where lexical candidates are generated is separate from the process during which the candidates compete. An exhaustive serial lexical search takes place during the first stage, which yields a shortlist of maximally 30 candidates that match the input processed thus far. These candidates are then hard-wired into an interactive-activation network, at which point the candidates begin to compete. In this interactive-activation network, words that are receiving support from the same section of the input are also linked to inhibitory connections, inducing competition. Figure 3.1 presents a schematization by Weber and Scharenborg (2012) of the recognition process for the word *sun* within the model of Shortlist.

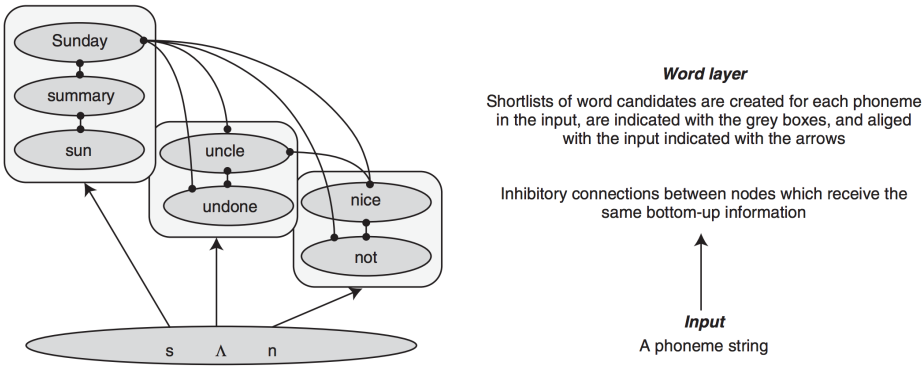


Figure 3.1: For each time point during which a new phoneme becomes available, a new shortlist (gray box) is created. Lexical candidates that overlap with one another at any position compete with one another. For *sun*, this would mean that all candidates inhibit one another (Weber and Scharenborg, 2012).

Activation of candidate words is modulated by the degree to which they fit the input, where lexical activation diminishes with mismatching information. The words receiving the highest activation inhibit candidates with lower activation during

competition, until the candidate with the best match is recognized. With each new phoneme that becomes available, the process is repeated, such that there is a separate shortlist and word layer for each incoming phoneme.

The Shortlist B (Norris & McQueen, 2008) model, as a later instantiation of the Shortlist model (Norris, 1994), proposes that auditory word recognition is based on the probability distribution of acoustic signals, whereby the likelihood of each incoming phoneme is predicated based on all phonemes that have been previously processed, independent of word-internal structure. Norris and McQueen (2008) suggest that the structure of the mental lexicon is based on whole words rather than individual constituents. Objections to such a conceptualization pertain to the issue of redundancy that is caused as a result of holding separate entries for morphologically related words (e.g., *cover*, *uncover*, *covered*; Wurm, 1997), but it has also been pointed out that such objections are unwarranted given the capacity of the human brain (Bybee, 1988; Sandra, 1994).

Despite this work, studies on the construct of cohort have yet to explicitly account for semantic representations and incorporate processing at the script level. Visual word recognition, in particular, involves various levels of representation that connect the visual-perceptual level to the semantic level, whether by orthographic or phonological routes in between. Most models of visual word recognition importantly assume that a visual word activates not only its own representations, but also those of similar words (Bowers, Davis, & Hanley, 2005). McClelland & Rumelhart (1981) explain that effects of cohort size and neighborhood density can be attributed to activation of a word in turn activating representations of matching letter units and those letter units also activating other words comprising these letters.

Due to overlap, activation of a specific word would also activate not only word units corresponding to that word, but also word units belonging to the word's orthographic neighbors. Thus, words with many more orthographic neighbors would result in activation of more word units than those with fewer neighbors, and activation of individual word units would build up to larger lexical activation. This explanation would also account for the facilitative effect found in visual

word recognition that is less consistently observed in spoken word recognition. By activating similar items that overlap in form, the additional visual mappings from orthographic information, not equivalently available in phonological representations, open up more channels for lexical access.

Several behavioral studies have attempted to propose a semantic origin to account for observed effects of orthographic cohort size, finding evidence for orthographic neighbors activating their respective semantic representations (Boot & Pecher 2008; Duñabeitia, Carreiras, & Perea 2008; Pecher, Zeelenberg, & Wagenmakers 2005; Rodd 2004). In a study using ortho-phonologically mediated associative priming, Duñabeitia et al. (2008) find that words are able to activate their ortho-phonological neighbors up to the semantic level. They used pairs like *oveja*–MIEL, (sheep-HONEY) wherein *oveja* is an ortho-phonological neighbor of *abeja* (meaning 'bee'), which is an associate of MIEL. In their experiment, participants responded significantly faster when ortho-phonological neighbors of associates preceded the target item, than when unrelated control words did so. The effect size was found to be as large as that of the effect with direct associative priming (e.g., *abeja*–MIEL).

For the aims of the present research, the aforementioned models offer an important theoretical construct that can help to account for how the different semantics of Hanja characters can be represented. In the same way that a cohort of lexical candidates is activated upon every additional phoneme that is heard, one way to conceptualize lexical processing in Sino-Korean is to consider the potential for a cohort of possible meanings to be activated for each constituent Sino-Korean morpheme. This might be the case particularly during auditory processing of Sino-Korean words, where individual morphemes are heard individually, rather than holistically perceived as in writing. As Chapter 2 demonstrated, the semantic contributions of Hanja characters are directly manifest in the meanings of the constituent Sino-Korean morphemes, which, in turn, contributes to the meaning of the full word.

For example, the word *marriage* is 결혼 [kjʌl.hon], where the first syllable means 'tie or knot', and the second syllable means 'marry', these meanings having been

derived from the original Hanja characters 結 and 婚, respectively. There is a degree of semantic transparency in the compositionality of the meaning of the Sino-Korean word, where each morpheme contributes directly to the composite meaning. In their seminal study, Marlsen-Wilson et al. (1994) investigated these effects of semantic transparency on lexical processing, the findings of which are presented and discussed in the following section. Their findings might help to inform how semantic transparency in Sino-Korean, due to the semantics of Hanja characters, affects processing, providing insight into the structure of the Korean mental lexicon.

3.2 Semantic processing

Having introduced the idea of cohort and its relevance to the semantic processing of multiple meanings in Sino-Korean, we now discuss in greater detail possible effects of semantic transparency within the larger context of semantic processing, drawing in aspects of semantic network theory and spreading activation. These individual strands will provide important background for understanding how the different semantic contributions of different Hanja characters are linked via the shared Hangul representation.

3.2.1 Effects of semantic transparency on processing

Marslen-Wilson et al. (1994) conducted a series of six experiments using a cross-modal paradigm, the motivation of which was to explore the relationship between an auditory prime and a visually presented target. They believed that the change in modality from auditory to visual presentation of stimuli would allow for better examination of more of the abstract representations involved in processing, the idea being such representations would in theory be representative of our endowed general linguistic knowledge. In addition to modality of presentation, Marslen-Wilson et al. (1994) also varied the level of semantic transparency in the relationship between the prime and the target. For example, suffixed derived words such as *government* and *govern* are semantically transparent, while *department* and *depart* are not. Their findings revealed semantic transparency to be a necessary condition for priming,

which meant that words such as *department* were unable to prime *depart*. This was taken as evidence to suggest that these two items were not stored under the same entry in the mental lexicon. That is, the lexical entry for *department*, in this case, would lend itself to the derivation of the semantically related derivates like *departmental*, while morphologically related derivates such as *departure* would start from the lexical entry for *depart*.

Processing of prefixed forms was also modulated by semantic transparency, as observed in words like *restrain* and *strain* which also do not prime. Longtin et al. (2003) also found comparable findings to those of Marslen-Wilson et al. (1994), where only semantically transparent prime-target pairs, such as *gaufrette*, 'wafer' and *gaufre*, 'waffle', elicited a priming effect. Morphologically related pairs sharing an etymological link, or pairs displaying accidental pseudo-morphological form overlap in the absence of a real morphological link (e.g., *baguette* - *bague*) also did not prime. These findings indicate that abstract representations which are not tied to a particular modality are specified for semantic and/or morphological information, making evident the effects of semantics on modality-independent processing in cross-modal priming. Thus, it appears that semantic transparency constrains the links between lexical entries, suggesting a certain degree of specification with regard to their semantic representation.

The nature of these representations seems more abstract from a phonological standpoint, as supported by the priming effects Marslen-Wilson et al. (1994) observed between morpho-phonological alternations that were also morphologically related (e.g., *sincerity* priming its stem *sincere*). This implies that different surface realizations are mapped onto an abstract underlying representation. The combined findings of Marslen-Wilson et al. (1994) and Longtin et al. (2003) offer important theoretical context for how semantic transparency may affect the semantic contributions of constituent morphemes in the processing of Sino-Korean words. Korean's bi-scriptal system, which uses both a logographic and an alphabetic script, encodes semantic information across these two different scripts. While the semantics of the Chinese logograms are transparent, in Korean, the encoding is made opaque by the cross-script

relationship that 'hides' the semantic information behind two script layers. The logographic nature of Hanja characters means that the semantics are directly encoded, and thus always transparent. However, given that a single Sino-Korean morpheme can encode the semantics of more than one Hanja character, it is not entirely clear whether the same effects of semantic transparency as observed by Marslen-Wilson et al. (1994) and Longtin et al. (2003) would be observed in the processing of Sino-Korean.

Research on embedded words can offer further insight into possible effects of constituent meanings on processing. Seminal findings in English from Shillcock (1990) show that embedded words, like *bone* in *trombone*, are sufficiently activated during cross-modal priming. Shillcock (1990) was interested in whether participants could activate the lexical entry corresponding to *bone* upon hearing the sentence *He carefully placed the trombone on the table*. Participants were presented with a word semantically related to *bone*, like *rib*, at the immediate offset of *trombone*. As a point of comparison, he also included sentences like *He carefully placed the bone on the table*. Shillcock found that the degree to which *trombone* was able to prime *rib*, compared to the degree to which *bone* was able to prime *rib* was comparable, suggesting that the lexical entry corresponding to the word *bone* is still activated even if only heard as a constituent unit of the word *trombone*.

Shillcock's (1990) findings additionally revealed that words like *wombat* were also able to prime *ball*. The two words are unrelated in meaning, so in order to account for such priming effects, and thus faster response times, the assumption has to be made that the lexical entry for *bat* was accessed as part of the entry for *wombat*. Of course, priming is a matter of degree and what should be of interest is the extent to which an item was primed (i.e., the size and the direction of the effect), which is to say that the observation of priming is not necessarily one which is clear-cut. To take this a step further, would it then also be possible for *football* to prime *dance*, the idea being that we activate all possible meanings of *ball* – a round object or a dancing event? Evidently, we are able to activate all meanings at the lexical level, but can this extend into the sub-lexical level? Questions raised in the present research take this

notion further by reconsidering how multiple semantic activation at the sub-lexical level might affect our conceptualization of the structure of the mental lexicon.

3.2.2 Spreading activation and semantic network theory

As seen in the literature so far, priming studies have been instrumental in forming our understanding of the different relationships in the lexicon as well as their cognitive realities. The various relationships include word association (Meyer & Schvaneveldt, 1971; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Shelton & Martin, 1992), taxonomy (Lucas, 2000; McRae & Boisvert, 1998), shared semantic features (McRae, de Sa, & Seidenberg, 1997), thematic relationships (Moss, Ostrin, Tyler, & Marslen-Wilson, 1995), phonological similarity (Radeau, Morais, & Dewier, 1989), as well as even shared affective valence (Rossell & Nobre, 2004; Spruyt, Hermans, De Houwer, & Eelen, 2004). In particular, the results from the seminal study by Meyer and Schvaneveldt (1976), where they tested processing differences between semantically related (e.g., *doctor – nurse*) and unrelated words (e.g., *library – nurse*), demonstrated that the semantic memory system was directly connected to feature-level representations of words. They attributed the priming they observed between semantically related words to spreading of 'excitation' between closely related items in the semantic system. The semantic relatedness between the prime and the target yielded extra excitation which in turn lowered the amount of time required to access the target word, hence the faster reaction times in their lexical decision task.

Spreading activation is used to explain how information represented in a semantic network is accessed, and to account for why words that are related to one another facilitate access (Anderson, 1983; Collins & Loftus, 1975; Posner & Snyder, 1975). The idea is that semantic knowledge is represented as a network of nodes (i.e., words) linked by connections (i.e., relationships) through which activation 'flows' during lexical processing. The nodes represent concepts, the meanings of which the network is attempting to capture, and the links are the relationships between concepts. Such models suggest that bottom-up activation from either visual or

auditory input results in lateral activation between word-level nodes in the network. The facilitative effects that are seen in priming tasks are therefore produced by gradual bottom-up activation which combines with lateral activation flow between nodes. Relationships between words are then structurally encoded in the architecture of the lexicon. In this particular view, often formalized as *semantic network theory*, word meanings are taken as collections of associated concepts, whereby the meaning of the word is defined as all information that comes to mind when the word is heard (Collins & Loftus, 1975; Collins & Quillian, 1972; cf. Rips, Shoben, & Smith, 1973; Smith, Shoben, & Rips, 1974).

Theories of spreading activation help to account for how word meanings are encoded in the mental lexicon and to understand patterns of behavior demonstrated by people in response to words. Figure 3.2 offers a schematization by Traxler (2012) to illustrate a set of nodes and the links between them.

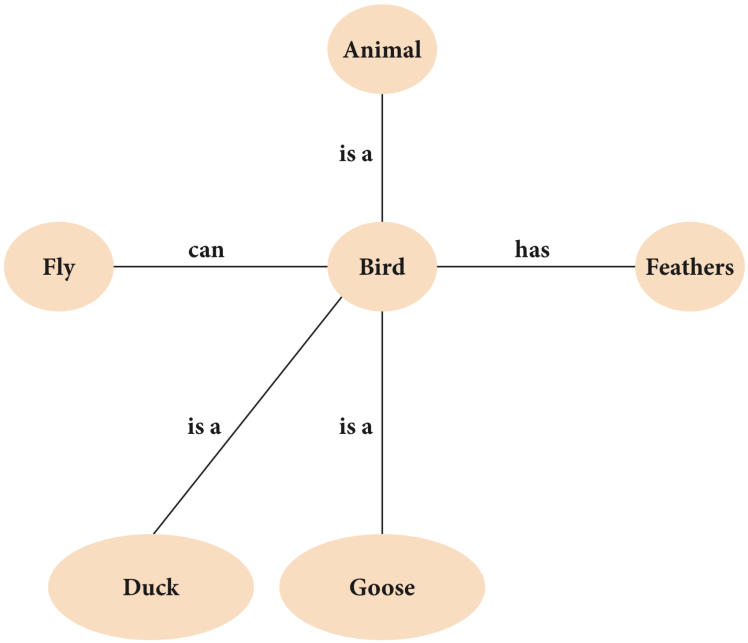


Figure 3.2: A collection of concepts operationalized to capture a word meaning and corresponding links for the word *bird* (Traxler, 2012: 83).

Figure 3.2 exemplifies a particularly important kind of link – the "*is a*" type, which encodes the relationships between general categories and the concepts that are subsumed within this category. For example, *swan* could be connected to the

waterfowl node by a unidirectional *is a* link, altogether expressing the concept that *a swan is a waterfowl*. With another *is a* link, the *swan* category could also be connected to the superordinate category node *bird*. Within semantic network theory, properties belonging to the superordinate nodes are inherited by the subordinate concepts (a swan is a waterfowl, a waterfowl is a bird, thus a swan is a bird), which means that there is no need to have to connect the specific concept of *swan* to the more overarching concept of *bird*.

Collins and Quillian (1970) found in their early work that sentences such as *A canary can fly* were able to prime responses to sentences like *A canary is a bird*, the reason for which is attributed to spreading activation from *canary is a bird* to *a bird can fly*. Other types of nodes and links can be used to represent other properties and attributes of specific concepts, such as *has* and *can* links, which can connect concepts to components (e.g., a *swan can fly*; a *swan has feathers*). Hearing *A canary can fly* involves implicitly activating the relationship *a canary is a bird* (Collins & Quillian, 1970). The meaning of a word is thus captured by the pattern of activated nodes and links.

The notion of spreading activation in effect describes the mental process by which a specific node in the semantic network is activated; when *swan* is heard, the *swan* node is activated by matching the phonological or orthographic information, after which point, this activation spreads to the other nodes that surround it. Activation of *swan* results in activation of its superordinate node, *bird*, as well as the various attributes that are connected to *bird*. Given this, there are two important properties to note: (1) spreading activation is automatic and occurs rapidly; (2) activation diminishes the further it spreads. More distantly connected nodes are less strongly and less quickly activated. Spreading activation diminishes substantially beyond one or two links in the semantic network; that is, semantic activation decays quickly.

Activation of one concept between two nodes that share many attributes in common leads to substantial activation in the set of properties that comprise the meaning of the other concept (e.g., *swan* and *duck*). The idea is that when one hears *swan*, all related and surrounding nodes, such as *waterfowl*, *bird*, *can fly*, and so on,

are also activated. Then, once *duck* is heard, concepts which have been pre-activated support faster recognition; in other words, there is less time involved in activating the part of the network that represents the concept *duck*. On the other hand, when *horse* is heard, activation spreads to nodes which are closely connected, but is substantially diminished by the time it reaches the part of the network representing concepts related to *duck*. The pattern of activation that represents the meaning of the word *duck* starts from a normal resting activation, which means that more time is needed for the network to activate the appropriate nodes and links. Spreading activation in a semantic network enables us to explain the patterns that are observed in semantic priming, where faster response times to primed words also correspond to decreased neural activity when targets are preceded by semantically related primes than when preceded by unrelated primes (Kuperberg, Lakshmanan, Greve, & West, 2008; Rissman, Eliassen, & Blumstein, 2003; Wagner, Desmond, Demb, Glover, & Gabrieli, 1997).[1] There is also often diachronic influence within a cohort, the effects of which are worth considering. This can be observed in Latin, for example, with *anas*, 'duck', and *anser*, 'goose', which together form a semantic network of connected nodes. However, over time, the latter loses initial aspiration (**hanser*, which brings into question how phonological changes over time can influence the composition of semantic neighbors which may share phonological features.

The mental process of spreading activation is considered to be automatic, and behavioral evidence indicates that word-to-word associations are activated quickly, in a way that involves no conscious effort and is outside of one's control. Neely (1977) conducted a study wherein participants were instructed to expect a particular kind of word after hearing a category label. While the category label was *body part*, subjects were informed that words referring to birds would follow the cue *body part*. Neely (1977) hypothesized that if subjects are able to control the activation of concepts, then they should focus their attention on birds immediately after hearing or seeing the cue *body part*. If subjects can control the spread of activation, concepts related to the cue *body part* (e.g., *arm*, *head*) should not be primed, and rather,

[1] It is worth noting that the notion of spreading activation is not restricted to semantics, but can also be seen in cases where there are phonological and structural overlap, particularly in rhyme priming.

members of the expected category *birds* should be primed. Neely (1977) observed no priming when he tested participants' responses to the expected targets related to the category *birds* immediately after the cue *body part*.

However, he found that concepts related to the cue *body part*, like *arm* and *head*, were primed. If there was a delay of a couple hundred milliseconds between the cue (*body part*) and the expected target, then there was priming for bird names. The explanation to account for such findings was that when subjects receive the cue (*body part*), they begin to formulate a short list of bird names they may potentially hear. The formulation of such a list requires some time, so there is no observed priming for birds immediately, but eventually when the list has been generated, there is a strong possibility of the target word being on the generated list, which, in turn, speeds up response times (Neely, 1977). The overall pattern of immediate responses to the cue, and delayed priming of bird names is consistent with fast and automatic activation spreading from the cue to related concepts, and slower, less automatic shift in attention to a list of potential bird names. Given Neely's findings, perhaps it is the case that when Korean speakers hear a Sino-Korean morpheme which is potentially underspecified for meaning, a list of potential meanings is generated, and the relative size of this semantic cohort may influence how far activation can spread to concepts which are semantically related or associated to one of the potential meanings.

3.3 Multiple semantic access

If Sino-Korean morphemes are able to encode a number of different meanings, then a topic of further interest concerns whether all meanings can be activated and thus processed at once.[2] This dovetails closely with the idea of semantic competition, especially as it interacts with spreading activation to multiple nodes. Correct selection of a meaning involves numerous cognitive processes that require successful integration of information from both contextual and lexical sources. Each

[2]The research described in this section is couched in terms of semantic ambiguity, which is not entirely relevant for the purposes of this thesis. The focus will be more on the idea of multiple semantic access, rather than the circumstances (ambiguity) which give rise to said multiplicity. The literature on semantic ambiguity resolution is relevant nonetheless for providing important background.

Sino-Korean morpheme derives its meaning from a potential cohort of multiple meanings due to the semantics of different Hanja characters which share the Hangul representation; whether all meanings from this cohort can be accessed at once remains a further question of interest in this thesis.

One challenge that arises is that it is often the case that a single word can refer to several, sometimes completely unrelated concepts, like *bark*, which can refer to the sound a dog makes or the outermost layer of a tree. Successful communication relies on using the appropriate contextual information to discern the intended meaning, while ensuring that all other irrelevant and contextually-inappropriate meanings are suppressed. How we resolve multiple meanings during semantic access has been the focus of a significant body of work in the field (Binder & Morris, 2011; Foss, 1970; Leinenger & Rayner, 2013; MacKay, 1966; Rayner & Duffy, 1986; Rodd, Cutrin, Kirsch, Millar, & Davis, 2013; Simpson, 1981; Swinney, 1979; Twilley & Dixon, 2000; Vu, Kellas, & Paul, 1998). Such works are motivated by the potential insights gained into the fundamental cognitive processes of selection, inhibition, and contextual integration. As a result of such research, evidence has been found in support of both modular as well as interactive accounts of lexical access (Duffy, Morris, & Rayner, 1988; Swinney, 1979; Twilley & Dixon, 2000; Vu, Kellas, & Paul, 1998), shedding light on the various factors that play a role in accessing and integrating meaning (Rayner, Pacht, & Duffy, 1994; Sereno, O'Donnell, & Rayner, 2006; Simpson, 1994). As the point of interest lies in whether there is exhaustive access of all semantic contributions of different Hanja characters for a particular Sino-Korean morpheme, findings pertaining to *initial meaning access* and *initial meaning selection* are discussed here.

3.3.1 Initial meaning access

Early models of *exhaustive access* contend that all possible meanings are automatically accessed during sentence comprehension regardless of contextual information, meaning frequency or dominance (Onifer & Swinney, 1981; Swinney, 1979; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979).

These studies showed that when a target was presented immediately after a word with more than one meaning, responses were faster for targets which were semantically related to either one of the available meanings of the prime, compared to those which were preceded by completely unrelated primes. Crucially, priming was found even when the context preceding the ambiguous word biased its meaning (e.g., *The accountant went to the bank ...*) and regardless of whether the biased meaning corresponded to the highly frequent ('money institution') or less frequent ('river bank') meaning. These findings were taken as evidence to support the claim that multiple semantic access is one of the first and automatic stages of resolving semantic ambiguity.

Subsequent studies, however, have revealed contrary evidence, challenging the idea of automaticity (Glucksberg, Kreuz, & Rho, 1986; Simpson & Krueger, 1991; Tabossi, 1988; Tabossi, Colombo, & Job, 1987; Tabossi & Zardon, 1993). Their studies found that in fact only one meaning is accessed; in a series of experiments, Tabossi et al. (1988, 1987, 1993) demonstrated that only the contextually appropriate meaning was primed if it was the dominant meaning, and if the context was semantically constraining. Similar findings have been corroborated by eye-movement data, showing that even in neutral contexts, multiple meanings cannot be accessed equally (Binder & Morris, 1995; Rayner & Duffy, 1986; Rayner & Frazier, 1989). The results revealed that readers were just as fast at reading ambiguous words as unambiguous words in neutral sentences when the ambiguous word had one particularly dominant meaning. Only when the ambiguous word had an equally frequent meaning did readers fixate for longer than they did on unambiguous words or ambiguous words biased towards one particular meaning. These findings were taken to disprove the claim that alternative meanings were activated equally in parallel.

The current consensus is that the extent to which all available meanings are engaged depends on the combination of contextual bias, contextual strength and meaning frequency, and the weight which is given to each of these factors varies from model to model (Duffy, Kambe, & Rayner, 2001; MacDonald, Pearlmutter, & Seidenberg, 1994; Simpson, 1994; Twilley & Dixon, 2000). How these factors

will then interact with the added semantic opacity contributed by the cross-script relationship in Korean may offer further insight. The findings presented here so far seem to suggest that all available meanings comprising the semantic cohort for a Sino-Korean morpheme have the potential to be exhaustively accessed, if factors such as context and frequency are considered equal. However, if all meanings are exhaustively accessed, the question then emerges to what extent all meanings remain activated. The point of meaning selection is discussed next.

3.3.2 Initial meaning selection

Early priming studies which found access to multiple meanings also revealed findings to suggest that listeners do not maintain these meanings for long, rather, they make a rapid selection within a few hundred milliseconds (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979). This conclusion was supported by the finding that priming was only observed in targets which were related to only one meaning of the word when the target was presented 200 ms after the ambiguous word. Such rapid initial selection was observed even for words for which meanings were balanced for frequency in neutral contexts, which indicates that listeners do not maintain all other possible meanings until the intended meaning is selected; rather, listeners integrate one meaning into their current understanding of the discourse relatively rapidly (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979). Seidenberg et al. (1982) proposed that such selection was the result of processing demands and limited capacity.

Later research, however, also revealed that initial meaning selection was cognitively demanding in itself, especially for words which have equally balanced meanings (Rayner & Frazier, 1989; Rayner & Duffy, 1986). The eye-movement data in this research demonstrated that readers fixated longer on balanced words than biased words with only one meaning when contextual information was presented at a later point in the sentence. When neither the frequency nor the contextual information can guide the selection process, these findings were taken to be indicative of processing

costs associated with meaning selection, especially in light of multiple semantic access being possible for balanced words.

Although, there is evidence to the contrary in this debate, with studies contending that multiple meanings of balanced words are maintained in working memory during sentence comprehension (Mason & Just, 2007; Miyake, Just, & Carpenter, 1994), these findings are highly contested (Twilley & Dixon, 2000; Duffy, Morris, & Rayner, 1988; Kambe, Rayner, & Duffy, 2001). Thus, the consensus that forms here is one which supports the claim that a single meaning is initially selected, even in the absence of a context biasing one meaning over the other. How then the non-selected meaning is handled, on the other hand, remains not entirely clear, with some models suggesting that they are completely suppressed (Gernsbacher & St. John, 2001; MacDonald, Pearlmutter, & Seidenberg, 1994), while others argue that they simply retain low levels of activation (McRae, Spivey-Knowlton, & Tanenhaus, 1998). This literature on multiple semantic access introduces further points of consideration in terms of how simultaneously activated meanings may compete with one another for selection.

3.3.3 Semantic competition

The notion of multiple semantic access introduces additional considerations in terms of possible effects arising from multiple meanings competing with one another, or semantic competition. How does richness or robustness of the number of semantic neighbors affect lexical processing? Can we manipulate the number of semantic neighbors in the same way that phonemic and orthographic neighbors have been manipulated before in previous studies? Depending on the language, routes to semantic access may be mediated directly by the orthography, or sub-lexically via phonological facilitation, or even in some cases, both lexical and sub-lexical paths may be available. Would additional routes bolster activation of representations of similar words? Do the representations of similar words constitute a separate semantic cohort that may in turn also have effects on processing?

Semantic cohorts have yet to be properly defined, and numerous studies have tried to manipulate the degree of richness in the number of semantic associates in various different ways: the number of semantic features (Pexman, Holyk & Monfils 2003; Pexman, Lupker, & Hino, 2002; Pexman, Hargreaves, Siakaluk, Bodner, & Pope 2008), the diversity of contexts in which words occur (Adelman, Brown, & Quesada 2006; McDonald & Shillcock 2001; Pexman et al. 2008), and the number of semantic neighbors in a semantic space as defined by corpora-based occurrences (Buchanan, Westbury, & Burgess 2001; Pexman et al. 2008). Other studies have made fruitful use of a score derived from free association norming studies, where participants are presented with a particular word and asked to note the first word that comes to mind. This score is then taken to be the number of semantic associates (*NOA*), which in turn, reflects semantic richness. Subsequent studies have thus shown that words with higher *NOA* values yield faster reaction times in lexical decision than those with low-*NOA* words (Bolota, Cortese, Sergent-Marshall, Spieler & Yap 2004; Locker, Simpson & Yates 2003; Mirman & Magnuson 2008).

An important question that emerges is to what extent semantic meanings can be probed beneath the lexical level, and to what extent this is viable across different scripts which are structurally distinct but belong to the same writing system for one language. The structure of languages such as English is not particularly conducive to such an investigation, given that oftentimes boundaries for its constituents are not particularly clear, nor do all perfectly demarcated constituents, such as syllables, yield meaningful units. It is also often the case that syllabic boundaries do not line up with morphemic boundaries. Korean, on the other hand, which uses an alphabetic syllabary, is similar to English in its use of an alphabetic system, but unlike English, has clear syllable boundaries. This will prove important for a language like Korean where the morpheme is particularly salient given the structure of its written form.

As seen in Chapter 2, the syllable typically corresponds to a morpheme, and the written conventions of Korean emphasize the morphological structure of words. This correspondence between the syllable and the morpheme is especially high in the Sino-Korean lexicon, where words tend to be compounds comprising two or more

constituent morphemes, each of which has the potential to encode the semantic contributions of different Hanja characters. Recall that due to non-exhaustive use of syllable combinations, there is repeated use of the same Sino-Korean syllable to encode different Hanja semantics. The question then emerges whether semantic access to all meanings is possible for such Chinese-derived Korean words, and if so, whether semantic access is mediated by the shared representation in Hangul. Furthermore, given the additional script-based dimension in Korean, it is not clear whether increased competition within a semantic cohort would facilitate or inhibit lexical processing.

Having set the larger theoretical stage for how semantic processing in Sino-Korean may involve semantic cohorts for individual morphemes and multiple semantic access, I now provide a brief overview of the existing literature on processing in bi-scriptal lexicons and processing of non-Roman scripts more generally, before finally moving on to processing in Korean, the main focus of this research.

3.4 Processing in a bi-scriptal lexicon

An important question that emerges is whether differences in encoding between different scripts affect language processing. Previous research has shown that reading in different scripts and writing systems recruits different cognitive resources (Nag, 2007; Almabruk, Paterson, McGowan, & Jordan, 2011; Coderre, Filippi, Newhouse, & Dumas, 2008); specifically Nag (2007) demonstrates that reading in an alphasyllabic script presents different challenges in reading compared to a phoneme-based alphabetic script. Differences in the numbers of symbols children are required to learn involves varying demands on cognitive resources. Cross-linguistic comparisons indicate that the process of reading acquisition depends significantly on the language, script, and writing system (Aro, 2006; Asfaha, Kurvers, & Kroon, 2009; Ellis, Natsume, Stavropoulou, Hoxhallari, Van Daal, Polyzoe, Tsipa, & Petalas, 2004; Georgiou, Parrilla, & Papadopoulos, 2008; Geva & Wang, 2001; Landerl, 2006; Seymour, Aro, & Erskine, 2003; Ziegler, Bertrand, Tóth, Csépe, Reis, Faísca, Saine, Lyytinen, Vaessen, & Blomert 2010; Ziegler & Goswami, 2005).

As Chapter 2 discussed, the Korean writing system comprises two different scripts, one which is alphabetic (Hangul) and one which is logographic (Hanja) (Kim-Renaud 2009; Taylor & Taylor, 1995).[3] The present research has important implications for how bi-scriptal borrowings are represented in the native lexicon and extends the existing literature on bi-scriptal processing, which is often limited to two different languages that use different scripts, rather than two different scripts belonging to one language. Previous studies on bi-scriptal processing are often within the context of bilingual processing, where the two languages involved use different writing systems (e.g., Japanese-English, Hoshino & Kroll, 2008; Russian-English, Marian & Spivey, 2003 Spivey & Marian, 1999), but the field has yet to fully examine cases where two different writing systems belong to the same language and how such a dynamic may affect lexical access.

A number of studies have explored the influence of phonology across two languages in different-script bilinguals, the findings of which broadly suggest that both languages, despite different writing systems, are recruited in processing (Bowers, Mimouni, & Arguin, 2000 for Arabic-French; Dimitropoulou, Duñabeitia, & Carreiras, 2011 for Greek-Spanish; Gollan, Forster & Frost, 1997 for Hebrew-English; Kim & Davis, 2003 for Korean-English; Nakayama, Verdonschot, Sears & Lupker, 2014 for Japanese-English; Voga & Grainger, 2007 for Greek-French; Zhou, Chen, Yang & Dunlap, 2010 for Chinese-English).

Broadly, studies focusing on the effects of script variation (Yang, Yang, & Hartanto, 2019; Liow, Green, & Tam, 1999; Hammond & Green, 1982) tend to pertain to two different languages, each having its own script which is structurally distinct from the other (e.g., English alphabet and Chinese logogram). Other studies have explored effects of transliteration and transcription on bi-scriptal processing, looking at artificial systems resulting from transliteration (e.g., Greeklish, transliterated Modern Greek using Roman characters; Dimitropoulou, Duñabeitia, & Carreiras, 2011). Each of these examples is the result of two distinct languages,

[3]The romanization of the names of the Korean scripts sometimes varies across the literature; here, we will use *Hangul* and *Hanja* to refer to the scripts.

each possessing a native script, interacting with one another, creating instances of bi-scriptal processing in bilinguals, or artificial amalgamations of the two scripts.

3.5 Processing of non-Roman scripts: Chinese, Japanese, and Korean

The following section details the literature on the processing of non-Roman scripts, focusing first on Chinese to better contextualize how Hanja characters are processed, and then Japanese, as the psycholinguistic research on Japanese scripts is relevant as the closest approximate of parallel literature exploring non-native Sino-script processing and multiple semantic access. Then I will turn to Korean, highlighting key findings from the existing work on Hangul and Hanja, while also revisiting some of the available models of the Korean mental lexicon.

3.5.1 Semantic processing of Chinese characters

The use of Hanja in Korean script processing dovetails closely with the existing literature on how morphemic ambiguity is resolved in Chinese, but it is unclear whether such patterns are transferable to Korean, given the cross-script relationship, and the added degree of opaqueness this relationship contributes. This parallel body of work comprising comparable studies on Chinese script helps to contextualize the present research, key findings of which are discussed here. Using a visual search task, Tsang and Chen (2010) found in their examination of morphemic ambiguity resolution in Chinese results consistent with previous findings showing coactivation of both dominant and subordinate meanings (Taft & Nguyen-Hoan, 2010; Taft & Zhu, 1997). Crucial to the present study is their finding showing activation of the subordinate meaning via the dominant meaning, contributing to the accruing evidence supporting the exhaustive-access assumption, whereby multi-meaning activation is viable, even under conditions where both frequency of meaning and context favor the dominant interpretation.

Given that Korean presently maintains use of both scripts, and considering the history of how Classical Chinese characters were brought into Korean, that

similar ambiguities and issues of resolution would occur at the morphemic level, albeit, discreetly, seems equally tenable in Korean. It remains unclear whether these findings could also be observed in Korean, given how Hanja semantograms contribute semantic information rather implicitly, across scripts. We predict this covertness to have some effect on semantic activation of "other" meanings, the investigation of which may also offer insight into how information that is accessed via a shared form, across different scripts, affects processing and the way in which it is stored in the mental lexicon.

Different priming paradigms have also been used in addition to the visual-world paradigm, to investigate morphemic ambiguity in Chinese, and many of these studies have yielded varying results. Tsang and Chen (2013b) found primes with subordinate meanings unable to facilitate recognition of targets in unmasked priming, while in their masked priming experiment, they find subordinate primes to facilitate recognition of targets with dominant meanings, but only if the meanings were biased towards the dominant meaning. They also tested items for which the frequencies of meanings were relatively balanced, and found no significant priming effect, and in fact, observed a trend of inhibition, which suggests that lemmas of the same morphemic form compete for activation, and that those of high frequency tend to have correspondingly higher levels of baseline activation (Taft & Nguyen-Hoan, 2010; Tsang & Chen, 2013a).

That is, even when the context favors a subordinate meaning, lemmas that are highly frequent can still be activated by any prime that shares a morpheme with it, but seemingly temporarily, as this was only observed in their masked priming study, suggesting a degree of automaticity. When prime presentation was prolonged, and more time was allowed for processing, making it much more conscious, context helped to enhance activation of the appropriate interpretation, but only when prime presentation remained relatively short was it possible for primes with subordinate meanings to facilitate recognition of targets with a dominant interpretation. Their explanation is such that when the subordinate lemma is activated, it also suppresses activation of the dominant interpretation that is doubly

unsupported by the context, thereby cancelling out the facilitation that was the result of temporary activation of the dominant lemma and the morphemic form that is shared between them (Tsang & Chen, 2013b).

In a larger theoretical context, these findings are consistent with the principles of lexical ambiguity resolution and the reordered access model (Duffy, Morris, & Rayner, 1988), which posits different meanings of an ambiguous word to be activated simultaneously, the initial strength of activation modulated by relative frequency. This means that the dominant meaning of a word (independent of relative frequency of all other possible meanings) will be more strongly activated than any of its subordinate meanings; however, a specific context that supports a more subordinate meaning may enhance activation of the meaning most appropriate to the present context.

Words for which the available meanings are balanced in frequency, on the other hand, will yield comparable levels of activation across all meanings. While we predict Korean will provide further evidence supporting multi-meaning activation, the language also offers a new lens through which to explore ambiguity and the different links that comprise the semantic network in the brain. Although arguments for particular distinctions have been presented here and in the earlier literature, the linguistic evidence under consideration is not entirely conclusive, and there is no clear consensus on how sub-lexical ambiguity is resolved and the ways in which Hanja may affect semantic access.

More broadly, a useful way to conceptualize morphemic processing in Chinese is a multi-level interactive-activation system where units of activation are hierarchically organized in levels that correspond to words, morphemes, and submorphemic parts (Taft & Zhu, 1995, 1997; Zhu & Taft, 1994; Taft, Liu, & Zhu, 1999). Figure 3.3 illustrates such a model, which presents orthographic units on the left, phonological units on the right, and the semantic units at the top. The level of "stroke" refers to the lines and dots that comprise a Chinese character. The idea of the model is that stroke units facilitate activation of the corresponding units representing radicals that contain those strokes, which in turn, feed activation to the units

representing the morphemes which contain those radicals, at which point, activation is then fed to the units representing the multi-morphemic word containing those particular morphemes.

Units at the top semantic level can be linked to units at all levels except the stroke level, as radicals, morphemes and polymorphemic words can all be associated with a meaning (although links from the radical level are not illustrated in Figure 3.3).

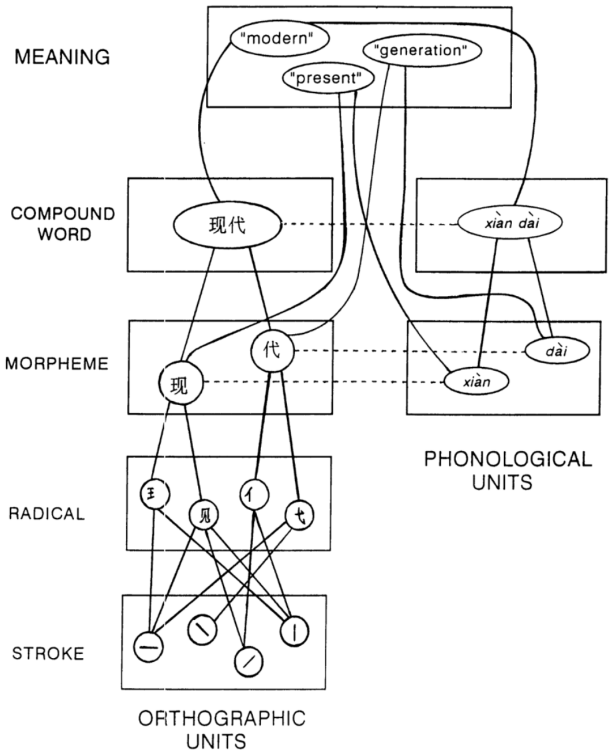


Figure 3.3: The architecture proposed by Taft et al. (1999: 95) and Zhu and Taft (1994) to account for morphemic processing in Chinese (e.g., 現代, meaning 'modern'), involving multiple pathways which interact with one another at the orthograpahic, phonological and the semantic levels.

Both orthographic and phonological representations exist at the morphemic and lexical levels. In theory, there could be phonological links at the radical level, but these are not integrated into the model, as the pronunciation of radicals is not necessarily indicative of the pronunciation of the morpheme (Fan, Gao, & Ao, 1984), meaning that phonological units at the radical level are not necessarily systematically mapped onto phonological units at the morpheme level.

During visual word recognition, Taft et al. (1999) suggest that units at the stroke level activate units at the meaning level by way of both the orthographic levels as well as the corresponding phonological units, as also corroborated by Perfetti and Zhang (1995) and Tan, Hoosain, and Peng (1995). The strength of the connection between units is modulated by the frequency with which that particular connection is used, and thus, more frequently used words are activated via stronger links, and therefore more quickly, relative to words which are lower frequency. However, there is an issue of redundancy in both the orthographic and phonological representations. Compound words in both cases are represented in full despite compositionality into constituent parts. Taft, Liu and Zhu (1999) argue that for semantically transparent compounds such as 手枪, meaning 'handgun', where 手 means 'hand', and 枪 means 'gun', it is redundant to have orthographic representations for both 手枪, as well as 手 and 枪 individually. They argue that even for semantically opaque compounds like 小心, where the constituent morphemes mean 'little' and 'heart', respectively, but combined means 'beware', it is redundant to posit an orthographic representation for 小心. More parsimoniously, the representations for 小 and 心 could individually activate their respective set of semantic features, but activate another set when combined. The same argument of redundancy applies for phonological representations.

The second issue pertains to whether homographic morphemes should have separate representations. According to the model presented in Figure 3.3, if the orthographic representations posited below the lexical level are in fact morphemic in nature, then homographic morphemes should have separate representations. [an] has two meanings ('quiet' and 'install'), for example, both of which are written as 安. They are different morphemes, and thus at the morphemic level of representation, they should be distinct. However, it may be necessary to posit a single representation for 安, in order to account for possible priming effects between 安 宁, meaning 'peaceful', and 安 装, meaning 'install'.

Given such concerns, Figure 3.4 presents a modified version, as proposed by Taft, Liu and Zhu (1999), of the model presented in Figure 3.3, whereby instead of

positing a whole-word level for orthography and for phonology, there is a set of modality-independent units, called lemmas (Levelt, 1989), linking together orthographic character units, phonological syllable units, and semantic units. By functioning as way-stations that develop when the same semantic information co-occurs with the same form information, either orthographic, phonological, or both, the lemma units account for the correlation between meaning and form and are of a similar conceptualization to that of Schreuder and Baayen (1995), who referred to them as concept nodes.

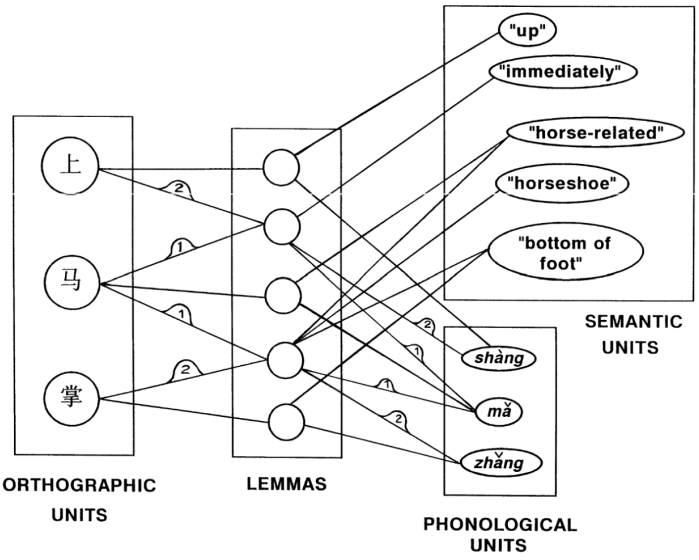


Figure 3.4: Revised model of the interactive-activation model which includes a level for lemmas, eliminating orthographic and phonological representations of complete compound words (Taft, Liu, & Zhu, 1999: 107).

Similar to distributed connectionist models of lexical processing which also propose emerging units through frequent co-occurrence (Seidenberg & McClelland, 1989), Taft et al.'s (1999) model posits a level of representation intervening between orthographic units and semantic units.

Each orthographic unit represents a character, which, in turn, sends activation to a lemma unit representing the free form that has a phonological pronunciation as well as a semantic meaning. For example, as Figure 3.4 shows, the character unit sends activation to a lemma unit that represents the free form with the pronunciation [ma] and the meaning 'horse', in addition to all other lemma units that represent

马 when combined with another character, such as 掌 [tʂaŋ]. The compound 马掌 ('horseshoe') is also linked to other semantic units which represent those that are semantically related to *horses* as well as the semantic units that are associated with 掌, which means 'bottom of foot'. This is in addition to any other semantic units that differentiate 马掌 from the string formed from a simple concatenation of 马 and 掌. This lemma unit is also connected to the phonological units for [ma] and [tʂaŋ].

The revised model eliminates the morphemic level of orthographic representation to enable homographic morphemes, such as 安, which means 'quiet', and 安, which means 'install', to be represented together in terms of orthography, but separately with regard to function. Taft et al. (1999) explain that when the meaning of a constituent of a compound word does not overlap with the meaning of the whole word (e.g., 安 meaning 'install' within 安宁, meaning 'peaceful', the 'horse' meaning of 马 in 马 上, meaning 'immediately'), the lemma that is associated with that meaning will be reset. In the case of 安宁, however, the lemma representing the 'quiet' meaning of 安 remains active.

Taft et al. (1999) assume that an orthographic character unit remains active so long as there is an active lemma associated with it, in which case, it would be possible for 安宁 ('peaceful') to prime 安装 ('install') via the shared orthographic unit despite the semantic discrepancy. However, something like 马上 ('immediately') would not prime 马掌 ('horseshoe'), because the orthographic unit for 马 would no longer be active. If these findings are applied to Sino-Korean processing, this suggests that multiple meanings of different Hanja characters have the potential be activated via the shared orthographic unit in Hangul, though it is not clear whether the activation posited by Taft et al. (1999) in their model could be sustained across an additional level of opacity posed by different scripts.

The effects of semantic transparency have also been explored in Chinese. Liu and Peng (1997) found there to be no facilitation of responses to targets that were semantically related to one of the constituents of an opaque prime, but facilitation for transparent primes. 马上, which means 'immediately' is a semantically opaque word, comprising two constituent morphemes that individually mean 'horse' (马)

and 'up' (上), respectively. The semantically opaque 马 上 ('immediately') was unable to prime the target word 绵 羊, meaning 'sheep', despite 马上 containing a constituent (马, 'horse') that was semantically related to 绵 羊 'sheep'.

On the other hand, the semantically transparent 背叛, 'betray', comprising 背 ('back') and 叛 ('rebel'), each contributing directly to the holistic meaning of the word, was able to prime a target word like 覆盖, 'cover', which while semantically unrelated to the prime 背叛 ('betray') is related to one of the constituents, 背 ('back'). These findings alongside Taft et al.'s (1999) model point towards the possibility for cross-script priming that is mediated by the shared Hangul representation. To better understand the dynamics of cross-script relations, I turn to a brief discussion of Japanese script processing, similarily to Korean, makes use two scripts, one which is native Japanese, and one which is Sino-Japanese. Such literature will further elucidate whether the contributions of Chinese, as manifest through the script, can in fact be activated via the script.

3.5.2 Script processing in Japanese

Japanese exhibits a relationship to Chinese that is comparable to that found in Korean, in that both languages have some version of a script derived from adopted logographic Chinese forms that maintains enough currency to be used alongside a more modern system. Due to such history, the literature on Japanese script processing offers a helpful parallel as context for the present research. In Japanese, the script that is equivalent in role to Hanja is known as *kanji*, where a concept or an idea is represented, and each character has multiple potential meanings only disambiguated by the context of the larger word of which it is a part. This is contrastive with *katakana* and *hiragana*, both of which are forms of the syllabic Japanese script and referred to as *kana*, wherein each kana character corresponds to a sound in Japanese (i.e., a phoneme).

There is ample literature on the character-to-sound correspondence, and whether kanji items of the same form, different sound and meaning, can activate one another (which can also depend on whether the kanji is situated in a specific

compound, or a standalone item). There is currently inconclusive evidence for whether reading Japanese kanji requires more processing time whether the characters involved have more than one pronunciation compared to when they only have one (Wydell, Butterworth & Patterson, 1995; Fushimi, Ijuin, Patterson, & Tatsumi, 1999; Kayamoto, Yamada, & Takashima, 1998), some suggesting that computation of phonology from orthography sits at the level of the whole word, rather than at that of its subcomponents, and others finding the complete opposite, that, in fact, computation can be affected at the subcomponent level (i.e., at the level of the individual kanji). Thus, the literature on phonological activation of multiple readings in kanji processing offers varying findings.

Also interested in the idea of multiple activation, Verdonschot et al. (2013) explored the possibility of a single kanji activating multiple pronunciations. Due to there being more than one possible pronunciation depending on whether the kanji is pure Japanese or of Sino-Japanese origin, the form can have multiple meanings. More than 60% of all kanji have multiple pronunciations. Consider the character for water 水; in Chinese it is pronounced [ʃueɪ3], with '3' referring to the tone, while in Japanese, it can be pronounced as the original Japanese word for water, e.g., /mizu/, or as the incorporated Chinese word for water, e.g., [ʃueɪ]. The Japanese reading is known as the KUN reading, and the Chinese-derived reading is known as the ON reading. Thus, the pronunciation can be reached via a path that is facilitated by meaning (i.e., pure Japanese), or by sound (i.e., Sino-Japanese).

Verdonschot et al.'s (2013) research investigates the types of processing costs associated with the multiple paths from orthography to phonology, especially in cases where the alternative reading is weak, taking into consideration the possibility that activation depends on the degree of salience of particular readings. Likewise, in Korean, certain Hanja-to-Hangul mappings are more frequent, thereby making certain semantic mappings more likely to be activated. To investigate, Verdonschot et al. (2013) conducted a series of studies, the most relevant of which are those where participants read aloud katakana targets preceded by masked kanji primes, the first with equal preference across possible readings, and the second with varying degrees

of preference. They posit that if native Japanese speakers are preparing multiple pronunciations when processing a masked kanji prime with two possible readings, then the naming of the transcriptions of those two readings in katakana when preceded by the congruent kanji prime should reveal facilitatory effects compared to when preceded by control primes. This is an interesting prediction, as it might also be expected that preparation of multiple pronunciations may yield inhibitory effects due to increased competition amongst the possible pronunciations.

The first study used kanji which, in compounds, yield balanced frequencies of occurrence in one or the other reading. He predicted that, if simultaneous activation of multiple readings proved possible, then those kanji which have approximately balanced readings are the most likely to induce this type of multiple activation. The target strings comprised katakana transcriptions of the pure Japanese and the Sino-Japanese readings for the kanji prime. For example, the kanji prime 町, meaning 'city' or 'block', was transcribed into katakana, once using the pure Japanese pronunciation, マチ [matɕi], and again using the Sino-Japanese reading, チョウ [tɕoː]. The findings from this study reveal that a single kanji prime can facilitate activation of multiple katakana targets, that is, both katakana strings (マチ and チョウ) yielded faster reaction times when preceded by the relevant kanji prime (町), compared to when preceded by a control prime.

The finding that both strings yielded comparable reaction times was taken by Verdonschot et al. (2013) as evidence for the prime 町 having activated both its pronunciations. There was some evidence, although not statistically significant, to suggest a stronger facilitation of the KUN, or the native Japanese, reading, which Verdonschot et al. attributed to the fact that kanji primes, when presented in isolation, tend to take the KUN reading. To confirm these findings, in the same study, they performed a follow-up experiment where, in addition to repetition primes having the same reading preference, they also included stimuli biased towards one of the pronunciations (i.e., the dominant reading occurs more frequently compared to the other reading). Based on this change, they predicted that when the selected kanji primes have a bias towards a specific reading, the prime is more likely to activate

that particular reading compared to the other (unbiased) reading, which will result in increased facilitation by that particular prime for its biased transcription target.

Verdonschot et al. (2013) found significant priming for ON readings even when the two pronunciations were of equal dominance (i.e., no strong preference bias for either pronunciation). They also found significant priming for the ON reading when the dominance was biased towards the ON reading, even when presented isolation, which recall, kanji primes tend to the KUN reading when presented in isolation. They explain that it is unlikely for the ON pronuciation to be the accurate pronunciation for stand-alone kanji primes, and thus the fact that the ON readings elicited significant priming suggests that all constituents receive activation.[4] Thus, based on the findings from the two studies, Verdonschot et al. (2013) concluded that for Japanese kanji with more than one pronunciation, both receive activation, and that the two studies confirmed parallel activation of both pronunciations.

Unlike Japanese, Korean does not have different pronunciations based on whether the word is pure Korean or Sino-Korean, so while different Hanja characters may contribute multiple different meanings, the multiplicity does not extend into different phonological readings for Sino-Korean. However, Verdonschot et al.'s (2013) findings of kanji characters activating multiple phonological readings, particularly when the readings are of equal dominance, offers an important baseline for the research presented in this thesis. Based on Japanese, it appears that multiple activation, insofar as the phonological domain is concerned, is possible for kanji forms that have both a pure as well as a Sino-based Japanese reading.

These findings of phonological co-activation suggest that individual kanji forms with more than one reading activate a potential phonological cohort comprising all possible readings. In Sino-Korean, as discussed in Chapter 2, all Korean can be expressed in Hangul, but depending on the origin of the word, certain words can be written in Sino-Korean. Thus, similar to Japanese and based on Verdonschot et al.'s findings (2013), a case could be made for multiple activation of Chinese

[4]Verdonschot et al. 2013 provide the caveat that further research using compound stimuli would be necessary to elucidate how phonological activation may differ depending on whether the character is presented in isolation or forms a constituent of a compound.

contributions in Korean, but unlike Japanese, the contributions of Chinese in Korean are meaning-based. Thus, rather than a phonological cohort of possible readings, each Korean form that encodes multiple semantic contributions from different Hanja characters carries a semantic cohort comprising these various different meanings.

Such focus on Japanese scripts, particularly kanji, underscores a more general special interest in the representation of kanji in the Japanese mental lexicon. One of the first studies to specifically consider the lexical representation of kanji words within the mental lexicon was Hirose (1992), who argued for a structure based around the first kanji constituent of a word. The idea was that words sharing the same first kanji were linked in clusters, and the first kanji functioned as a retrieval cue. Importantly, words sharing the same kanji in the second constituent slot were not linked. Hirose (1992) conducted a series a lexical decision task using constituent-morpheme priming, where two-kanji compound targets were preceded by the first-constituent, the second-constituent or an unrelated kanji. His findings suggested that while there was significant priming for both the first- and second-constituent conditions, compared to the unrelated primes, significantly greater priming was found for the first-constituent, related to the second-constituent condition. These findings led Hirose to postulate a form of lexical retrieval for compound words that involved a serial search mechanism which unfolded left to right, the first constituent working as a retrieval cue, theoretically consistent with Forster's (1976) serial search model.

Hirose's (1992) model, most importantly, does not appear to be able to account for how such arrangements based on clusters of words sharing the same first constituent interact with semantic relationships. For example, consider the native Japanese compound 青空 [ao.dzora], meaning 'blue sky', which shares the same first constituent as 青物 [ao.mono], meaning 'green vegetable', as well as the same phonological reading of the first constituent. [ao] is the native Japanese KUN reading of the kanji character 青; however, according to Hirose's model, these native Japanese compounds would also share a cluster with Sino-Japanese compounds like, 青年 [seː.nen], meaning 'adolescence' and 青銅 [sei.dō], meaning 'blue bronze',

which use the Sino-Japanese ON reading of the kanj character 青, [sei]. Another consequence of Hirose's model appears to be compounds that are semantically related but do not share the same first constituent (such as the Sino-Japanese 登山 [to.zan] and the native Japanese 山登り [jama.no.boɾji], both of which mean 'mountain climbing', being represented in different clusters.

Building from the seminal beginnings of Hirose's (1992) model, subsequent accounts have assumed a more connectionist approach, comparable to that of McClelland and Rumelhart's (1981) interactive-activation and competition model (cf. Seidenberg & McClelland, 1989; Seidenberg & Gonnerman, 2000). These connectionist models rely on the interaction between semantics and form overlap to explain priming effects observed between items which are morphologically related. In these models, morphological information does not play a role in the representation and processing of extant words; instead, any effects of morphology can be attributed to the combined effects of semantics and phonology, as an epiphenomenon. Furthermore, within these models, effects of base frequency are reduced to automatic activation of words in the lexicon that overlap significantly in form and meaning with the base word. As a result, morphologically related words will be activated.

In his lemma-based model of the Japanese mental lexicon, Joyce (2002), using a connectionist framework, draws on the structure posited by Taft, Liu, and Zhu (1999) in their conceptualization of the Chinese mental lexicon, as discussed in the preceding section on the processing of Chinese characters. Recall that Taft et al.'s model (1999) posits a layer for lemma unit representations, as abstract forms of a lexical entry, in order to mediate the connections between both orthographic and phonological representations and semantic representations. This was motivated in Taft et al.'s model in large part due to the high frequency of homographs in Chinese as well as the high level of semantic transparency of compound words, as Chinese characters are logograms and thus directly encode meaning. Joyce (2002) explains that as such issues also arise in Japanese, an approach parallel to that of Taft et al.'s is most effective.

Joyce (2002) presents in his model, which dovetails closely in structure to Taft's (1991, 1994) extensions of McClelland and Rumelhart's (1981) model, a level where lemma units are represented between the semantic level of representation and the phonological and orthographic levels of representation, crucially enabling the model to take into account both kanji and kana processing within a single model. Joyce explains that by extension of having a model that integrates processing of both scripts, the model can also represent the processing of both KUN and ON-readings of Japanese.

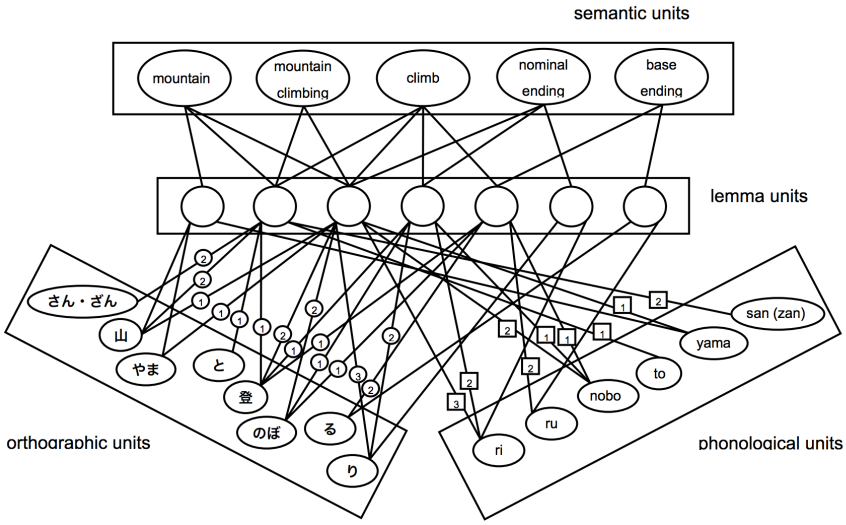


Figure 3.5: A model of the Japanese mental lexicon, based on connections mediated by lemma-unit representations (Joyce & Ohta, 1999; Joyce, 1999; 2002: 89), which shows the representation of two words which are synonymous for 'mount climbing', the Sino-Japanese 登山 [to.zan] and the native Japanese 山登り [jama.no.borji].

Figure 3.5 shows that orthographic representations link directly to the relevant lemma units, such that the kanji 山 [to.zan], the hiragana やま [jama], and the katakana ヤマ [jama] can all be linked to the meaning of the 'mountain climbing' via its lemma unit.

Both representations of the Sino-Japanese 登山 [to.zan] and the native Japanese 山登り [jama.no.borji] are assumed to activate the meaning of 'mountain climbing', but as Masuda and Joyce (2018) clarify, their lexical status as independent words is represented by the separate lemma units. The idea is that the Sino-Japanese compound 登山 and its ON reading are activated when the constituent kanji

characters 登 [to] and 山 [zan] are combined as Sino-Japanese morphemes in that specific order. For the native Japanese compound 山登り and its KUN reading [jama.no.boɾji], they would only be activated when the compound word is a combination of the native Japanese morphemes 山 [jama] and 登り [no.boɾji].

For findings such as those observed by Verdonschot et al. (2013), Joyce's (1999) model appears to be able to capture the potential intricacies that arise from different readings of Japanese, though it is unclear whether Joyce's model can account for multiple activation, as found by Verdonschot et al. (2013). Joyce's (1999) model may be able to cope with the multi-script dimension of the Japanese writing system, but whether it can account for multiple access by way of a cohort comprising potentially both the KUN and ON phonological readings depending on the kanji is not conclusive. Joyce's model, unlike Hirose's (1992), which made use of a serial search mechanism, assumes a mechanism of spreading activation as the main underpinning of lexical retrieval, especially for two-kanji compound words. More specifically, Joyce (2002) explains that for compound words comprising two kanji characters, there is parallel activation from both orthographic representation units for its constituent kanji to the compound word's lemma unit, thereby activating the word. Furthermore, Joyce posits that Hirose's observed effects of constituent-morpheme priming can also be accounted for by his model by assuming that any remaining activation within the network of lemma units linked to the orthographic units of constituent kanji facilitates activation of those lemmas, leading to faster response times compared to unrelated primes.

Based on Joyce's (2002) model, there seem to be a number of useful parallels in the structure of the Japanese mental lexicon that share the potential for application in a model of the Korean mental lexicon. For example, a lemma-based level may prove useful when trying to account for the semantic multiplicity of Sino-Korean morphemes. Each Sino-Korean morpheme could be an individual morphemic lemma on which the semantics of different Hanja characters can be mapped. What still remains unaccounted for within the structure of Joyce's model is exactly how the different semantics of Hanja characters are represented. As Chapter 2 explained,

almost every Sino-Korean morpheme encodes the semantics of not just one Hanja character, but of many, which means for each morpheme, there is potential semantic cohort of meanings competing for activation. There is no particular level within Joyce's model where such contributions could be represented. One crucial difference between the Japanese and the Korean writing systems is that Japanese makes more integrated use of both native Japanese and Sino-Japanese, meaning that both kanji and kana are used relatively frequently. This is not the case in Korean, where Hangul remains the primary script, with Hanja characters used less frequently.

Furthermore, while all Korean words can be written in Hangul, not all can be written in Hanja, which means that any model of the Korean mental lexicon may not be straightforwardly holistic, but may require a concerted effort to integrate the two sub-lexicons of pure Korean and Sino-Korean. Crucially, however, Joyce (1999; 2002) notes that his model reflects the fact that the knowledge of the different KUN and ON readings within the adult native-speaker is largely dependent on usage frequencies, hence the rich network of links from phonological access representations to lemma units.

Joyce's point regarding the effect of individual differences across adult speakers on reading ability draws attention to an important consideration for any model of the Korean mental lexicon as well, as proficiency in Hanja characters may be varied, and thus how these characters are represented in the mental lexicon may not be directly tied to the script itself, but rather the concepts that the characters evoke modulated by proficiency. The models that have been discussed so far in this chapter of both the Chinese and the Japanese mental lexicons appear to converge on the need for some lemma-based level, an aspect that may be necessary for a model of the Korean mental lexicon, given the degree of overlap between Japanese and shared use of Chinese characters. To better understand how the status of Hanja characters during Sino-Korean lexical processing is understood, the final focus of this literature review will be on the existing work on script processing in Korean and current models of the Korean mental lexicon.

3.5.3 Processing Hanja and Hangul

Building on the literature on Sino-Japanese processing, a comparable question that emerges is to what extent do Hanja characters play a role in the processing of Sino-Korean words? Despite more than 60% of the Korean lexicon comprising Sino-Korean words, studies have yet to properly delimit the contributions of Hanja characters in processing. This area of Korean psycholinguistics is relatively under-researched, with the current state of the literature focused more heavily on literacy, rather than linguistic processing. While research on the effects of Hanja literacy on the processing of Hangul is important in its own right, there still exists a gap in the theoretical background of Korean psycholinguistics regarding a fully-fledged account of its mental lexicon. This is not to say that the structure of the Korean mental lexicon differs drastically from those of other languages and deviates from current psycholinguistic findings on the mental lexicon, but that Korean offers certain idiosyncrasies that require more careful treatment.

Although a comparable literature on the processing in Japanese, which shares a similar history of Chinese influx, (Okano, Grainger, & Holcomb, 2013; Feldman & Turvey, 1980; Iwata, 1984; Yamada, Mitarai, & Yoshida, 1991; Yamada, 1992) there is limited literature of the equivalent in Korean. The idea that the two scripts interact in some meaningful way is simply presumed with no detailed understanding of how this may be unfolding. Much of the research on processing of languages with different scripts has focused on bilinguals; the interest primarily lies in understanding how a bilingual speaker's two languages are interconnected, and by virtue of this set-up, how their two scripts, should they vary in orthographic system, interact.

Studies show that three constituent processes underlie reading and visual word recognition across writing systems: orthography (O), phonology (P), and semantics (S) (Coltheart, 1980; Lesch & Pollatsek, 1993; Lukatela & Turvey, 1991; Daneman & Reingold, 1993). Previous psycholinguistic investigations on Korean scripts focus primarily on the varying degrees to which these factors play a role in the larger process of visual word recognition, and how they relate to one another (Cho & Chen, 1999; Simpson & Kang, 1994; Kang & Simpson, 2001; Kim, Taft, & Davis,

2004; Shen & Forster, 1999; Frost, 1998; Grainger & Ferrand, 1994; Grainger, Diependaele, Spinelli, Ferrand, & Farioli, 2003).

As it combines more than one type of system in its available scripts, Korean is of particular interest, lending itself well to this pursuit of teasing apart the phonological, orthographic and semantic factors affecting processing. Hangul, an alphabetic system, which encodes phonological information directly, is assumed to access meaning by a phonologically mediated route, while Hanja, a logographic system, which directly represents meaning but not sound, is thought to activate meaning directly from the orthography, rather than through a phonologically-mediated path (Nam, 1995; Park, 1995; Cho & Chen, 2005; Simpson & Kang, 1994; Park, 1990). These studies show that Hanja and Hangul play an interconnected role in processing at all different levels, motivating our research to determine the finer-grained details of how they interact.

Simpson and Kang's (1994) study provides some insight into how the two scripts may affect one another, looking specifically at how context of neighboring scripts may affect processing. They compared naming latencies for high- and low-frequency Hangul words embedded in a list that also contained other Hangul words (apart from the high-/low- frequency words), Hanja words, or Hangul non-words. Their findings showed a frequency effect for alphabetic Hangul words, which was affected by the composition of the list in which the words were embedded. That is, if the Hangul words were embedded in a list consisting predominantly of Hanja characters, they observed a large frequency effect, and if the Hangul words were embedded in an experimental list comprising primarily Hangul words, the effect diminished completely. Simpson and Kang (1994; Kang & Simpson, 2001) used these findings to show a Korean reader's ability to modulate their use of the lexical and sub-lexical pathways depending on the experimental environment.

The idea of lexical and sub-lexical routes originates from the ongoing question of whether word recognition is based on solely the visual analysis of the input or mediated by phonology (Seidenberg, 1985). Dual-route theories of reading (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, 2005; Pritchard, Coltheart, Palethorpe, & Castles, 2012) assume two possible routes to visual word processing, one which is

more direct, where access to meaning proceeds directly from orthography, is termed the lexical route, and the other, which is mediated by phonological information, is termed the non-lexical or sub-lexical route. The lexical route activates phonological information by way of matching orthographic input to a known word that is stored in the mental lexicon. The phonologically-mediated, sub-lexical pathway involves activation of a phonological representation in order to facilitate semantic access. The path is considered sub-lexical as it entails assembly of a phonological code based on grapheme-to-phoneme correspondence rules. Structurally different writing systems may recruit either a lexical or a sub-lexical route, or a combination of both (Frost, Katz & Bentin, 1987), and whether one plays a more dominant role relative to the other can be modulated by such effects as word frequency (Jared & Seidenberg, 1991; Seidenberg, 1985) and levels of reading proficiency (Cho & Chen, 1999; Doctor & Coltheart, 1980; Rusted, 1989; Kang & Simpson, 1996).

Since Hanja is likely to be processed via the lexical route, the presentation of a large number of these words causes the reader to be biased towards the use of this lexical path for all words, and thus a frequency effect for Hangul is observed (Lee, Lee, & Kim, 2016). That is, within the context of our research, it seems that the Hanja interferes by biasing the reader towards a particular processing pathway, and as a result, the inhibitory effects causes those Hangul words with higher frequency to be named faster, as those with higher frequency may be less susceptible to inhibitory effects by virtue of being highly frequent words. Likewise, because phonological information is easily accessible from Hangul, these words can be processed very rapidly and accurately via a sub-lexical route, mapping graphemes onto a corresponding phoneme.

Thus, when all words are listed in Hangul in the list, this particular route dominates, and subsequently no effects of frequency are observed. Simpson and Kang's findings are pertinent to our research, as they highlight an interaction between the scripts in processing as well as the readers' sensitivity to script changes. Something to consider in the future is whether the items in the list shared any semantic relationship; it may be the case that regardless of which of the two

pathways is recruited for processing, given that semantic effects are very robust (Meyer & Schvaneveldt, 1971; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Shelton & Martin, 1992), the fact that two items are semantically related may elicit faster response anyway, making it difficult to discern whether such findings are in fact due to phonological information being more or less accessible.

Simpson and Kang's studies (Simpson & Kang, 1994; Kang & Simpson, 2001) are representative of the larger investigation on Korean visual word recognition (Park 1999; Cho & Chen 2005; Cho & Chen 1999; Park 1996; Park & Iyotsna 1995; Tavassoli & Han 2001; Yi 1993; Park 1995), which conventionally rests within the perimeters of exploring the implications of this dual-route model, wherein two pathways, one lexical and the other sub-lexical, are assumed to be possible (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Seidenberg, 1985). These studies hint at the inter-relatedness of the scripts, underscoring the ways in which the presence or activation of one can affect the other, but do not offer a detailed account of their relationship.

Previous research by Cho and Chen (1999) also suggest that the role phonology plays in semantic processing was significantly dependent on the script involved. This is in line with previous findings that also confirm that phonological activation during semantic access is heavily script-dependent (Chen, Flores d'Arcais, & Cheung, 1995; Katz & Frost, 1992). These findings support the view that the pattern of lexical processing and memory are modulated by such factors as proficiency in reading and properties of the orthography, which is consistent with models that posit dual-routes, rather than those which suggest either phonological mediation or direct access to be important to visual word recognition.

Subsequent studies have built on these results, with some finding additional effects of script-switching (Cho & Chen, 2005). Others have carried out a closer examination of the pathway from orthography to phonology, and the varying degrees to which these aspects can be considered autonomous in lexical processing, particularly in Korean (Kim, Taft, & Davis, 2004), given the relationship to both Hanja and Hangul. Some of this research has also been couched in terms of its

importance to further dedicated research on reading cross-orthographically in the hope of gaining more insight into both universal and orthography-specific patterns of comprehension (Cho & Chen, 2005). What remains under-explored in this line of research is how the relationship between the two scripts contributes certain specific complexities during semantic processing.

As Chapter 2 showed, any given Sino-Korean morpheme has the potential to encode the semantics of more than one Hanja character, which means that each different meaning represents a competing member belonging to a possible semantic cohort of meanings, which is built across two different scripts. A model of the Korean mental lexicon should, thus, to some extent, be nuanced enough to capture these cross-script semantic cohorts. Very few models of the Korean mental lexicon exist, but in the following section, I offer an overview of the most recent and exhaustive account of the Korean mental lexicon.

3.6 Current models of the Korean mental lexicon

The current literature is limited in the way of proposed theoretical models of the Korean mental lexicon, with only one model that provides some representation of Hanja characters. Yi (2009) provides the following model, which (1) conceptualizes Hanja characters at the morphological level, and (2) positions this level of morphological representation above the level of lexical units.

Yi (2009) argues that a key feature of this model is a greater weighting of inhibitory connections between lexical-level representations which have the same syllable in the same position. This means that individual differences in lexical processing of different Sino-Korean words can be accounted for by these weighted connections between whole-word units and morpheme units. Increased exposure to Sino-Korean words, or Hanja characters, would not only create new links, but also reinforce existing links between Sino-Korean words and constituent morphemes (Yi, 2009).

Yi contends that models often take for granted a three-level structure which posits characters, morphemes and words on separate levels of representation. These layers are organized hierarchically with character representations at the lowest

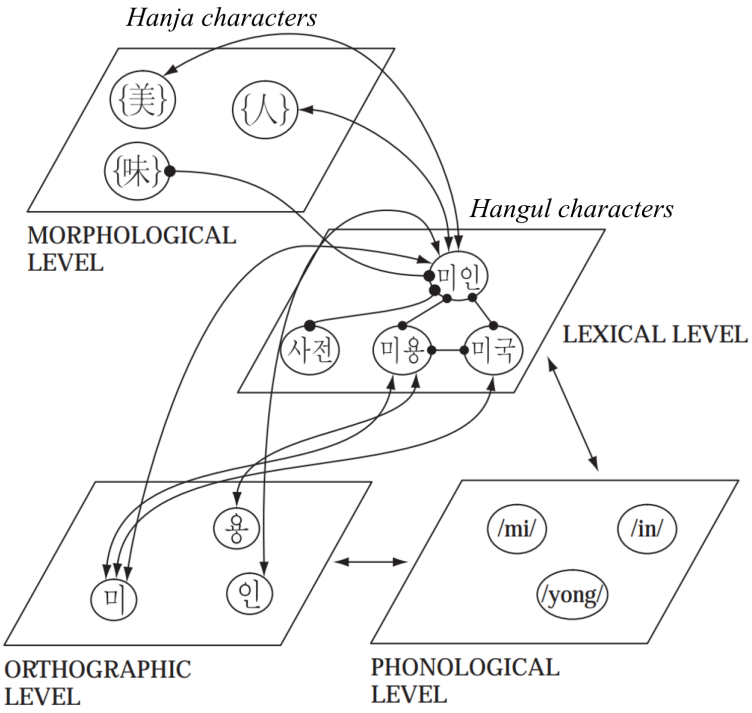


Figure 3.6: A model of the Korean mental lexicon for the representation and processing of Sino-Korean words, as proposed by Yi (2009: 407).

level and lexical representations at the highest level, with a layer for morphological representations in between. However, he proposes an alternative structure whereby the morphemic level of representation exists above the word level of representations. This is in part motivated by findings from his own studies on the morphological processing of Sino-Korean which show little evidence of morphological facilitation for Sino-Korean words that share a common morpheme (Jung & Yi, 2004; Yi, 1999; Yi & Yi, 1999). However, this in itself is problematic, since using prime-target targets that share a common morpheme would necessarily mean form overlap, obscuring proper interpretation of the results. Using 반항 [pan.haŋ], meaning 'resistance', as a target, Yi and Yi (1999) had two main experimental conditions, one in which the prime was morphologically related by sharing a constituent morpheme (**반**칙 [ban.ts^{h}ik], meaning 'violation', and one in which the prime was an orthographically related word sharing a syllable (e.g., 반장 [ban.tsaŋ], meaning 'captain'.

In addition, Yi and Yi (1999) also had a prime which was neither morphologically

or orthographically related, as a control. In both experimental conditions, there is clear overlap in form between prime and target, and while the morphologically related condition makes use of Hanja character (反, 'to turn') from which the meaning of 반 is derived, it neglects the fact the form 반 in 반장 [ban.tsaŋ], meaning 'captain', makes use of another meaning derived from a different Hanja character, namely 班, which means 'to turn'. He takes a supra-lexical perspective by reasoning that when primes and targets share a common morpheme, any observed morphological activation will facilitate processing of the target by way of excitatory connections between the relevant morphemic representations and the appropriate whole-word representation (Giraudo & Grainger, 2000). This seems to miss the point that for Sino-Korean, each morphemic lemma is underpinned by a rich network of semantic meanings, as contributed by different Hanja characters.

Despite such shortcomings, Yi's (2009) model represents an important step towards establishing an exhaustive account of the Korean mental lexicon, but takes for granted the semantic relationship that arises from the use of two different scripts, from the morphological to the lexical level. While the model focuses on the morpho-orthographic aspects governing the structure of the Sino-Korean lexicon, it fails to capture the various semantic networks that underpin processing. Yi's (2009) model can mostly be reduced to the mapping of different levels, with little description of the actual mechanics to explain how different levels interact.

For example, it is not very clear how exactly the contributions of Hanja, however defined (i.e., orthographic or morphological), play a role in processing. The model is useful insofar as it informs us of the level at which Hanja characters are represented, but in a similar tradition as the existing literature on the processing of Hanja characters (Cho & Chen, 2005; Cho & Chen, 1999; Kim, Taft, & Davis, 2004) makes few predictions for whether Hanja plays an active role in Sino-Korean processing, thereby taking for granted that the two scripts interact in some meaningful way without a proper understanding of how such interactions may be unfolding.

3.7 Main research objectives

The model Yi presents does not make room for the possibility of multiple semantic access, and more crucially, does not account for the multitude of meanings that are clearly available to each Sino-Korean morpheme, given the repeated use of homophones to represent more than one meaning. That is simply a fact of the Sino-Korean sub-lexicon that any model of the Korean mental lexicon must account for in their theoretical construct. There appears to be no such structure in Yi's model that would at least attempt to account for the semantic cohorts that exist for each Sino-Korean morpheme. Yi's model requires further refinement to capture the complex nature of the role of Hanja characters in cross-script semantic cohorts and Sino-Korean semantic processing. As part of developing a revised model of the Korean mental lexicon, I report on the findings of four behavioral studies investigating the precise status and representation of Hanja characters in the mental lexicon. In order to provide a finer-grained analysis of script dynamics between Hanja characters and Sino-Korean, I introduce the notion of cross-script semantic cohorts.

The complication that arises due to Hanja characters and their semantic contributions presents interesting challenges for models of the mental lexicon, particularly with regards how its structure handles semantic information that is tied to scriptal information. Building on previous works as described here, the following chapters detail a two-fold investigation of the status and the representation of Hanja characters in the mental lexicon, exploring (1) whether the semantic information in Hanja characters is represented in the mental lexicon, and (2) whether the relative number of encoded meanings, that is, the semantic cohort across scripts, modulates semantic access, given the potentially numerous semantic representations. Given the script-based properties of Korean, is it possible for activation to spread across scripts? Is such activation mediated by the scripts at play? Despite the high proportion of Sino-Korean words, the status of Hanja characters in the Korean mental lexicon, and more broadly semantic encoding in a bi-scriptal lexicon, has yet to be properly understood.

4

Status of Hanja in the mental lexicon

Contents

4.1 Introduction 76
4.2 Cross-script semantic representations 77
4.2.1 A cohort across scripts: Sino-Korean 82
4.2.2 Experimental paradigm and priming 87
4.3 Experiment 1: Visual Hanja fragment priming 92
4.3.1 Methods 92
4.3.2 Results 95
4.3.3 Discussion 95
4.3.4 Methodological considerations 96
4.4 Experiment 2: Visual Sino-Korean full-word priming . 98
4.4.1 The effects of varying cross-script semantic cohort size . 99
4.4.2 Methods 100
4.4.3 Results 103
4.4.4 Discussion 106

4.1 Introduction

The preceding two chapters provided an introduction to visual and auditory word recognition, with a special focus on semantic processing, highlighting the often overlooked role a language's writing system might have in such processing. Languages with a linear form of writing may present little challenge to the current state of the field, but those with structurally different scripts or those which do not use Roman characters, may offer new insight into how semantic meaning is encoded and accessed in the mental lexicon. What are the semantic networks that support

a bi-scriptal lexicon? We explore this question through the lens of Korean, which, as we saw in Chapter 3, encodes the semantic contributions of a non-native script (Hanja), using a native script (Hangul). This offers an interesting window into how semantic representations are encoded across multiple different scripts, representing an important departure from the existing literature on bi-scriptal processing, which very often centers on the study of bilinguals, who use the scripts of two different languages by virtue of speaking two different languages that use two different scripts (e.g., Japanese and English).

For Korean, however, such is not the case; the writing system inherently comprises two scripts, and further, we are not interested in bilinguals of Korean and Chinese, but rather monolingual speakers of Korean who are literate in the Korean writing system. The experiments here are dedicated to understanding such cross-script semantic processing, bringing to light new considerations for what it means to be a bi-scriptal lexicon. This chapter discusses the details of the first two experiments, which both employ a semantic priming lexical decision task. Both experiments use visual primes to investigate the status of visual Hanja; in the first experiment (Section 4.3), Hanja characters are presented directly, while in the second experiment (Section 4.4), Hanja characters are presented in the native script of Hangul. The chapter concludes with a brief discussion of the theoretical implications of the two experiments, and in doing so, lays the groundwork for the further investigation which is undertaken in Chapter 5.

4.2 Cross-script semantic representations

The representation of meaning is absolutely central to language, whether spoken or written. There are obvious and significant differences between the two domains of word processing that relate to the time course during which the sensory information becomes available. Subsequently, this has important implications for how semantic meaning is not only represented, but also accessed. Words presented auditorily can only be perceived to the listener in a strictly incremental way, as more information becomes available. In visual word recognition, on the other hand, the incoming

sensory information is presented holistically, which means that processing is not necessarily dictated by a sequential, left-to-right configuration.

Although the exact processes underpinning lexical access will vary between domains, the two must eventually converge on the same representation of information specific to that lexical entry. To what extent these representations take into account the contributions of script-based information is not clear. This is perhaps due a heavy focus in the literature on Indo-European languages, or rather languages using Roman-based scripts more generally, where the writing systems tend to be alphabetic, abiding by linear, sequential presentation of letters, and thus relatively more straightforward than those which have structurally different properties, both in terms of letter arrangment but also in how meaning is encoded.

In Korean, semantic information is encoded by a writing system which presents two structurally different scripts. As a result, numerous one-to-many mappings are generated across the two different scripts, creating not just a cohort of meanings, but a cohort of meanings which is in part derived from cross-script relations, a type of cohort that has yet to be defined. This research is an important step towards understanding semantic processing in a bi-scriptal lexicon, where the same semantic information might be represented in two different ways, using two different scripts. Studying languages that use a writing system consisting of structurally different scripts offers a useful lens through which to explore this domain of research. In pursuing these questions, we shed new light on how we might define a cohort, by exploring those which are based on cross-script relations. Could the structure of the writing system for a particular language help to draw meaningful links between otherwise non-related words?

This chapter aims to address the aformentioned question by examining Korean and its bi-scriptal writing system. The first part of our investigation into the status of Hanja characters in the representations of Sino-Korean words in the Korean mental lexicon is reported here. We begin by exploring the status of written Hanja in the mental lexicon. Investigating a bi-scriptal language, where a single language uses two structurally distinct writing systems, can offer new insights into lexical

access and the structure of the mental lexicon. Most importantly, in a language like Korean, where semantic information is encoded across two different scripts, one which is alphabetic and native to the language, and one which is logographic and non-native to the language, offers additional novel insights.

An account of the Korean mental lexicon that exhaustively models how Sino-Korean words are represented has yet to be proposed, and given the composition of the lexicon, it would seem remiss of any proposal to overlook proper treatment of Hanja characters. Prior research (Cho & Chen, 1999; 2005; Cho, McBride-Chang, & Park, 2008; Kang & Simpson, 2001; Kim & Davis, 2001; 2002; Kim, Taft, & Davis, 2004) on Korean and its scripts presupposes that Hanja characters are represented in the Korean mental lexicon. Our investigation is centered on empirically testing this as a means of solidifying an idea perhaps taken for granted, that Korean speakers encode such information.

If so, how is such information encoded? Korean casts a different light on how the mental lexicon may be organized. Understanding whether such information on Hanja characters is encoded will also have significant implications for the structure of the Korean mental lexicon, an account of which must reflect the divided composition of the lexicon between pure Korean words and Sino-Korean words. Various conceptualizations of the mental lexicon exist for other languages that share a similar history of Chinese influence, such as Japanese (Kess & Miyamoto, 1999; Bowers & Michita, 1998; Hino, Kusunose, Miyamura, & Lupker, 2017; Leong & Tamaoka, 1998; Verdonschot, 2011; Verdonschot, La Heij, & Schiller, 2010), and Yi's(2009) model of the Korean mental lexicon fails to consider the processing effects of Hanja.

The link between writing systems and language processing often goes unexplored, but such links have the potential to offer novel insights. Much of the literature about writing systems focuses on Western European languages, or languages with writing systems of Latin descent (Seymour, Aro, & Erskine, 2003 (exclusively Indo-European languages); Ziegler et al., 2010 (Finnish, Hungarian, Dutch, Portuguese, and French); Aro 2006; Asfaha, Kurvers, & Kroon, 2009; Georgiou, Parrilla, & Papadopoulos,

2008; Landerl, 2006; Ziegler & Goswami, 2005), ignoring the structurally diverse set of writing systems attested cross-linguistically. Languages like English, for example, use an alphabet which maps sounds to letters, while languages such as Chinese use a system of logograms, where each character (e.g. 下) represents a meaning ('down').

Korean draws on properties of both, in its simultaneous use of its own native sound-based writing system, Hangul, and non-native logographic system, Hanja, borrowed Chinese characters transcribed into Hangul according to Korean sound rules. Korean offers ideal testing grounds to investigate not only a bi-scriptal language, but a case where two structurally different writing systems co-exist within one language. Improving our understanding of how different languages map sound onto written form will offer critical insights into both language processing as well as literacy and learnability. A language like Korean introduces theoretical questions concerning how a bi-scriptal lexicon represents semantic information.

Even the structural idiosyncrasies of Hangul alone create interesting challenges; for example, unlike many languages with an alphabetic writing system, recall that Korean uses a non-linear arrangement of characters, creating block-like syllables. The featural design of individual Hangul characters heightens the significance of each compositional stroke, which represents important phonetic elements such as place or manner of articulation. Thus, Hangul can encode information at a number of different levels, spanning from the lexical, to the sub-lexical, right down to individual strokes. This information is intrinsically tied to Hangul and its relationship to Hanja, but is this relationship borne out in other ways during processing? We know that how words are linked is relevant to lexical access; we very often co-activate words because of some relationship they share.

What about in Korean? To what extent do such relationships, such as that of script, which for Korean appears to cut across all different levels of the lexicon, influence the structure of the mental lexicon? Marslen-Wilson (1987; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980) shows that in spoken word recognition, auditory lexical retrieval relies on incremental auditory input. Every possible word that is consistent with the first one or two speech segments, or phonemes, is activated,

creating a cohort of possible words. Although this notion of cohort is underlyingly driven by phonological motivations and its role in auditory processing, languages such as Korean, which offer an intricate system of cross-script links, offer new and robust grounds on which to extend the notion of cohort to non-phonological domains.

We predict, given the holistic composition of the Korean lexicon, more than half of which is Sino-Korean in origin, that (1) not only is Hanja represented in the Korean mental lexicon, but also that (2) the processing of Sino-Korean actively relies on such representations. A subsequent prediction is that the semantic processing of Sino-Korean necessarily involves probing the mental representations of the semantics of Hanja. This means that semantic access to Sino-Korean words engages cross-script representations; that is, the semantic concepts of individual Hanja characters, encoded by Hangul. These representations may be tied to the scripts themselves, i.e., the actual Hanja logograms and their corresponding representation in Hangul, or they may be an abstract representation, rooted in a more conceptual understanding of the semantics.

Either way, the semantic information as encoded in the Sino-Korean word is expressed and ultimately relayed across different scripts, from Hanja to Hangul. The potential for semantic information to be represented across two different scripts, one which is native to the language, and one which is not, is an understudied link between the systems we use to write, and the systems we use to process language. The seemingly parallel literature on the processing of cognates (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Gollan, Forster,& Frost, 1997; Grainger & Frenck-Mestre, 2008; Peeters, Dijkstra, & Grainger, 2013; Perea, Duñabeitia, & Carreiras, 2008) across different scripts is in many ways orthogonal to this discussion, as such work involves different scripts belonging to two different languages, and cognates occurring across those two distinct systems, but in the present research, we focus on semantic contributions, which, while they are expressed using two different scripts, are still confined to one language system – Korean.

We investigated how the semantics of Hanja may be represented in the mental lexicon, either as script-based form representations, or conceptual representations.

We conducted a series of four behavioral studies using a lexical decision task with semantic priming, two using visual primes to probe the possibility of form representations, and two using auditory primes to investigate the possibility of conceptual representations. We first report on our experiments using visual primes; our first experiment tests whether printed Hanja characters can activate Hangul representations. Based on the results of this first study, we conducted a second study wherein Hanja characters were fully lexicalized as Sino-Korean words, rendered in Hangul. The semantic contributions of the Hanja characters are directly encoded in the Sino-Korean word, which means that while the Hanja characters are no longer explicitly rendered, their meanings are still indirectly manifest through the Hangul representations. This is crucial, as it means that semantic access necessarily taps into the underlying representation of the Hanja characters.

4.2.1 A cohort across scripts: Sino-Korean

Semantic access to a particular cohort may be mediated by a phonological representation, but the composition of the cohort itself could be based on non-phonological features, such as semantic relatedness, or specific relations between scripts. To begin our investigation, consider the following figure:

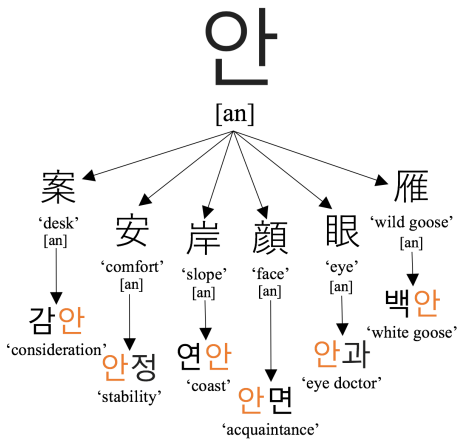


Figure 4.1: The meanings of the different Hanja characters, all of which approximate the same phonological sound (discounting the tonal value), represented as 안 in Hangul, which can partner with other syllables to create lexical items that relate semantically to one of the meanings encoded by the relevant Hanja logogram.

Figure 4.1 demonstrates the repeated use of the Sino-Korean syllable 안 [an] to express the various meanings of different Hanja characters. As described in Chapter 2, Chinese characters were borrowed into Korean in varied ways – for their meaning, their sound, or both – but crucially, these borrowings very often resulted in multiple meanings being encoded by the same syllabic morpheme, creating homographs, which by virtue of comprising the same Hangul characters, were also homophones.

We have observed that simultaneous activation of all possible meanings of a word is possible, regardless of whether the context biases selection for one over the others (Swinney, 1979; Onifer & Swinney, 1981). This suggests that we activate an initial semantic cohort of candidates which presently match the semantics of the available input. In Korean, we predict that each Sino-Korean morpheme activates a cohort of meanings, as contributed by the different Hanja characters which share the same Hangul representation. Thus, simultaneous activation of all possible meanings for individual fragments is also expected. A fragment in the context of visual word recognition in this thesis is hereafter defined as an individual morpheme. Given the sub-lexical one-to-many mapping that arises from the relationship between Hanja and Hangul, we are interested in whether multiple semantic access in fact does occur in Korean, whereby other available meanings are activated given the share representation in Hangul. How does semantic multiplicity that is the result of cross-script morphemic ambiguity affect lexical access and processing, and how is it resolved in Korean?

If we take as our starting assumption the idea that a semantic cohort of possible meanings is activated for each Sino-Korean morpheme, then we predict that all possible words containing that particular homograph would also be activated, which would still be consistent theoretically with Marslen-Wilson's Cohort Model (1987; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980). That is, from the initial cohort that is activated based on all meanings that match with the morpheme 안, an extended semantic cohort comprising all words containing the homograph is also activated. This is visualized in Figure 4.2.

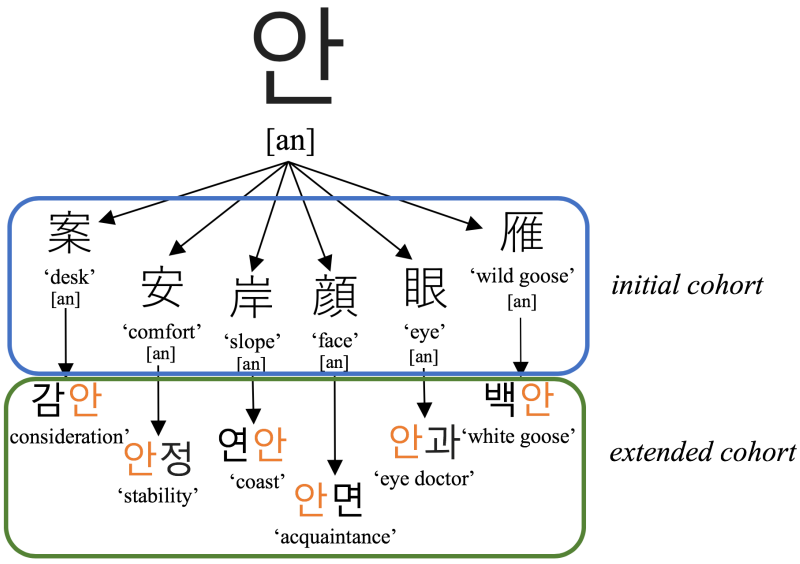


Figure 4.2: Semantic access across scripts, as mediated by the shared representation in Hangul, activates a starting cohort, which also activates an extended cohort containing all Sino-Korean words that contain that particular homograph, and thus the specific semantics of the appropriate Hanja character.

In order to facilitate such access, we posit a pathway that is mediated by the shared representation in Hangul. Consider the following two nodes in the mapping of 안, which are highlighted in black in Figure 4.3, while the remaining nodes are grayed out:

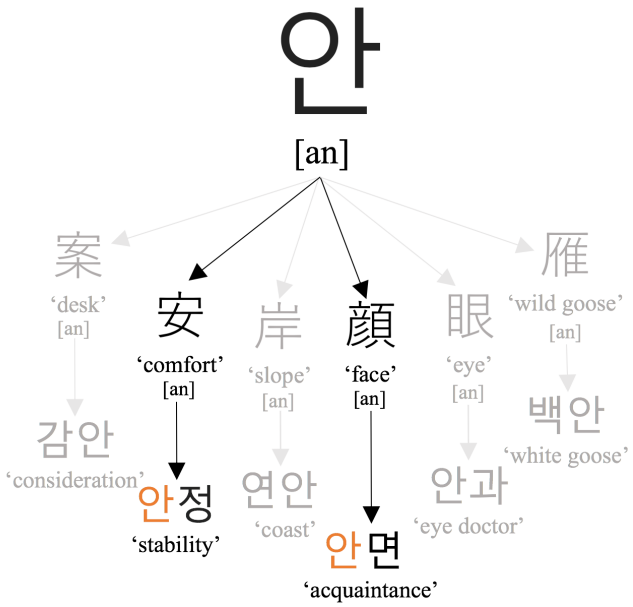


Figure 4.3: Two branches of the schema representing the many meanings of 안.

We assume that all meanings that are available for the fragment 안 are activated, based on which, we then predict that semantic access within the cohort is possible. That is, within the cohort of meanings available for 안, i.e., the various different Hanja characters that share 안 as their Hangul representation, it is possible to activate the semantics of one Hanja character from spreading activation to another.

Consider Figure 4.3, which highlights two specific nodes between which we would expect cross-activation (the selection of these two nodes is arbitrary, as we would expect activation from any given node); we predict, if all members of the cohort are activated, as per the Cohort Model (Marslen-Wilson & Welsh, 1978; Marlsen-Wilson, 1987; Marslen-Wilson & Tyler, 1980), that activation of the Hanja character 安 ('comfort') activates fellow cohort member 顔 ('face') and all other subsequent members in the extended cohort of words that contain this particular meaning of 안 ('face'), e.g., 안면 [an.mjʌn] ('acqaintance'). This is to say that retrieval of the different Hanja characters is mediated by the shared orthographic representation in Hangul; the idea is that, in processing the meaning of any given Hanja character, speakers necessarily activate the corresponding representation in Korean, or Hangul, which in turn enables access to all other members of the cohort, i.e., other Hanja characters that also map onto the same Hangul representation. This would be evidence for not only cross-script semantic access, but also semantic representations that actively rely on tapping into information that is encoded across two different scripts.

Given the relationship between Hanja characters and their corresponding Hangul form, we are interested in the following structure, where a cohort of meanings for a fragment can be activated, and more interestingly in Sino-Korean, across two different scripts:

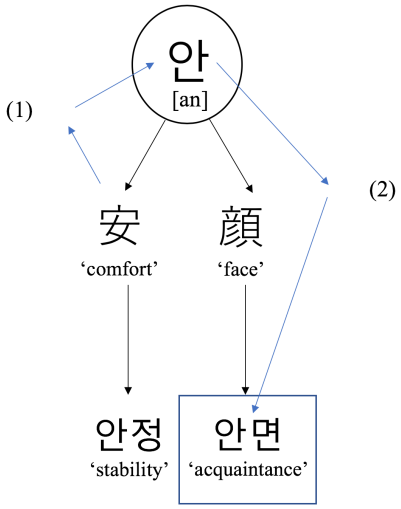


Figure 4.4: Potential semantic access across scripts, as mediated by the shared representation in Hangul.

Figure 4.4 illustrates that visual presentation of the Hanja character 安 (‘comfort’) will activate the Hangul representation 안, and in doing so, also activate the Hanja character 顔 (‘face’), which shares the same Hangul representation (안). Additionally, in reading the printed Hanja character, it may also be the case that the phonological representation [an] is also activated, which may in turn activate a cohort of Hanja characters that share the same phonological representation. Whether the intermediary representation mediating semantic access is the Hangul representation 안 or the phonological representation [an] may depend on individual differences, but the role of Hangul, whether in form or in sound, is crucial in this type of processing.

Under this view, in processing Sino-Korean, at each level of representation, a cohort of candidates is activated, which in turn activates another set of candidates that are linked via phonological, semantic or script-based relations. This notion of cohort is not new, especially within the context of spoken word recognition, but its use in this particular domain of script processing, however, is. As Figure 4.4 shows, the Hanja character 安 (‘comfort’) activates both the phonological representation [an] as well as the orthographic representation <안>, which activates all other Hanja characters that also share the same orthographic and/or phonological representation. In doing so, at this point, we also activate all of the meanings of these Hanja characters (recall the nodes in gray in Figure 4.3). Also at this point,

we activate all words that match the phonological sequence [an] and thus contain 안 in their orthographic representation.[1] Thus, in addition to the semantic cohort that is activated for the fragment, a cohort of semantically related words, i.e. full lexicalizations of the Hanja character in Hangul, is also activated.

This research represents a new direction in Korean psycholinguistics, which means little in the way of existing literature. As Chapter 3 outlined, much of the present work has focused on the effects of Hanja on Korean processing centers principally on the domain of literacy, the degree of exposure to Hanja characters being the main independent variable (Tavassoli & Han, 2001; Yoon & Lim, 2011; Cho & Chen, 1999; 2005; Cho, McBride-Chang, & Park, 2008; Kang & Simpson, 2001; Kim & Davis, 2001; 2002; Kim, Taft, & Davis, 2004). While these studies offer important insights into how levels of proficiency in Hanja interact with ability to comprehend Korean in a mixed script context, the overarching question of how exactly the two scripts are represented in the mental lexicon remains unanswered.

Such an understanding is of use in its own right as it draws attention to the role of writing systems in the mental lexicon, an under-researched area, but also contextualizes more broadly any observed interactions between the two scripts in the area of literacy and reading comprehension. That is, while more recent exposure to a script may enhance its role in processing, more generally, the fact that script-specific information is encoded in the first place offers an important precedent for any future analysis. Thus, using psycholinguistic methods and measures, we investigate (1) the idea of cohorts for smaller lexical units, and (2) cohorts drawing on information encoded bi-scriptally.

4.2.2 Experimental paradigm and priming

A number of different experimental tasks are employed to investigate psycholinguistic processing, the two most common of which are *lexical decision* and *naming*. A lexical decision task involves participants deciding whether a string of letters forms a real

[1] At this point in time, we are referring exclusively to the Sino-Korean lexicon, which means that the pure Korean, i.e., not Sino-Korean, homograph of 안 is not activated. The intersection of the Sino-Korean and pure Korean sub-lexicons and the cohort of morphemes that are shared across the two will be discussed in Chapter 6.

word or a non-word. That is, participants press one button (YES) in response to the word NURSE, and another button (NO) in response to the non-word MURSE. Here, the experimenter measures the reaction times (and error rates) of the participants. An important consideration in lexical decision tasks is the problem of the *speed–accuracy trade-off*, where it is the case that the faster the participants' response, the more errors they are likely to make (Pachella, 1974). Asking participants to be accurate yields increased accuracy but slower reaction times, while encouraging them to respond as quickly as possible results in faster reaction times at the cost of making more mistakes (Pachella, 1974; cf. Wagenmakers, Van der Mass, & Grasman, 2007).

Priming is particularly important in word recognition, and involves presenting material before the word to which participants are asked to respond. The general idea of priming is that if two things are similar to one another and involved in the same level of processing, they will either assist with or interfere with each other. For example, the word *doctor* may be easier to recognize if the word *nurse* has been presented before, since the two are related in meaning. In contrast, people tend to respond more slowly when a target word is preceded by an unrelated word like *hula-hoop*; this has been shown by the works of Meyer & Schvanevedlt (1971; 1976), and Moss et al. (1995). This effect is known as semantic priming (Collins & Loftus, 1975; Posner & Snyder, 1975; Meyer & Schvaneveldt, 1971; 1976; Moss, Ostrin, Tyler & Marslen-Wilson, 1995; Kuperberg, Lakshmanan, Greve, & West, 2008; Rissman, Eliassen, & Blumstein, 2003; Wagner, Desmond, Demb, Globver, & Gabrieli, 1997; Balota & Spieler, 1999).

Semantic priming results from the spreading of activation in the lexicon. Since *doctor* and *nurse* share common attributes, activating one of the concepts necessarily leads to activation in the set of properties comprising the meaning of the other concept (Anderson, 1983; Collins & Loftus, 1975; Neely, 1977; Posner & Snyder, 1975). Thus, if the participant is presented with *doctor*, *medicine*, *can heal*, *stethoscope* are also activated. Then, when the participant is presented with *nurse*, the pre-activated concepts support the lexical decision the participant has to make. There is a shorter period of time required for the network to activate the parts of the semantic network

that represent the concept presented in the target, i.e. *nurse* (Meyer & Schvaneveldt, 1971; Posner & Snyder, 1975; Meyer, Schvaneveldt & Ruddy, 1974; Anderson, 1983; Collins & Loftus, 1975; Neely, 1977; Hutchinson, 2003).

Mediated priming is another type of priming relevant to our study. Spreading activation is believed to lessen beyond one or more links in the network, and studies using mediated priming offer evidence for such findings. To explain, mediated priming involves word pairs like *lion* and *stripes*, and the idea is that *lion* draws a connection to *stripes* through the mediating word *tiger*. Thus, presentation of the word *lion* causes activation to spread to *tiger*, which in turn should activate *stripes* (Anderson, 1983; Collins & Loftus, 1975). Following this pathway of spreading activation from *lion* to *tiger* and finally to *stripes*, the word *lion* should prime the response to the target word *stripes*.

Indeed, activation spreading beyond the directly related concepts (like *lion* and *tiger*) is supported by findings that show a small priming effect for *stripes* primed by *lion* (Balota & Lorch, 1986; McNamara & Altarriba, 1998). However, *lion* primes *stripes* much less than it primes a concept that is directly related like *tiger*, which is expected, since with increased distance, the spread of activation diminishes (de Groot, 1983; Balota & Lorch, 1986). This is because semantic network theory, as discussed in Chapter 3, explains that there is a limited amount of total activation that can be spread over the network; otherwise, activation would spread beyond immediately connected nodes, spreading throughout the whole network. Thus, nodes that are directly connected to the prime are strongly activated, while those which are less directly connected are less strongly activated.

The special focus on the role of scripts and orthographic representations warrants inclusion of form-based priming as yet another form of priming. Here the idea is that it is reasonable to predict that a word like *contract* should prime *contact*, due to similarity in visual form. This overlap means that they are orthographically related, resulting in a phenomenon known as orthographic, or form-based priming. However, form-based priming is difficult to show (Humphreys, Besner, & Quinlan, 1988; Forster & Veres, 1998), as orthographically related words also tend to be related in

other ways, such as phonologically. That is, those which share the same orthographic onsets, e.g. the initial string of C-O-N in both *contract* and *contact*, also share the same phonological onset of [kɒn], making it difficult to discern whether any observed priming can be attributed to the orthographic or phonological overlap. Equally, words sharing orthographic shape are also likely to share morphological structure, thereby making them morphologically related in addition to orthographically; consider *darken-dark*, for example. With regards to orthographic cohorts, form-related primes have been shown to inhibit activation, slowing down recognition of the target (Colombo, 1986). This has been attributed to the possibility that words visually similar to one another compete for selection during processing.

In summary, priming enables us to probe the relationship between two words, by way of observing the type of effect the prime has on subsequent processing. The reaction times from the lexical decision task represent an index of how accessible the word's entry is in the lexicon. Manipulating the relation between the prime and the target can help to further inform our understanding of visual word recognition. In our particular case, seeing the type of effect the Hanja prime has on the accessibility of the Hangul target via the time it takes to respond, can offer important insight into how the scripts interact and are organized in the mental lexicon. If reaction times are slower compared to control conditions using an unrelated prime, this means that more time that is needed to activate the relevant properties in the semantic network, suggesting inhibition. If reaction times are faster, this suggests less time required to activate the necessary properties, and thus an effect that facilitates target activation.

The amount of time that is needed to make a lexical decision can be contingent upon a number of different factors. Nonwords that are similar to real words, for example, often yield slower response times compared to non-words which are significantly different (Vitevitch & Luce, 1998; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Coltheart, Davelaar, Jonasson, & Besner, 1977; Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015). Likewise, real words that are higher in frequency usually elicit faster reaction times than those which are lower in frequency (Morton,

1969; Forster, 1989; Forster & Bednall, 1976; Taft & Forster, 1975; Duchek & Neely, 1989; Ziegler, Jacobs, & Klüppel, 2001).

These response times and the various factors that affect speed of activation provide important insight into how the mental lexicon might be structured. Applying these same methods, we use a lexical decision task with semantic priming to explore the status of Hanja characters in the mental lexicon. The experiments presented here are all methodologically identical, with primes varying along two dimensions – modality (audio versus visual) and lexical size (full word versus fragment). If primes are visually presented, they also vary in terms of script (Hanja versus Hangul). This enables us to probe the precise nature of how exactly the semantic contributions of Hanja are represented, testing whether observed effects are stable across input modalities.

With all such methods in mind, the paradigm of use for all studies presented in this thesis will be one of mediated semantic priming. In our first study, we present a Hanja character as the prime prior to the Hangul character, which is the target to which a response (via a lexical decision) has to be made. By mediated, we mean that the activation of a word form is mediated by that of another word form which is in some way semantically related (contributed by the scripts relations of the Korean writing system). If the priming results in quicker processing, this is called **facilitation**, but if priming yields slower processing, this is referred to as **inhibition**. In this study, we are interested in seeing how the scripts affect one another in processing, and whether priming of the Hanja will facilitate or inhibit activation of the Hangul target.

We begin our investigation by looking into whether cross-script activation, from Hanja to Hangul, is viable during semantic access. We test whether visually presented Hanja characters alone can activate Sino-Korean words **(Experiment 1)**, as well as whether Sino-Korean lexicalizations of Hanja characters can provide semantic access to words semantically related to other Hanja characters in the same cohort **(Experiment 2)**.

4.3 Experiment 1: Visual Hanja fragment priming

In this first study[2], we investigated how semantic information is encoded across two structurally-different scripts, by exploring how the various semantic contributions of Hanja characters are represented. We hypothesize that the shared representation in Hangul mediates semantic access, enabling multiple activation. The idea is that the written form of the Hanja character activates the Hangul representation, activating all other Hanja characters which are represented using the same Hangul form. We posit that the various semantics of the different Hanja characters can be accessed via the shared representation in Hangul. Given the focus on script-based representations and the question of to what extent script plays an active role in processing, we presented both primes and targets visually in an intra-modal lexical decision task, looking to see how visual presentation of Hanja affects activation of Sino-Korean.

4.3.1 Methods

4.3.1.1 Materials and design

For each target in this experiment, there are two prime conditions. Exemplar stimuli are presented in Figure 4.5.

1. *Related prime.* These primes, which are Hanja characters, are not directly related to the meaning that is presented in the target, but they do share the same representation in Hangul with another Hanja character that *is* directly related in meaning to the target. The idea is that the relationship between the prime and the target is mediated by an 'intermediate' step, which is the representation in Hangul that is shared between the two Hanja characters.

2. *Unrelated prime.* These primes are not semantically related to the meaning presented in the target nor is there any intermediate step to facilitate any relatedness.

[2]All studies presented in this thesis received Research Ethics Approval from the Central University Research Ethics Committee (CUREC) at the University of Oxford, Ref No: R52481/RE001.

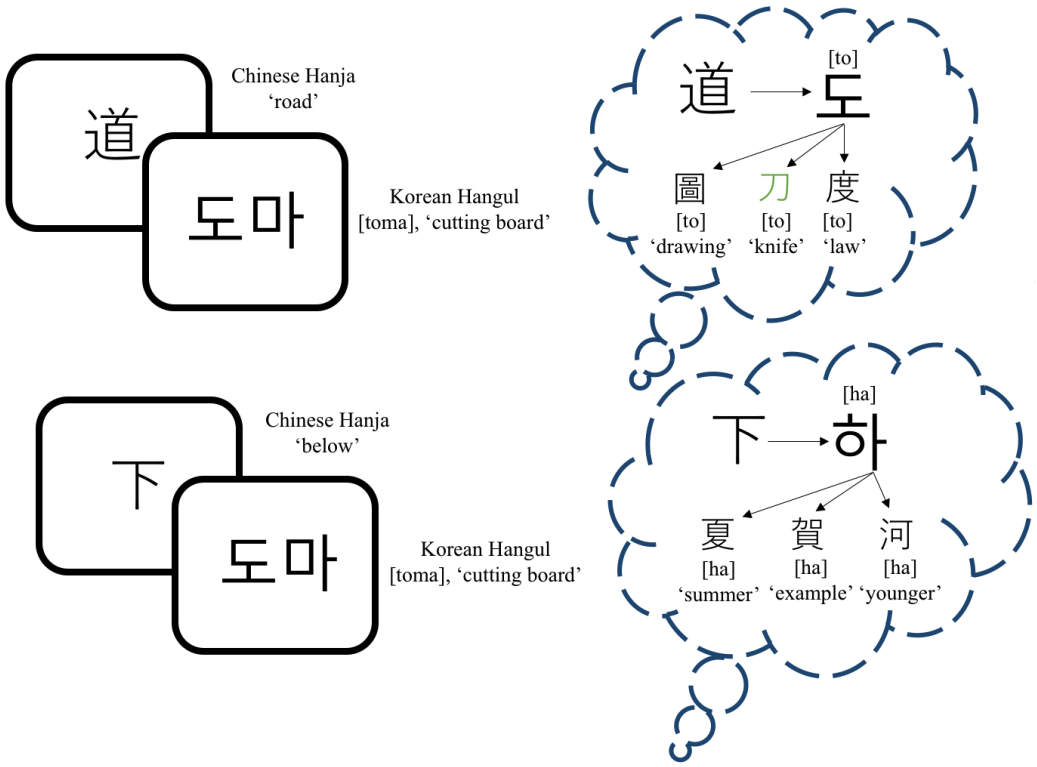


Figure 4.5: Overview of test stimuli.

We also devised a set of non-word stimuli, which obviously bear no relation in meaning to the primes preceding them. The total number of targets was 48, 24 of which were non-words to provide a balanced design in the number of word and non-word stimuli. Of the other 24 real word targets, 12 were preceded by semantically related primes, and the other 12 were preceded by unrelated primes. Two experimental lists were created in a Latin Square design, in which each target and either its related or unrelated prime occurred only once.

4.3.1.2 Participants

44 Korean native speakers participated in this first study. The participants were students or post-doctoral researchers at the University of Oxford who were from Korea and had received some (i.e., primary and secondary) education, if not all (including university), in Korea. These selection criteria ensured that they had a basic understanding of Hanja characters provided through the school curriculum. All

participants had corrected-to-normal vision and no hearing (or other) impairments. They all spoke English as a second language (L2), and were currently living in an environment (Oxford) where English is the predominant spoken language. The participants gave their informed consent and were compensated appropriately for their participation.

4.3.1.3 Procedure

The study was conducted in the *Language and Brain Laboratory* at the University of Oxford. Participants were tested individually, or in pairs. Upon arriving, participants were provided with a sheet of information explaining the study in Korean and asked to sign a consent form. The participants were first asked to complete an information-gap activity, either with the experimenter or another participant if they came in pairs. Due to the participants being in a predominantly English-speaking environment, the information-gap activity attempts to put participants into the appropriate language mode (i.e., 'Korean mode') before completing the study, as increasing evidence corroborates the importance of language mode (Grosjean, 1985; 1997; Bialystok, 2015; Green, 2011; Prior & Gollan, 2011; Luk & Bialystok, 2013). For the activity, participants were given a picture (see Appendix A) to describe in Korean so that the experimenter could draw and replicate the picture as closely as possible. If the participants arrived in pairs, they were each given a picture, the two pictures being nearly identical, and asked to find the differences between their respective pictures, using solely verbal communication. The duration of the information-gap activity was capped at ten minutes.

Afterwards, the participants were seated behind an allocated computer and provided with written instructions, which were then summarized by the experimenter, at which point participants were also invited to ask any remaining questions. They were told to respond only after seeing Hangul, and further instructed that if the Hangul formed a real word, to press YES, and if it did not form a real word, to press NO, using the button box, as quickly, but also as accurately as possible. The button boxes were placed in the participants' hands such that they were able to

respond YES with their dominant thumb. All subjects participated in a ten-item practice session to ensure that they had understood the task.

A Latin Square design was used, ensuring that every participant only saw each prime and target once, while allowing for maximal use of all prime-target combinations. Forty-eight stimuli were presented to each participant. Each Hanja prime was presented for 500 ms, followed by a blank screen for 600 ms. Each Hangul target is then presented for 650 ms, and the participants are given 1850 ms to complete the lexical decision task, with an inter-trial interval of 700 ms. The stimuli were presented with experimental software designed by Reetz and Kleinmann (2008), and all trials were pseudo-randomized.

4.3.2 Results

Participants with an error rate higher than 25% and items that elicited high error rates across all participants were excluded. Trials for which reaction times exceeded the cut-off point of ±2 standard deviations (SD) from the overall mean reaction time were also excluded from the analysis. Only correct responses were submitted for further data analysis. The total loss made up 20.4% of data.

Based on these data, a linear mixed effects analysis was performed for the reaction times, with the fixed factor *Prime Type* (Related versus Unrelated) and the random factors *Participant* and *Target*, using the statistical analysis package JMP (SAS). The main effect of *Prime Type* did not reach significance ($p = .359$)], meaning that a priming effect could not be established for real word targets.

4.3.3 Discussion

In this first study, no priming effects were observed for the Related condition compared to the Unrelated condition, which may be attributable to a number of factors, but two main causes are of interest. The participants may have highly variable levels of proficiency in Hanja characters, further exacerbated by limited exposure to Hanja characters day-to-day in printed form, living outside of Korea. Based on these findings, because of the high variability in speaker ability, it is difficult

to gauge whether the Hanja characters are having any effect on the processing of Korean, let alone argue for a cross-script interaction.

Given that the script alone cannot elicit any priming, this suggests that the relationship is not purely mediated by the visual script. Insofar as the Hanja characters are concerned, perhaps their mental representations are phonological, from which the semantic information can be derived. By positing a relationship that is not completely based on the visual script, we can also account for speakers who may have experienced language attrition and thus have diminishing use of Hanja characters, especially in terms of how they are written, but may still retain knowledge of what the characters can mean. Thus, is seems likely that rather than visual representation of Hanja in the mental lexicon, such representations may be more conceptually-based (or phonologically-based).

4.3.4 Methodological considerations

The reasons for the observed results in this first study could be due to a number of factors, but most pressing may be those which concern our participant demographic. While they were all native speakers of Korean (and Korean by heritage), they varied significantly not only in their day-to-day exposure to Hanja characters in print, but also in their general proficiency in Hanja. Participants were were required to have completed all of their primary and secondary education in Korea, to minimally ensure that they had received some degree of formalized education in Hanja. Our justification was that speakers of Korean educated in the UK for their primary and secondary school years were unlikely to have received such education, unless under special circumstances (i.e., independently taught).

Despite taking into consideration such factors, the issue is likely to remain that regardless of whether one has working proficiency in Hanja, having learned it in school in Korea, living in the UK, surrounded by English as the ambient language, will not necessarily provide the daily exposure needed to actively access the Sino-Korean representations in the mental lexicon. While Hanja is doubly unlikely to be seen living in a non-Korean dominant environment, such as Oxford,

even the exposure to Hangul will be generally limited. In Korea, while Hangul is the primary script in use, Hanja characters still appear with relatively regular frequency, whether it be on public signs (Figure 4.6), or in other forms of printed media (Figures 4.7 & 4.8).



Figure 4.6: A road sign in Seoul, which displays the street name (*Itaewon*) three different ways – in Romanized Korean (*Itaewon-ro 19-gil*), Hangul (*이태원로 19 길*, offset in yellow), and Hanja (underlined in red).

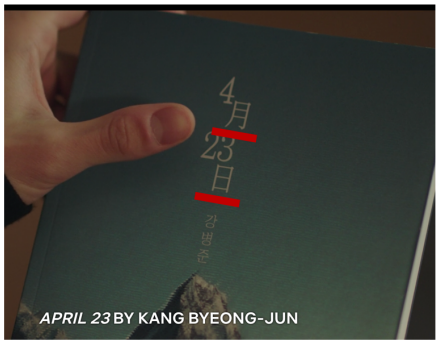


Figure 4.7: A cover of a book titled *April 23rd*, with the Hanja characters used to express *month* and *day* (underlined in red).



Figure 4.8: Much of Korean television is 'annotated' with additional commentary, presented in Hangul (offset in yellow), and very often featuring Hanja characters (underlined in red; n.b., this text is not to be mistaken for closed captioning. Unlike closed captioning, in which speech is simultaneously presented onscreen in written form, this particular form of captioning, which has been described as *graphic overtitles* (Park, 2009), does not reiterate what is being said onscreen. Rather, such captioning tends to express unspoken voices, such as the internal thoughts of a character, comments by the producer, or a predicted response from the audience. Importantly, these unspoken voices tend to contrast with the spoken language and/or imagery that is being heard or observed onscreen).

As observed in this still, Hangul and Hanja can be found printed not only in text like books and newspapers, but also in more image-based forms of media, such as television, underscoring the pervasive nature of Hanja characters in day-to-day media content.

Given the limitations due to individual differences in Hanja proficiency, for the following studies, we changed all our prime stimuli to Korean, so that both primes and targets could be presented in Hangul, in order to increase accessibility. This meant that rather than presenting the Hanja character directly, there were presented indirectly via Sino-Korean words, which contained a Sino-Korean morpheme which encoded the meanings of more than one Hanja character. Furthermore, to ensure greater uniformity in our participants in their day-to-day exposure of printed Hangul and spoken Korean, we ran all remaining experiments in Seoul, South Korea.

4.4 Experiment 2: Visual Sino-Korean full-word priming

Limited daily exposure to and general lack of proficiency in Hanja characters may make semantic access of Hanja characters doubly difficult for speakers of Korean who are not based in Korea. It is important to bear in mind that we are not interested in how the level of proficiency and literacy in Hanja characters affects Sino-Korean processing; rather, we are interested in the possibility of speakers tapping into a more abstract representation of the semantics of Hanja, particularly given the varied familiarity with Hanja characters. That is, in order to mitigate against the implications of a theory that posits representations of Hanja predicated on literacy, we shift to a more conceptual-based undestanding of the semantics, although still rooted in script (Hangul) nonetheless.

In presenting the Hanja characters as full lexicalizations of Sino-Korean words, Hanja characters are no longer explicitly present, but rather indirectly, as manifest in the Hangul representation of the Sino-Korean words, which bear the original semantic contributions of the Hanja characters. In continuing with our investigation of Hanja as a script-based representation, primes and targets continue to be presented

visually in an intra-modal lexical decision task. The main difference between the current study and Experiment 1 is that primes are no longer Hanja fragments, but rather full words in Hangul, and in making this change, both primes and targets are presented in Hangul.

4.4.1 The effects of varying cross-script semantic cohort size

Another important aspect that Experiment 2 takes into account is the relative size of the semantic cohort. In Experiment 1, we were primarily interested in whether Hanja characters alone could activate Sino-Korean words. A cohort cannot be formed on the basis of a single Hanja character, but a semantic cohort can be built around a Sino-Korean fragment, as the Hangul representation of the Sino-Korean very often encodes the meanings of multiple Hanja characters. As we saw earlier in Figures 4.2 and 4.3, a single Sino-Korean morpheme can activate a cohort of possible meanings, which can also subsequently activate an extended cohort of all Sino-Korean words containing that particular morphemic syllable.

Crucially, depending on the Sino-Korean morpheme, the relative number of available meanings can vary significantly. If we take the morphemes 미 [mi] and 학 [hak] as examples, we can observe the relative difference in the number of available meanings between them (Figure 4.9). The question then emerges whether having fewer or more mappings bolsters semantic access: Does a larger semantic network surrounding a given fragment speed up or slow down lexical processing?

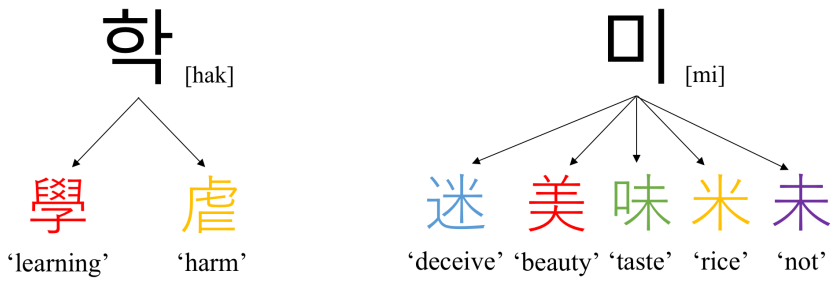


Figure 4.9: The representations of [mi] and [hak] in Hangul can map onto a multitude of different Hanja characters, thereby encoding a number of different meanings, the relative number varying between [hak], which has fewer, and [mi], which has more.

It is important to note the emphasis here is less on the precise number, and more on the fact that multiple meanings may be available in any given case. The exact number of meanings any one Sino-Korean morpheme carries is difficult to define, as Hanja characters are borrowed into Korean.

The precise size of a semantic cohort for any one Sino-Korean morpheme is difficult to define, as boundaries are largely artificial due to borrowings being in flux. It is hard to pinpoint the exact moment at which a borrowing has fully entered a language; thus rather than treating cohort size as discrete or categorical, we will conceptualize semantic cohort size as a continuous variable, quantifying each additional mapping as a new member of the cohort. As primes will now be presented in Hangul (due to the aforementioned limitations of Hanja accessibility), we will be able to take into account relative differences in size of the semantic cohorts, enabling us to explore in greater detail how the composition of a cohort affects processing.

4.4.2 Methods

4.4.2.1 Materials and design

All prime and target stimuli were disyllabic Korean words, and all primes were Sino-Korean. We created 36 prime-target pairs based on a set list of Hanja characters taught in Korean secondary schools to maximize the likelihood of participants knowing the Sino-Korean homographs. Each prime contained a Sino-Korean morpheme, the homographs of which mapped onto two Hanja characters, and thus two different meanings.

These Hanja characters were among those identified by the Korean Ministry of Education as Chinese characters to be learned for educational purposes (Grant, 1982; Cho, Sohn, &Yang, 2002; Korean Association for the Promotion of Hanja Education (KAPHE) database), and therefore were considered to be recognizable for the average native Korean speaker. To note, no frequency counts are available for individual Sino-Korean morphemes, nor is there any frequency measure for actual Hanja characters in text; thus, it was not possible to take into account potential effects for frequency due to such limitations.

It was important that primes and targets did not share any orthographic or phonological overlap. Since we were interested in the possibility of multiple semantic activation as mediated by the shared Sino-Korean morpheme, any form overlap between the prime and target would obscure whether observed priming effects are due to speakers tapping into multiple meanings, or because of the repeated morpheme. As a result, we used mediated priming in our paradigm to tap into the relationship between meanings sharing form representation.

This method has not commonly been used before and thus represents a novel approach to understanding processing in a bi-scriptal system whereby semantic representations exist across different scripts. Each target item was preceded by one of three possible primes that did not share its visual or phonological properties: (1) Directly Related to the target: that is, the meaning of the homograph as instantiated in the prime is also related to the meaning presented in the target; (2) Indirectly Related to the target: that is, the target is related to the other Hanja mapping that is not directly lexicalized in the prime itself, thus semantic access is mediated by the shared Hangul representation; or (3) completely Unrelated to the target (Figure 4.10).

Note that in Experiment 1, we only had two experimental conditions, wherein the equivalent Related condition involved a semantically-mediated relationship via the shared Hangul representation. So, in fact, the prime was indirectly related to the target, as the meaning presented in the target was semantically related to the meaning of another Hanja character, but one which shared the same Hangul representation as the Hanja character presented in the prime. It would have been useful to have a condition where the Hanja meaning presented in the prime was directly manifest in the meaning presented in the target, so we include this as a third condition in Experiment 2, in order to establish a full comparison.

As our design depended on the specific Hangul-to-Hanja mappings and thus each stimulus necessarily had to contain a specific Sino-Korean morpheme, the word-frequencies of the resulting Korean words were controlled to the greatest extent

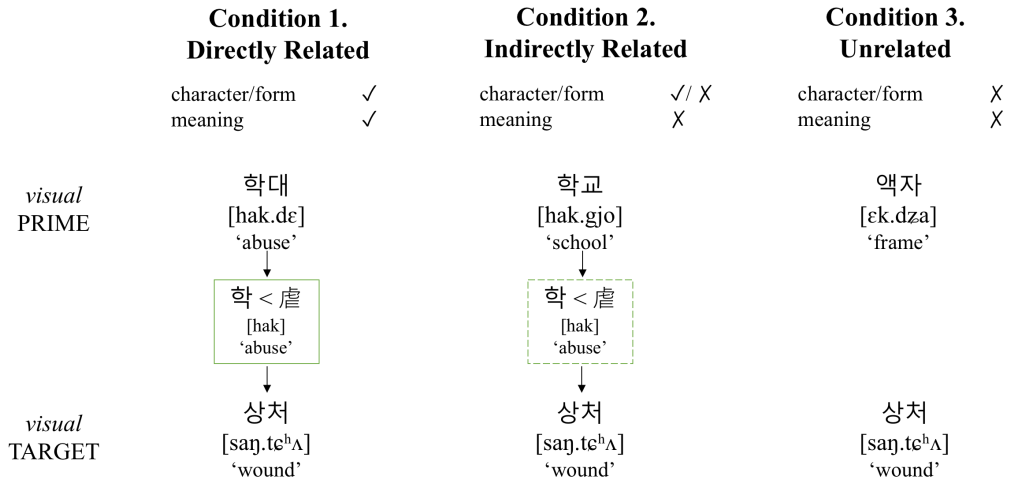


Figure 4.10: Experimental conditions. Depending on the condition, the prime-target pair involved a direct or indirect semantic relationship that required mediated semantic access via shared homograph (e.g., *hak*).

possible given these constraints.[3] Pairwise t-tests across the three conditions showed no difference in word frequency, and an additional one-way ANOVA that included our three main experimental conditions also confirmed no significant difference in frequency ($F(2, 107) = 0.468$, $p = 0.628$). Judgements on the semantic relatedness between the primes and targets in the Directly and Indirectly Related conditions were checked against independent ratings on a 5-point scale[4] by 30 native Korean speakers who did not participate in the main experiment. In order to provide a balanced design in the number of word and non-word stimuli, 36 additional prime-target pairs with non-word targets were included, resulting in a total of 72 items divided into 12 blocks of six trials.

4.4.2.2 Participants

A total of 197 Korean undergraduate students from three different universities in Seoul, Korea (Hansung University, Kyung Hee University and Sookmyung Women's University) participated in the experiment. All participants were native speakers

[3]Once primes and targets had been selected (based on the relevant Sino-Korean morpheme), we used a frequency database collated by the National Institute for Korean Language to control for frequency as best as possible (National Institute of Korean Language, 2008).

[4]To confirm the degree of semantic relatedness between stimuli, primes and targets from both the Directly and Indirectly Related conditions were given to subjects who were asked to rate the items on a scale from 1 (completely unrelated) to 5 (highly related).

of Korean, had normal or corrected-to-normal vision, and had no language or reading impairments. They provided informed consent and were compensated for their participation.

4.4.2.3 Procedure

The experiment took place in quiet rooms located at the different universities. The experiment was controlled using *Presentation* software (Neurobehavioral Systems) and displayed on a 14-inch Dell Latitude 7480 laptop. A Latin Square design was used to ensure that every participant only saw each prime and target once while allowing for maximal use of all three possible prime–target combinations. On each trial, participants were first presented with the visual stimulus for a duration of 300 ms. The targets were presented for a duration of 300 ms, after an interstimulus interval (ISI) of 250 ms. Participants were given up to 1500 ms to make a lexical decision.

Capitalization of letters does not exist in printed Hangul, in the same way that we observe in English, but is a convention that is often used in psycholinguistic studies to distinguish between different stimuli. Thus, to differentiate between the prime and target stimuli for this experiment, targets were presented in font-size 85pt, while primes were presented in 65pt. Participants were then instructed to respond only to the Hangul string presented in the larger font. Before the experiment, participants completed a practice session comprising ten practice items. Participants were instructed to decide whether the string presented on the screen was a word in Korean, using the button box provided. YES responses were made with the participant's dominant thumb.

4.4.3 Results

Only correct responses were submitted to further analysis, and all responses falling outside 2 SDs of the group mean reaction time to targets were excluded. Twelve participants had to be excluded due to high error rates and further analysis of frequency measures showed two items to be outliers and were thus also excluded.[5]

[5]Post-hoc analysis revealed that frequencies for two of the targets lay outside our cut-off for outlier rejection (average + 2*SD).

Overall, 16.42% of the data was excluded from the analysis. The remaining reaction times were submitted to a linear mixed model analysis (Baayen, 2008). *Participant* and *Target* were included as random effects using the statistical package *R* (R Core Team, 2018) and the downloadable package *lme4* (Bates, Maechler, Bolker, & Walker, 2015) for mixed effects modeling. Treatment coding with the reference level Unrelated was applied to our fixed effect, which was defined as *Prime Type* (Directly Related, Indirectly Related, Unrelated).

Through model comparision, it was established that a model containing random intercepts for *Participant* and *Target* provided the best fit for our data, while inclusion of *Cohort* as a nested factor within *Prime Type* did not improve model fit. Thus, our maximal model[6] did not include *Cohort* as a continuous variable. Our results are summarized in Table 4.1 and visualized in Figure 4.11:

Table 4.1: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	483.460	7.337	123	65.891	$< .001$*
Directly related	-7.337	2.401	5687	-3.056	.002*
Indirectly related	-3.947	2.414	5686	-1.635	.102

[6] lmer(RT ~ PRIME TYPE + (1 | PARTICIPANT) + (1 | TARGET)).

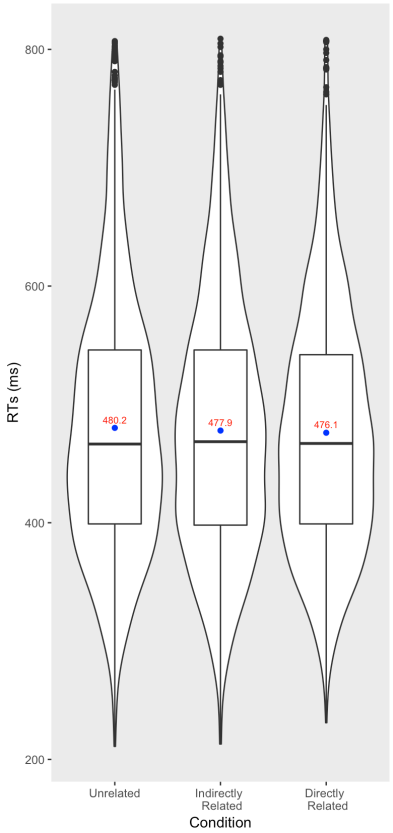


Figure 4.11: Distribution of reaction times (RTs) for all three experimental conditions.

Responses were significantly faster in the Directly Related condition compared to the Unrelated condition, while responses in the Indirectly Related condition were not significantly different from those in the Unrelated condition. Generally, the distribution of reaction times across the three different conditions is about the same, as visualized in the violin plots in Figure 4.11, which seems to suggest that participants responded to targets at about the same speed regardless of the status of the prime with which they were presented. Although the effect of *Prime Type* reached significance in the Directly Related condition, which we would expect, based on the semantic priming literature (Meyer & Schwaneveldt, 1971; Moss, Ostrin, Tyler & Marslen-Wilson, 1995; Perea & Gotor, 1997), the effect size (∼7 ms) is small; this is also evident in Figure 4.11, which shows little variation in the mean reaction times (in red) across all three conditions.

4.4.4 Discussion

These results seem consistent with those from Experiment 1, where Hanja characters were also hypothesized to play an active role in the semantic access of Sino-Korean, but no significant results emerged. Given the findings from Experiment 1, we had surmised that such findings were in large part due to the presentation of Hanja characters, particularly given the variable level of awareness amongst Korean speakers. But recall that we had no basis for comparison in Experiment 1, as we only had two conditions – one which was completely unrelated in meaning and another which required a mediated relationship; we had no condition for which there was non-mediated direct semantic priming.

Our collective findings from both experiments demonstrate that, while the Directly Related condition elicited reaction times that were significantly faster than those observed in the Unrelated condition, these reaction times are actually not all that much faster (∼1 ms) than those observed in the Indirectly Related condition. This is to say that the Indirectly Related condition may in fact elicit behavior that is very similar to that which is observed in the Directly Related condition, but such effects are not borne out in the statistical analysis. While it is clearly the case that there was enough experimental power to detect even an effect size of ∼7 ms in Directly Related condition in Experiment 2, it is not particularly clear what is driving the significance. The semantic priming literature (Meyer & Schwaneveldt, 1971; Moss, Ostrin, Tyler & Marslen-Wilson, 1995; Perea & Gotor, 1997) offers robust evidence for a mental lexicon structured around semantic relatedness. This literature shows that targets which are preceded by a directly related prime are activated significantly faster than those preceded by an unrelated prime, so it seems curious that in Experiment 2, a larger effect size was not observed in the Directly Related condition.

One immmediate consideration that comes to mind is whether cross-script activation, as investigated here, or the potential for spreading activation across different scripts, could simply be accounted for by phonological priming. Although that may be a possibility, given that our participants were adult speakers of Korean,

and thus literate to some extent in Hanja characters by virtue of being native speakers, it seems unlikely that such information would be completely ignored. Once a speaker is equipped with a certain set of knowledge, particularly information that is part and parcel of a core aspect of the language, namely its script and writing system, it seems counterintuitive to presume that such information would not be integrated in some way during lexical processing.

The objective of this thesis, and more specifically, the first two studies here, is to carve out additional yet necessary space for the role that script relations had to play, given the deeply-rooted nature of the scripts' complexities to the Korean lexicon. The next step is to better understand how the interrelated links between the scripts work with or against the role of phonology during Sino-Korean lexical processing. An understanding of how the multiple semantic contributions, as brought about by the idiosyncrasies of the Korean writing system, are tied to the structure of the mental lexicon, represents an important first step, and will also help to elucidate the extent to which phonology plays a role. The specific way in which the contributions of Hanja characters are represented in the mental lexicon are likely to vary from speaker to speaker, so whether phonological information is encoded at all within such representations is not necessarily an assumed given, and in fact, remains an equally open question. Thus, it is important to investigate other aspects of the script, such as the way in which the semantics is encoded, particularly given inherently logographic nature of Hanja characters.

Another possibility is that although we controlled for the selection of our targets to the best of our ability, there might still be by-item variance that is driving the smaller than expected effect size in the Directly Related condition. In other words, perhaps it is the case that in fact for certain Sino-Korean morphemes, semantic access to the various Hanja meanings is more or less accessible or viable, thereby affecting lexical access to the specific target. While access to certain targets is facilitated by their semantically related prime, other targets do not show facilitation, so the small effect size could be the result of this interaction, where the facilitative priming is being 'cancelled out' by inhibitive priming. We explore how this might

change with auditory presentation of Hanja characters and their Sino-Korean lexicalizations in the next chapter. This discussion will be revisited in Chapter 6, where we will review in full the findings from all four experiments, and situate them in a larger theoretical context concerning the structure of the Korean mental lexicon and the status of individual Hanja characters.

5

Conceptual encoding of Hanja

Contents

5.1 Introduction 109
5.2 A cohort of cross-script meanings 110
5.3 Experiment 3: Auditory Hanja fragment priming 119
5.3.1 Methods 120
5.3.2 Results 122
5.3.3 Discussion 123
5.4 Experiment 4: Auditory Sino-Korean full-word priming 124
5.4.1 Methods 125
5.4.2 Results 127
5.4.3 Discussion 132

5.1 Introduction

Our preliminary results from Chapter 4 offer mixed evidence for a distinct representation of Hanja, insofar as printed Hanja characters are concerned. When Hanja characters are presented visually, whether represented directly by the Chinese characters themselves, or indirectly via the Hangul representation, the results are varied. On the evidence of the two studies reported in the previous chapter, the findings suggest that spreading activation across different scripts, i.e., from Hanja to Hangul, is neither robust nor particularly viable, but chances of lexical access through mediated semantic priming are increased when streamlined into one script, Hangul. The access to Hanja is itself less explicit, but the mediated access via the Hangul

script facilitates some degree of semantic priming, as observed in the findings from the intra-modal experiment.

Following on from the experiments in Chapter 4, this chapter discusses the details of the final two experiments investigating the status of Hanja characters in the mental lexicon. Similar to the experiments presented in Chapter 4, the ones presented here also employ a semantic priming lexical decision task; however, unlike those presented in Chapter 4, which used a visual intra-modal paradigm, the experiments here use a cross-modal paradigm with auditory primes. Thus, Chapter 5 explores the status of Hanja characters in auditory word recognition.

The first experiment (Section 5.3) presents Hanja characters as auditory fragments. Fragment in the context of auditory word recognition in this thesis is hereafter defined as a morphemic syllable. The second experiment (Section 5.4) presents Hanja characters as fully lexicalized spoken Sino-Korean words. The chapter finishes with a brief discussion of the theoretical implications of the two experiments as well as providing an overarching comparison across all four experiments of the status of Hanja characters as modality-dependent representations. This collectively sets the stage for our final analysis of the structure of the Korean mental lexicon to be presented in Chapter 6.

5.2 A cohort of cross-script meanings

The question of whether the semantics of Hanja are represented in the Korean mental lexicon invites a broader theoretical question concerning the semantic network that surrounds an individual fragment. The Hanja character alone does not instantiate a full word, and neither does its corresponding Hangul representation, but the meaning of the Hangul representation is dependent on the individual semantic contributions of these Hanja characters. This particular dynamic gives rise to important questions regarding what individual fragments, which do not have independent meanings by themselves, are capable of activating, couched in a larger semantic network. Korean, through its writing system, offers a concrete way of answering these questions.

More to the broader implication of this work, what this research shows is the immense degree of variability in input our mental architecture is capable of handling; much of the research thus far shows that we can cope with the semantic links and structures in English and other structurally similar languages using alphabetic systems (Nelson, Bennett, Gee, Schreiber, & McKinney, 1993; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Balota et al., 2006; Collins & Loftus, 1975; Hutchinson, 2003; McClelland & Rumelhart, 1985; Collins & Quillian, 1972; Rips, Shoben, & Smith, 1973; Posner & Snyder, 1975; Rhodes & Donaldson, 2008) but in looking at Korean and understanding how Korean handles multiple meanings, underscores the enormous flexibility in this architecture. However, the structures we can use to account for English can also be used to account for phenomena in Korean.

Although grounded in auditory processing, the notion of cohort can perhaps be extended to non-phonological domains. Looking to Korean, its two scripts and the residual effects of early use of Chinese characters offers a new perspective on how we might stretch this flexibility, by investigating the different ways parameters for membership to a particular cohort can be set. While many Chinese loanwords exist in the Korean lexicon due to early cultural influences, the effects of Sino-Korean on the lexicon are magnified by the duality of writing-systems and provides a particularly interesting lens through which to investigate how different words may also be linked.

Marslen-Wilson's (1987) original conceptualization of cohort is based on the idea of words activating all other words that also begin with the same initial sequence of phonemes. Based on this phonological similarity, words sharing this string of phonemes, together, form a group, or cohort. Although rooted in phonology, the construct of cohort offers robust grounds on which to explore other ways in which words can be linked, extending the idea to non-phonological aspects of the lexicon, such as scripts. The present research adds to the ongoing discussion on the organization of the mental lexicon by examining a different kind of cohort – one which is based on relationships between scripts. This is particularly relevant when thinking about how a bi-scriptal language may be accounted for by existing architectures for language processing.

Consider Figure 5.1, which illustrates the Sino-Korean monosyllabic morpheme *mi* [ㅁ|] mapping onto five different meanings, based on the five different Hanja characters:

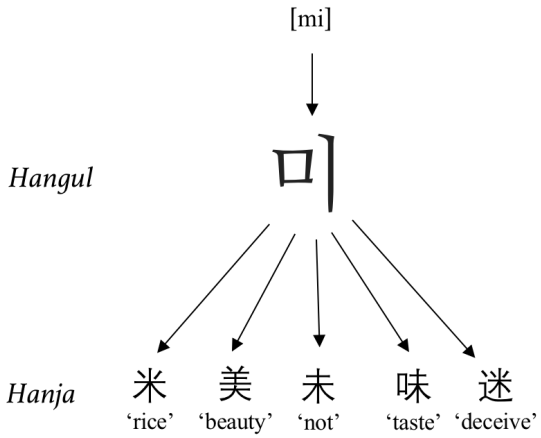


Figure 5.1: The set of possible meanings of *mi*. The representation of *mi* in Hangul can map onto a multitude of different Hanja characters, thereby encoding the number of different meanings.

It is important to bear in mind that in all five instantiations of *mi*, the Hangul form and the phonetic shape are identical [mi], but they are attributed different meanings. Unlike the shared syllable representation in Hangul, the Hanja characters are different in form. Thus, the five different instantiations of [mi] are homophones of one another and can only be differentiated without context when rendered in Hanja, and not when written in Hangul. When the Sino-Korean morpheme appears in a larger word form, the full word bears a meaning semantically related to the constituent morpheme, as illustrated in Figure 5.2:

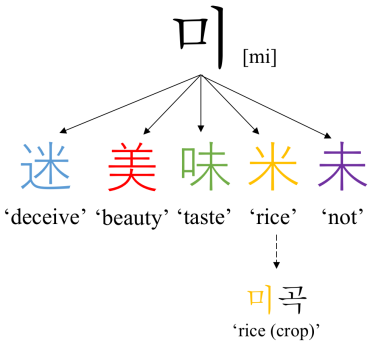


Figure 5.2: The semantic mapping of *mi* from morpheme to full word. Identical homographs in Korean map onto different Hanja characters; however, although identical, each homograph of *mi* encodes a different meaning, and each lexical instantiation is semantically related to the corresponding Hanja character (associations by color).

As a result, homographs of the same morpheme, while looking and sounding the same, each encode different meanings, and whether processing of Sino-Korean words is sensitive to such internal semantic multiplicity remains unclear.

While Figure 5.2 shows five possible meanings, this is not uniformly the case across all Sino-Korean morphemes; in fact the number of possible meanings can vary considerably. The Sino-Korean syllable *hak* (Figure 5.3) demonstrates a case whereby one Korean syllable maps onto one of two possible meanings.

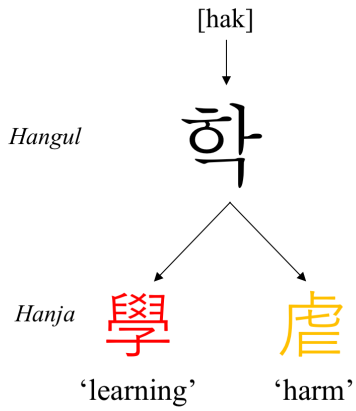


Figure 5.3: The set of possible meanings of *hak*.

Figures 5.2 and 5.3 illustrate the variability in the number of available meanings. Returning to this idea of variability and the extent to which the mental architectures are capable of coping, the type of variability we see here in Sino-Korean is not easily observed cross-linguistically, whereby sub-lexical units elicit a degree of multiplicity in meaning. Thus, in studying Korean and its idiosyncratic way of mediating multiple meanings, we are able to gain new perspectives and insights. To note, however, the emphasis here is less on the precise number, and more on the fact that multiple meanings may be available in any given case. The exact number of meanings any one Sino-Korean morpheme carries is difficult to define, as Hanja characters are borrowed in Korean.

Based on earlier findings from Swinney (1979; cf. Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Lucas, 1987; 2000) and Shillcock (1990), we would also expect exhaustive activation of all possible meanings

for any given Sino-Korean morpheme. However, this phenomenon of lexical embedding, whereby short words are embedded within longer words, is fairly common cross-linguistically. In fact, McQueen et al. (1995) estimate that about 83% of spoken polysyllabic words comprise minimally one embedded word, with embeddings frequent in initial, final and medial positions. More recent research has shown that when the task requires semantic information the meanings of embedded word stems are also activated during semantic processing (Hasenäcker, Solaja, & Crepaldi, 2019).

The question then arises of how the speech perception system is able to identify, seemingly so effortlessly, the embedding word, rather than the sequence of the many embedded words present within the input. Spoken language poses additional challenges given how noisy the speech channel can be, making already poorly marked boundaries even more opaque. One possible answer is that listeners are sensitive to and thus exploit sub-lexical cues that are probabilistically associated with word boundaries, such as metrical stress, phonotactic regularities, and acoustic-phonetic variants (cf., Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003).

Korean poses different challenges, but the idea that each constituent morpheme in a Korean word potentially encodes, or embeds, more than one meaning, is similar in principle to the idea of smaller words embedded in larger words. The difficulty caused by opaque boundaries in languages like English may be mitigated against by the highly visible block-like arrangement of the syllable in Korean. Thus, while the boundaries are clearer in terms of embedded units in a larger word in Korean, there is a greater degree of competition at the sub-lexical level due to the semantic contributions of the Hanja characters. In English, on the other hand, a word like *seat* may co-activate of a cohort consisting of such word forms as *sea*, *eat*, and *eat*, but while these embedded words are not necessarily strictly syllabified in the larger word *seat*, it is less likely to be the case, that each of the word forms – *sea*, *eat* and *eat* – carry more than one meaning, and thus compete for identification.

Processing Sino-Korean words, like *misul* [미술], 'art', necessarily involves coping with morphemes which have more than one meaning, like *mi*. Given the evidence for

multiple and exhaustive access (Swinney, 1979; Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982), we would expect coactivation of all five meanings (Figure 5.1), but eventual suppression of the inappropriate meaning based on context. In any given representation of Sino-Korean, how are the contributions of Hanja characters represented, and how are such effects modulated by a number of cohort members? Figure 5.4 provides a window into this comparison of lexical processing between Sino-Korean morphemes of varying semantic cohort size. The morpheme on the left in Figure 5.4, [hak], provides fewer semantic links than [mi], showing fewer connections through which activation of semantic associates could be facilitated.

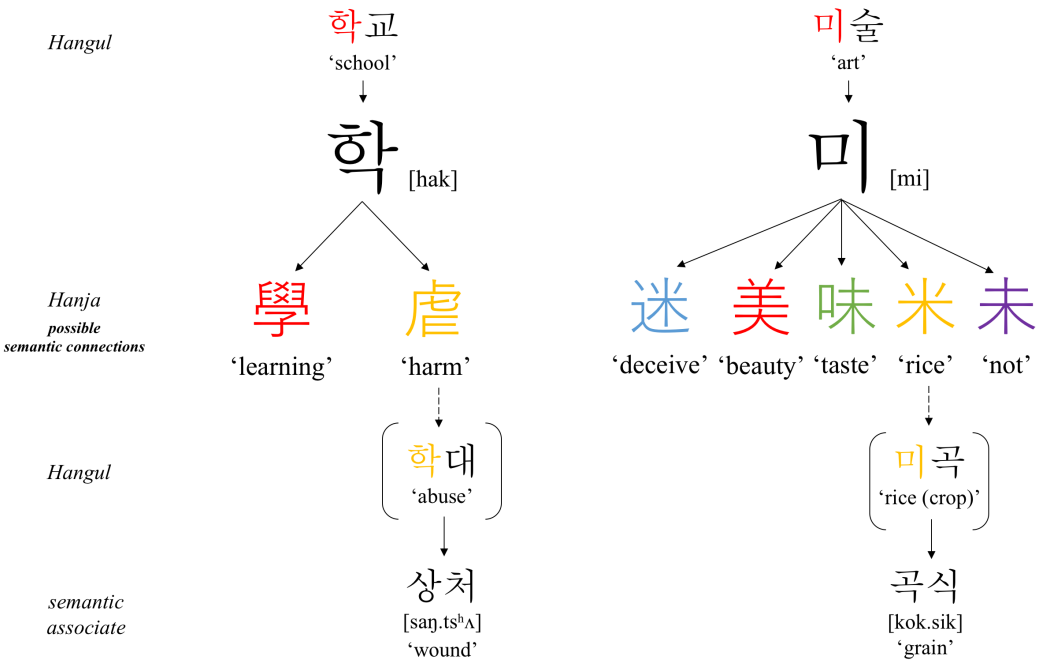


Figure 5.4: Semantic access mediated by shared homographs *hak* or *mi*. The shared homograph mediates semantic access to different semantic connections, as contributed by the different Hanja characters. Homographs – *hak* or *mi* – vary in the number of semantic connections, which may influence the viability of lexical access to semantic associates via mediated routes.

Recall in Figures 5.2 and 5.3 the relative difference in the number of meanings a given Sino-Korean morpheme can encode. The morpheme *hak* can map onto two different meanings: 'learning' and 'harm', so is it possible for a Sino-Korean word like [hak.kjo], 'school', which contains [hak] as mapped onto the 'learning' meaning, to activate a word that is semantically related to the secondary meaning,

'harm', such as 'abuse', as instantiated by 학대 [hak.dɛ]? If mediated access to secondary and tertiary meanings is possible, is there a processing 'advantage' in having more or fewer meanings available in the cohort?

Is it possible for Sino-Korean morphemes to activate not just the immediate meaning supported by context, but also secondary and tertiary meanings? Our hypothesis is that lexical access from one meaning to another is mediated by the shared homograph representation in Korean (i.e., 학), but the strength of this mediation is likely to be modulated by each new semantic connection made by every additional Hanja character that contributes a meaning to the shared homograph. The mediated relationship that facilitates semantic access across all available meanings will depend on the cohort, or set of meanings, available to each Sino-Korean morpheme. We hypothesize that semantic access from 'school' to 'abuse' to 'wound' is viable, as mediated by *hak*, and likewise, from 'art' to 'rice (crop)' to 'grain', via *mi*. We conducted a cross-modal lexical decision task with semantic priming, presenting Sino-Korean words, rather than individual fragments, containing a morpheme with more than one meaning, to investigate whether such morphemes can facilitate lexical access to a word semantically related to another available meaning, and how cohort size may influence semantic access.

Following from our hypothesis, the prediction is that because of their close semantic relationship to Hanja characters, Sino-Korean fragments alone should be able to prime any target that is semantically related to one of the activated meanings in the cohort, and that perhaps an increasing number of cohort members bolsters lexical access due to a greater number of semantically supportive links. Equally, lexical access of the Sino-Korean word may take longer for the same reasons; the constituent morphemes are individually linked to more meanings, and thus compiling all meanings and selecting the appropriate one based on the lexical and even the broader sentential context may require more time. Several studies have explored the effect of number of meanings or senses on lexical processing. The early visual word recognition literature suggests that when all words are matched for frequency, those with many meanings are recognized faster than those with fewer,

in an effect known as the "ambiguity advantage effect" (Rubenstein et al., 1970; Jastrzembski, 1981; Kellas et al., 1988; Millis & Button, 1989; Hino & Lupker, 1996; Azuma & Van Orden, 1997). Ambiguous words benefit from the multiplicity of their entries in the lexicon, which speeds up processing.

It is important to note, however, that such reasoning fails to account for structural differences between different types of ambiguity, as observed in the distinction between homonymy (ambiguity between unrelated lexemes) and polysemy (ambiguity between related senses) (Swinney, 1979; Borowsky & Masson, 1996; Rodd et al., 2002; Klepousniotou, 2002.)(See Eddington & Tokowicz (2015) for an exhaustive review of the semantic ambiguity literature, discussing how different types of ambiguity (e.g., polysemy and homonymy) affect processing, while taking into account a number of different factors such as semantic similarity, meaning or sense frequency, task, timing and modality.) To clarify, the focus of this research is not directly related to the topic of ambiguity, per se, but rather the more general idea of having access to multiple meanings; how then the resulting ambiguity is resolved is slightly orthogonal. Thus, the sub-fields of polysemy and homonymy resolution within the body of ambiguity research are only adjacently related, but they offer a larger context in which to couch the type of multiplicity we see in Sino-Korean.

Korean offers yet a different take on the matter, as the semantic distance that separates the primary and secondary meanings of a Sino-Korean morpheme can vary significantly from that which separates the secondary from the tertiary. That is, the different meanings the different Hanja characters contribute can be more or less closely related to one another, falling on a continuum ranging from homonymous to polysemous forms. Meanings that are related to one another to varying degrees are all encapsulated by a single Sino-Korean representation. The larger set to which these different meanings belong, for any particular Sino-Korean morpheme, comprises a number of different semantic relations of varying semantic strength (more or less related), the number also varying significantly. With Sino-Korean morphemes, not only is the number of corresponding meanings rather variable, but also these meanings are less systematically related to each other than a comparable set of

native Korean homonyms/polysemes. Thus, the semantic relations between the meanings for a particular Sino-Korean morpheme are not necessarily categorically related, which may pose challenges for how they are represented in the mental lexicon, given semantic network theory, as discussed in Chapter 3.

Furthermore, Sino-Korean morphemes are not ambiguous in the same way as whole words. The appropriate meaning of a particular word can be resolved from the larger sentential context, but the potential ambiguity arising from an individual morpheme in Korean is resolved almost immediately given the larger context of the word. Yet, it is uncertain whether semantic access in this particular case is bolstered by the existence of more or fewer semantic relations (e.g., comparing [hak] and [mi], Figure 5.3). Consistent with the aforementioned literature, we, too, predict that since all meanings are activated in the first place, ahead of contextual effects, accessing the appropriate meaning of the given Sino-Korean morpheme should be faster. The more members there are in a cohort, the more reliable links there are with which to bootstrap semantic selection. The fewer the members, the fewer 'resources' or cues, outside of broader contextual support, there are to exploit. Those which have larger cohorts essentially benefit from a larger pre-activated network.

Another important aspect to consider in all of this is that the degree to which all members of a given cohort are engaged may be task-dependent; that is, naming tasks may allow for a decision using only the initial status of the cohort, while others may require complete checking of the cohort, such as in a lexical decision task. The potential effects of similarity also point to letter-level encoding, a level of abstract linguistic information that also needs to be accounted for by models of visual word recognition. This to say that the notion of the cohort, however it is defined and at whichever level it is posited to operate, is not necessarily concrete in nature; there is much that remains to be learned about its construction, composition, and effects, both within one language and across the board. Thus, experimentally, cohort size has been found to have important effects on lexical access in both spoken and visual word recognition, though the exact size and direction of the effect remains inconclusive, with studies only revealing general tendencies, rather than consistent

patterns, and appear relatively varied across different paradigms and studies (Chen et al., 2011; Dufour & Frauenfelder, 2010; Frauenfelder & Tyler, 1987; Gahl & Strand, 2016; Gahl, Yao, & Johnson, 2012; Grainger, 1992; Gwilliams et al., 2018; Henderson et al., 2013; Luce & Pisoni, 1998; Müller, Duñabeitia, & Carreiras, 2010; Munson & Solomon, 2004; Vitevitch et al., 1999; Vitevitch & Rodríguez, 2005).

We build on this discussion by continuing our investigation of the status of Hanja in the Korean mental lexicon, which for purposes of parsimony, we assume makes use of the same architecture used in language processing more broadly. Our previous experiments in Chapter 4 probed the possibility of a script-based representation of Hanja in the mental lexicon. But given the mixed results, we continue our investigation by considering auditory representations of Hanja, moving to a cross-modal paradigm in our lexical decision task. We build on our prior findings from Experiments 1 and 2 to explore a more conceptual-based encoding of the semantics of Hanja.

5.3 Experiment 3: Auditory Hanja fragment priming

In Experiment 1, we observed that Hanja characters, when presented visually, were unable to activate indirectly-related semantic meaning in Korean. In the present study, we changed the modality of the primes, from visual to auditory, while keeping the presentation of the targets the same, to investigate exactly how the semantic information is encoded, if it is not via script. This change in modality also introduces a new aspect to our investigation: Hanja characters that would be homophonous if not for their different tone assignments end up as homophonous in Korean, since Korean is not a tonal language. Importantly, this means that the many different meanings individually encoded by distinct Hanja characters are mapped onto the same Korean homophone. The change from a visual intra-modal to an auditory-visual cross-modal task allows us to examine whether representation of Hanja is stable across modalities.

5.3.1 Methods

5.3.1.1 Materials and design

To mitigate against the potential effects of variable proficiency and language mode, this experiment was carried out in Seoul, South Korea, where, importantly, speakers have more day-to-day exposure to printed Hanja characters. The stimuli were also controlled more carefully, taking into account the relative number of meanings, or the size of the semantic cohort, for individual fragments. For each target in the present experiment, there are two conditions. Exemplar stimuli are presented in Figure 5.5.

1. *Related prime.* These primes are semantically related to the meaning presented in the target.

2. *Unrelated prime.* These primes are not semantically related to the meaning presented in the target.

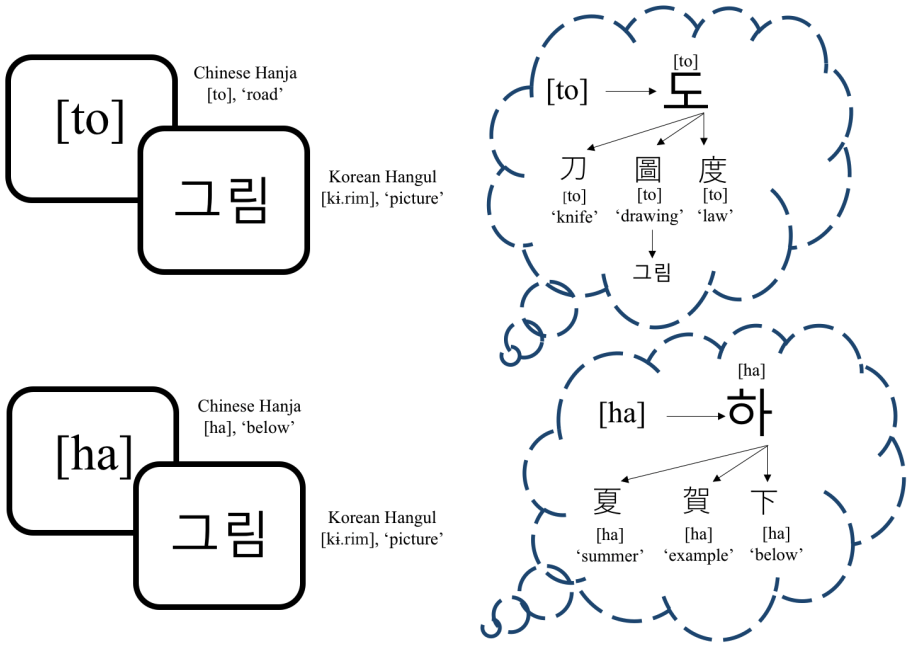


Figure 5.5: Overview of test stimuli.

The auditory stimuli for this experiment were recorded by a female native speaker of Standard Seoul Korean. The corresponding audio files were then edited using

Praat (version 5.1.32). They were cut into individual syllables at the nearest zero crossings, with every effort taken to minimize audio clicks. In this study, the total number of targets per list was 144, of which 72 were real word targets, 36 preceded by a semantically related prime, and the other 36 by an unrelated prime. These trials were divided into 12 blocks of six trials.

5.3.1.2 Participants

134 Korean native speakers participated in this experiment, all of whom were students at one of the three universities at which the experiment took place, had normal or corrected-to-normal vision with no history of neurological disorders, and no reading impairments. All participants provided consent for participation and were compensated for their time.

5.3.1.3 Procedure

The experiment, similar to Experiment 2, took place in quiet rooms located at three different universities in Seoul: Sookmyung Women's University, Hansung University and Kyung Hee University. The experiment was controlled using *Presentation* software (Neurobehavioral Systems) and displayed on a 14-inch Dell Latitude 7480 laptop. A Latin Square design was used to ensure that every participant only saw each prime and target once while allowing for maximal use of all possible prime–target combinations. For each trial, participants were first presented with the auditory prime stimulus via headphones (Sennheiser PX200-II).

These were then immediately followed by visually presented target words (ISI = 0 ms) with a display time of 300 ms. The target words were presented in font-size 85pt. Participants were given up to 1500 ms to make a lexical decision. Participants were asked to decide as quickly and accurately as possible whether or not the item on the screen was a word in Korean, using the button box provided. All YES responses were made with the participant's dominant thumb. Each experimental session was preceded by a practice session comprising 10 practice items to help familiarize the participants with the task.

5.3.2 Results

Three participants were excluded from the analysis (incorrect responses > 25%), and three targets had to be removed (incorrect responses > 25%). Only responses within two SD of overall mean reaction time to targets were included. Incorrect responses were also excluded. Further analysis of frequency measures showed four additional items to be outliers[1] and these were not submitted for further data analysis. Overall, 15.90% of the data was excluded from the analysis.

A linear mixed effects analysis was performed for the remaining reaction times using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2018). *Participant* and *Target* were included as random effects (Baayen, Davidson, & Bates, 2008), and treatment coding with the reference level Unrelated was applied to our fixed effect, which was defined as *Prime Type* (Related and Unrelated). Through model comparision, it was established that a model containing random intercepts for *Participant* and *Target* provided the best fit for our data, while inclusion of *Cohort* as a nested factor within *Prime Type* did not improve model fit ($\chi^1(6) = 0.162$, $p = .688$). Thus, our best fit model[2] did not include *Cohort* as a continuous variable.

We report the coefficients, with p values approximated by the normal distribution (Barr et al., 2013) in Table 5.1.

Table 5.1: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	453.049	5.698	292	79.509	< .001*
Related	-6.091	1.492	7733	-4.083	< .001*

As Table 1 indicates, responses were significantly faster in the Related condition than in the Unrelated condition. We visualize the distribution of reaction times in the following violin plots in Figure 5.6.

[1] Further analysis revealed that frequencies for two of the targets lay outside our cut-off for outlier rejection (average + 2*SD).

[2] lmer(RT ~ PRIME TYPE + (1 | PARTICIPANT) + (1 | TARGET)).

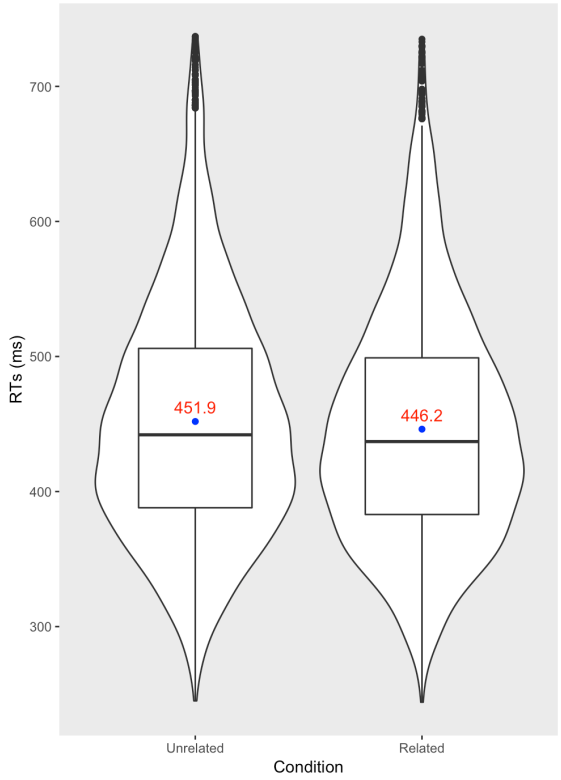


Figure 5.6: The distribution of reaction times in both the Related and Unrelated conditions.

The finding that responses were significantly faster in the Related condition compared to the Unrelated condition suggests that auditory presentation of the Hanja character alone activated an exhaustive enough cohort of possible meanings, such that one of activated members of the cohort semantically matched the meaning of the word presented in the target. This is consistent so far with the findings in support of multiple and exhaustive semantic access (Swinney, 1979; Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Lucas, 1987; 2000), which suggests that all possible options are entertained and more than one meaning is activated.

5.3.3 Discussion

We found in our present study that including *Cohort* as a nested factor within *Prime Type* did not improve model fit. This suggests that the existence of more or fewer meanings has no particular effect on identification; that is, it neither bolsters

nor inhibits activation by way of competition. This seems plausible for fragments, as there is no neighboring morpheme to further contextualize and thus restrict the range of activated meanings. Figure 5.7 visualizes the stable effect of cohort on the distribution of reaction times across increasing cohort size.

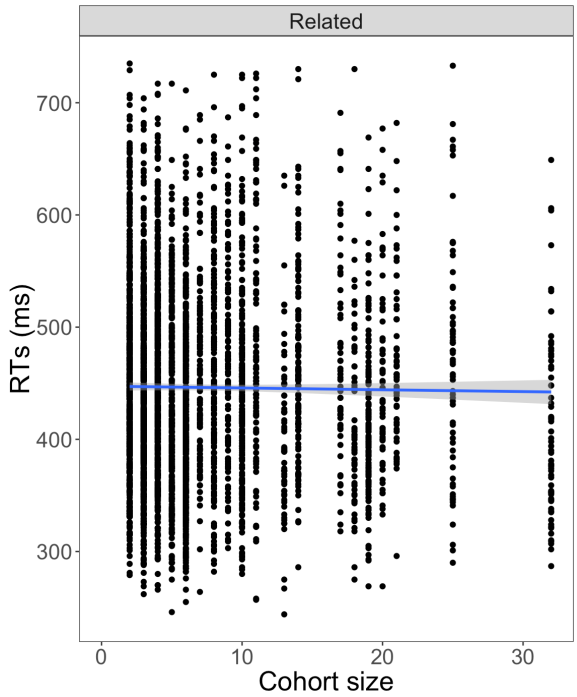


Figure 5.7: Distribution of reaction times (RTs) for the Related condition for Experiment 3, the blue line representing a fitted linear regression, and the shaded grey showing SE. An additional meaning or member that is added to the cohort does not affect reaction times.

The nested effect of *Cohort* may affect Sino-Korean morphemes differently when lexicalized. Thus far, auditory presentation of Hanja appears to yield stronger priming effects compared to visual presentation of Hanja (Experiment 1), even if effects of cohort are not observed.

5.4 Experiment 4: Auditory Sino-Korean full-word priming

Building on the findings of our previous experiment, here we test the hypothesis that information from Hanja characters is encoded in the mental representations of Sino-Korean words, not necessarily as specific characters, but as semantic concepts.

Thus, the representations are not confined to the entry in the mental lexicon relating to the character itself, which would necessitate literacy in Hanja. This hypothesis predicts, as corroborated by the results from Experiment 3, that Sino-Korean morphemes activate more than one concept simultaneously, and that the shared representation in Hangul mediates access between multiple concepts.

5.4.1 Methods

5.4.1.1 Materials and design

All prime and target stimuli were disyllabic Korean words, and all primes were Sino-Korean. We devised 36 prime-target pairs based on a set list of Hanja characters taught in Korean secondary schools to maximize the likelihood of participants knowing the Sino-Korean homographs. Each prime contained a Sino-Korean morpheme, the homographs of which mapped onto two equivalent Hanja characters, and thus two different meanings.

The selected Hanja characters were among those identified by the Korean Ministry of Education as Chinese characters to be learned for educational purposes, and therefore deemed recognizable to the average native Korean speaker. No frequency counts are available for individual Sino-Korean morphemes, nor is there any frequency measure for actual Hanja characters in text. However, the word-frequencies of the resulting Korean words were controlled for as well as possible given these constraints.[3]

The same considerations in the design of stimuli with regard to a mediated relationship were applied here as in Experiment 2. Recall the following paradigm in Figure 5.8, first presented in Chapter 4 in our explanation of Experiment 2.

[3]Once primes and targets had been selected (based on the relevant Sino-Korean morpheme), we used a frequency database collated by the National Institute for Korean Language to control for frequency as best possible (National Institute for Korean Language, 2008).

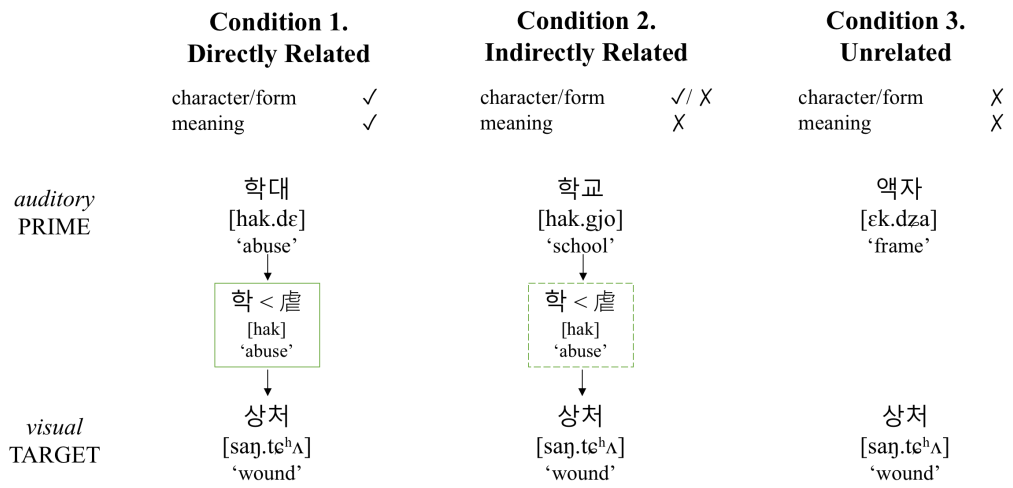


Figure 5.8: Experimental conditions. Depending on the condition, the prime-target pair involved a direct or indirect semantic relationship that required mediated semantic access via shared homograph (e.g., *hak*).

We also conducted a survey of 30 native speakers of Korean to ensure that relevant judgements of semantic relatedness were consistent for our experimental conditions. Pairwise t-tests across the three conditions showed no difference in word frequency, and an additional a one-way ANOVA over our three main experimental conditions also identified no significant difference in frequency ($F(2, 98) = 0.131$, $p = 0.877$). For purposes of a balanced design in the number of word and non-word stimuli, 36 additional prime-target pairs with non-word targets were included, resulting in a total of 72 items divided into 12 blocks of six trials. The auditory stimuli were recorded by a female native speaker of Standard Seoul Korean.

5.4.1.2 Participants

189 native speakers of Korean participated in this experiment, none of whom had taken part in any of the previous studies. All participants were undergraduate students from three different universities in Seoul, Korea (Hansung University, Kyung Hee University and Sookmyung Women's University). Participants were native speakers of Korean, had normal or corrected-to-normal vision, and had no language or reading impairments. Participants provided their informed consent and were reimbursed appropriately.

5.4.1.3 Procedure

The experiment took place in quiet rooms located at the different universities. The experiment was controlled using *Presentation* software (Neurobehavioral Systems) and displayed on a 14-inch Dell Latitude 7480 laptop. A Latin Square design was used to ensure that every participant only saw each prime and target once while allowing for maximal use of all three possible prime–target combinations. On each trial, participants were first presented with the auditory prime stimulus via headphones (Sennheiser PX200-II). The targets were presented visually immediately after the prime for a duration of 300 ms. Participants were given up to 1,500 ms to make a lexical decision. Before the experiment, participants completed a practice session comprising 10 practice items. Participants were instructed to respond to every item on the screen by deciding whether it was a word in Korean, using the button box provided. YES responses were made with the participant's dominant thumb.

5.4.2 Results

Four participants were excluded from the analysis (incorrect responses > 25%), and two targets had to be removed (incorrect responses > 25%). All reaction times outside ±2 standard deviations from the overall mean were also excluded. Only correctly classified targets were submitted to analysis. Overall, 14.53% of the data was excluded from the analysis. A linear mixed effects analysis was performed for the remaining reaction times using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2018). *Participant* and *Target* were included as random effects (Baayen, Davidson, & Bates, 2008), and treatment coding with the reference level Unrelated was applied to our fixed effect, *Prime Type* (Directly Related, Indirectly Related, Unrelated).

Our best model fit[4] contained a continuous variable *Cohort* as a nested factor within *Prime Type*, which was included to investigate the effects of cohort size. By comparing models with and without the relevant nested factor, we found that the factor *Cohort* nested within *Prime Type* was significant at $\chi^2(2) = 8.88$, $p =$

[4] lmer(RT ~ PRIME TYPE / COHORT + (1 | PARTICIPANT) + (1 | TARGET)).

.012. We report the model coefficients representing simple effects, with p values approximated by the normal distribution (Barr et al., 2013), in Table 5.2.

Table 5.2: Results of linea mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	453.216	6.218	69	72.883	$< .001$*
Directly related	-13.136	3.077	5255	-4.269	$< .001$*
Indirectly related	2.31	3.08	5238	0.750	.453
Directly related × *Cohort*	0.128	0.275	4374	0.467	.641
Indirectly related × *Cohort*	-0.643	0.275	4375	-2.342	.019*

Responses were significantly faster in the Directly Related condition than in the Unrelated condition. While responses were not significantly faster in the Indirectly Related condition compared to the Unrelated condition, the nested interaction of *Cohort* in the Indirectly Related condition produced a significant coefficient estimate (β = –0.643, t = –2.342, p = .019). Only in the Indirectly Related condition did targets with larger cohorts elicit significantly faster responses times than those with fewer. This can be seen in Figure 5.9 (left panel), which illustrates the incremental decrease in mean reaction times as the number of members present in the cohort increases, across all targets.

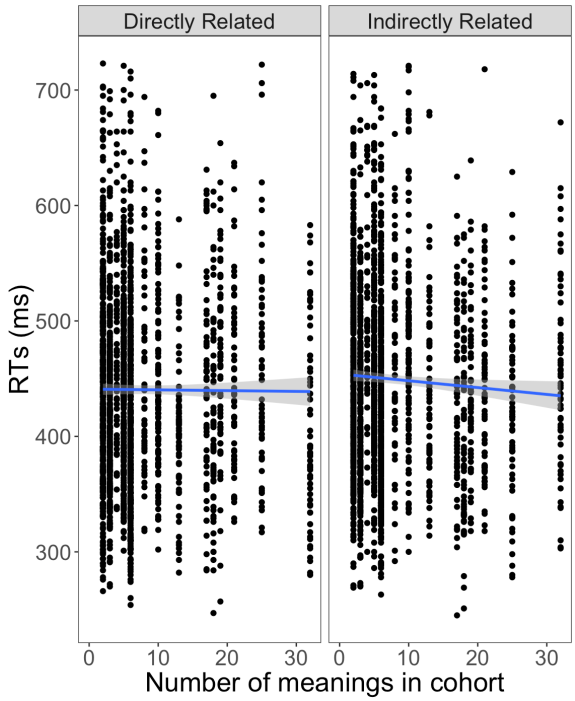


Figure 5.9: Distribution of reaction times (RTs) for the Indirectly and Directly Related conditions, the blue line representing a fitted linear regression, and the shaded grey showing SE, for Experiment 4 (audio-to-visual cross-modal priming). Only in the Indirectly Related condition did targets with larger cohorts elicit significantly faster reaction times than those with smaller, with each additional member significantly decreasing response times.

The right panel shows that in the Directly Related condition, reaction times remain stable across all cohort sizes, which is also shown by the results presented in Table 5.2, showing that the nested interaction with Cohort did not reach significance.

Given the significant effect of semantic cohort size, these results raised additional questions regarding the possible role of semantic salience. For example, it may be the case that one of the available meanings is more or less salient relative to the other, or that the two meanings are equally salient. Possible differences in semantic prominence may have potential effects on lexical access, especially if meanings which are more or less salient are embedded in a larger cohort of competing meanings. We investigate here via post hoc analysis whether semantic salience interacts with semantic cohort size. From this, there are two predictions: (1) if one meaning is particularly salient relative to the other available meaning, activation is unaffected by semantic cohort size; (2) if meanings are equally salient, activation is affected by increasing semantic cohort size.

When devising the stimuli, we controlled as best as possible to ensure that Korean speakers would be familiar with the two meanings of the Sino-Korean morpheme used for the Directly and Indirectly Related conditions, based on the corresponding Hanja characters. However, semantic salience was not directly controlled for given these restrictions, hence the post hoc analysis we present now. To gauge semantic salience of the two meanings, we conducted a survey where 30 native speakers of Korean (none of whom participated in any of the main experiments in this thesis) were asked to decide whether, of the two meanings, one was more salient, or if both meanings were equally salient.

These judgements were then coded into the dataset. Due to the analysis being post hoc, the distribution of tokens within each category is not uniform. Figure 5.10 offers visualizations of the data according to semantic cohort size and semantic salience, which reveal some interesting trends.

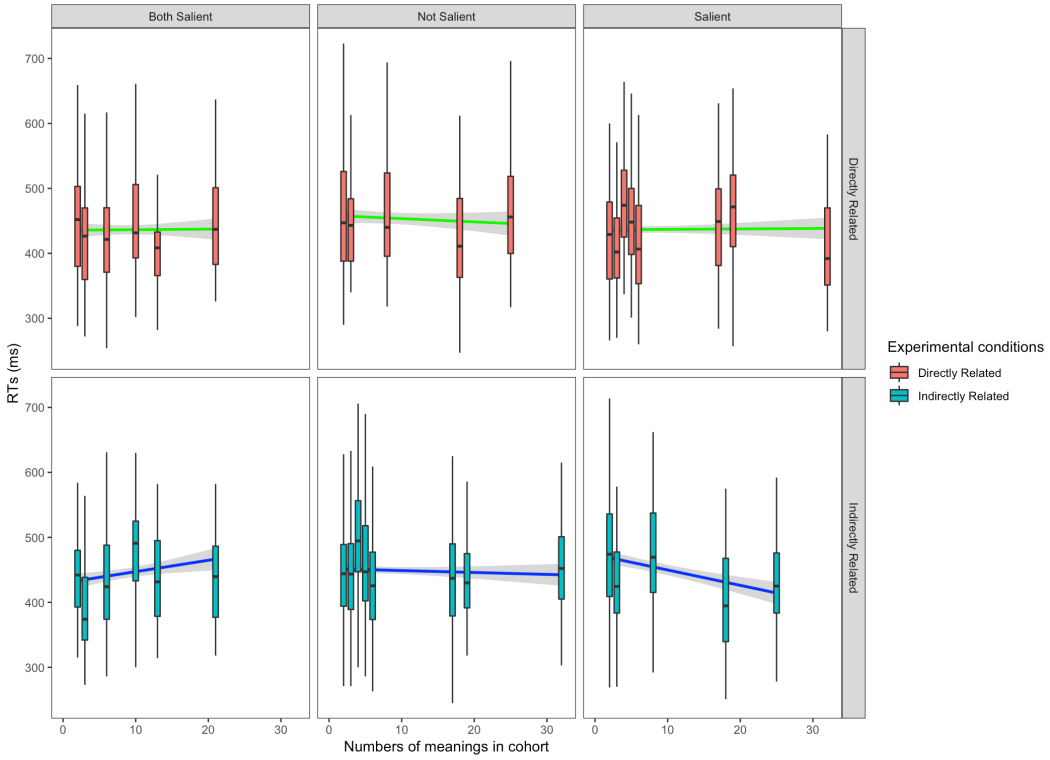


Figure 5.10: Distribution of reaction times (RTs) for the Indirectly and Directly Related conditions in Experiment 4 (cross-modal), according to semantic salience of the presented meaning, and the relative number of other meanings with which the presented meaning is competing. The blue line represents a fitted linear regression, and the shaded grey shows SE.

The first set of panels presented under the label "Both Salient" show the reaction times to targets that were preceded by primes which had two equally salient meanings; that is, the two meanings used across the two experimental conditions, Directly and Indirectly Related, were judged to be equally salient. The second pair of panels presented under the label "Not Salient" presents the reaction times to targets that were preceded by the less semantically dominant meaning of the prime. The final two panels given under the label "Salient" show the reaction times to targets that were preceded by the more semantically salient meaning of the prime.

Reaction times are predicted to be significantly faster in the later two panels, than those displayed in the other four quadrants, as targets are preceded by primes that are not only semantically related, but also semantically salient. The trends demonstrated in Figure 5.10 broadly suggest that in the Directly Related condition, reaction times are relatively stable with increasing semantic cohort size as well as varying levels of semantic salience (observed in the top three quadrants). This is less apparent in the Indirectly Related condition, where the bottom three quadrants seem to illustrate high levels of variability in the interaction between semantic salience and semantic cohort size. As a caveat to these trends, all four quadrants in Figure 5.10 highlight the uneven distribution of data points in the different cohort sizes. In many ways, this is inevitable and perhaps an artifact of the nature of new borrowings into a language, as it is hard to define when a borrowing is truly complete, making the boundaries between different cohort sizes less obvious.

The best-fitting model model[5] contained the interaction between *Cohort* and *Saliency* (Salient, Not Salient, and Both Salient)[6] as a nested factor within *Prime Type*, with *Participant* and *Target* included as random effects. By comparing models with and without the interaction, we found that the nested interaction between *Cohort* and *Saliency* within *Prime Type* was significant at $\chi^2(6) = 21.04$, $p = .002$. We report the coefficients, with p values approximated by the normal distribution (Barr et al., 2013) in Table 5.3.

[5] lmer(RT ~ PRIME TYPE / (COHORT*SALIENCY) + (1 | PARTICIPANT) + (1 | TARGET)).

[6] These levels are rank deficient in the model due to low frequency in the dataset and are thus dropped, hence absent from the model output.

Table 5.3: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	453.268	6.487	62	69.875	$< .001$*
Directly related	4.963	7.341	4850	0.676	.499
Indirectly related	-0.442	7.277	4826	-0.061	.952
Directly related × *Not salient*	-23.780	9.637	4660	-2.468	.014*
Indirectly related × *Not salient*	11.394	8.381	4677	1.360	.174
Directly related × *Salient*	-18.880	8.411	4687	-2.245	.025*
Indirectly related × *Salient*	-11.993	9.541	4618	-1.257	.209
Directly related × *Cohort*	-2.096	0.697	4685	-3.007	.003*
Indirectly related × *Cohort*	-1.074	0.691	4655	-1.554	.120
Directly related × (*Cohort* × *Not salient*)	2.99	0.874	4557	3.419	$< .001$*
Indirectly related × (*Cohort* × *Not salient*)	0.509	0.778	4690	0.653	.514
Directly related × (*Cohort* × *Salient*)	2.346	0.782	4707	3.001	.003*
Indirectly related × (*Cohort* × *Salient*)	0.665	0.868	4522	0.767	.444

Responses were significantly faster in the Directly Related condition than in the Unrelated condition, while reaction times for targets in the Indirectly Related condition were varied. Only in the Directly Related condition did the interaction between *Cohort* and *Saliency* reach significance ($\beta = 2.346$, t = 3.001, $p = .003$).

5.4.3 Discussion

5.4.3.1 Effects of modality on cross-script cohorts

In Experiment 2 in Chapter 4, when we presented Sino-Korean words visually, *Cohort* was not included as a nested factor within *Prime Type*, as it did not improve model fit. Figure 5.11 visualizes the distribution of reaction times for targets in both the Indirectly and Directly Related conditions for Experiment 2.

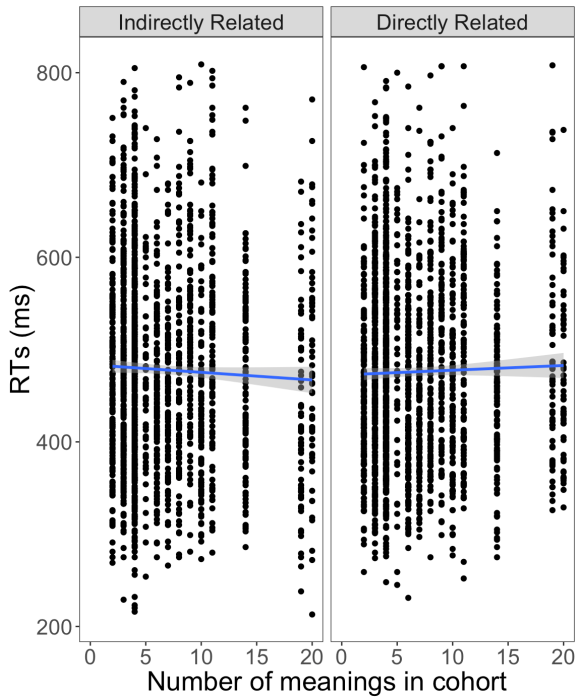


Figure 5.11: Distribution of reaction times (RTs) for the Indirectly and Directly Related conditions, the blue line representing a fitted linear regression, and the shaded grey showing SE, for Experiment 2 (visual intra-modal). Based on our model, neither condition elicited significant effects of cohort.

While the blue line seems to indicate downward and upward trends (in the Directly and Indirectly Related conditions, respectively), the trends are too weak to be detected statistically. This is also likely to be the result of non-uniform distribution of tokens across different semantic cohort sizes. This is simply due to language-specific constraints and having to create stimuli within the remit of what is available in Korean. This is further complicated by the larger constraint of ensuring that the Sino-Korean morpheme used falls within the boundaries of speakers' working knowledge of Hanja. We also tried modelling the size of the semantic cohort as a discrete, rather than continuous, variable, but found that this created artificial boundaries. Borrowings are not necessarily a static occurrence, which made it difficult to group morphemes according to number of available meanings.

Given that the only difference between Experiments 2 and 4 was the way in which primes were presented, the collective findings from both experiments suggests that whether we observe the effects of cohort may be dependent on modality of presentation. That is, the effects of cohort may be more or less pronounced

depending on whether the Hanja character is presented visually or auditorily. As we had surmised from our previous studies on visual presentation of Hanja characters (whether directly or indirectly through Sino-Korean), the representation of Hanja in the mental lexicon is not likely to be dependent on the representation of the original Chinese characters, but rather the equivalent representation in Hangul. Furthermore, visual presentation of the Sino-Korean word, where there is a local context (i.e., the neighboring morpheme) supressing incompatible meanings, may result in fewer activated, or more weakly activated meanings in the cohort, and is thus less susceptible to the effects of cohort. However, when the Sino-Korean word is presented auditorily, incremental processing allows for more immediate activation of the local semantic network of meanings, before the local context (i.e., the neighboring morpheme) intervenes.

Looking specifically at the findings from Experiment 4, the interaction between *Saliency* and *Cohort* yielded varied results, most likely again due to uneven distribution of tokens. Perhaps a more-controlled design focusing on these factors would elicit significant interactions, warranting further investigation. As it stands, the results in Table 5.3 seem to broadly suggest that words which are directly related to one another are more prone to the effects of increasing cohort size and differences in semantic salience on reaction times. When words are more distantly related by way of semantic mediation, or in our case, when the semantic relationship is made more opaque by cross-script links, there is little effect of cohort size or salience in lexical access.

Recall the discussion on semantic network theory from Chapter 3; whether there are more or fewer additional meanings attached to a particular node, or if one meaning is more or less salient, is not particularly important if that part of the semantic network is weakly activated. Targets in the Indirectly Related condition should be expected to be weakly activated by virtue of the greater distance spreading activation must travel to reach that part of the semantic network, since the connections in between are not direct, but rather, mediated. This analysis is slightly inconsistent with the analysis we presented earlier on the effects of

cohort size, particularly in the Indirectly Related condition, where it was found to have a significant effect. Although our best-fitting model included the interaction between *Saliency* and *Cohort*, our research aims were not guided by an interest in understanding the role of meaning salience. That is, the experiments in this thesis were not designed to answer such questions, so the fact that the best-fitting model is one which includes the interaction can probably be taken with a grain of salt. Regardless, this exploratory analysis reveals interesting trends nonetheless, which importantly suggest that factors other than semantic cohort size are at play (discussed more in Chapter 6).

5.4.3.2 Beyond phonological cohorts

To contextualize our research, we first introduced the idea of cohort within its original context of spoken word recognition. While it was conceptualized as a phonological construct, its application is evident in non-phonological domains. For example, orthographic cohorts have also been found to be a potential factor affecting speed of activation. Orthographic cohorts operate much in the same vein as phonological cohorts. When a word is encountered, the first incoming orthographic code activates all lexical entries containing this information (Johnson & Pugh, 1994). These models assume a stage of initial encoding of the graphical information, which is then recoded into an abstract orthographic form. Information about the identity of the target letter as well as two immediately adjacent letters (i.e., the neighboring context) is encoded. As successive pieces of orthographic information becomes available, lexical candidates that are not consistent with this information lose activation. The cohort of possible candidates is resolved when a single matching item remains.

In visual word recognition, much like in spoken word recognition, orthographic neighbors are generated by replacing one letter from the target word in the same letter position (Coltheart, Jonasson, Davelaar, & Besner, 1977) (e.g., orthographic neighbors for the word *cap* include *gap*, *cup*, *car*). A large selection of studies suggests significant influence of orthographic cohort size on visual word recognition (Andrews, 1997 and references therein for a complete review). Findings from lexical decision

tasks reveal a facilitatory effect of larger orthographic cohort sizes: words with many orthographic neighbors yield faster reaction times and increased accuracy, compared with words with fewer competitors in their orthographic cohort (Andrews, 1989; 1992; 1997; Carreiras, Perea, & Grainger, 1997; Pollatsek, Perea, & Binder, 1999).

The effects of orthography have also be shown to be closely related to those of phonological cohorts. In their 2005 study with readers of French, Grainger et al. found that the effect of larger phonological cohorts helped to facilitate activation of words with correspondingly large orthographic neighborhoods, but proved to inhibit activation for words with fewer cohort neighbors. Such findings are also consistent with those from studies exploring effects of frequency on orthographic cohort size; for example, Grainger et al. (1992) find in an earlier study that effects of inhibition are reduced when there are more neighbors of higher frequency. The interaction between orthographic and phonological cohorts suggests the potential for representations to occur across multiple different levels, and models of both spoken and auditory word recognition should be able to account for these multiple levels of different representations.

The early works of Luce and Pisoni (1998) on spoken word recognition, however, suggest a different pattern; their findings from a perceptual-identification task reveal words with *fewer* competitors in its cohort to be recognized more accurately than those belonging to denser cohorts. However, other effects may also be at play; Luce and Pisoni's findings also revealed potential effects of frequency: words in a cohort of highly frequent lexical candidates elicit delayed recognition, compared to words that are embedded in a cohort of items of lower frequency. Taken together, the identification of low frequency words in densely-packed phonological cohorts required more time than those which have higher frequencies and fewer phonological competitors. They report similar findings in both naming and lexical decision tasks, with words having larger cohort sizes yielding slower responses than words with smaller cohorts, referring to this effect as *lexical competition*. The perceptual difficulties that emerge are due to the high density of phonological cohorts, where there is increased competition amongst all possible contenders for activation.

Subsequent studies by Yates et al. (2004) controlling for possible effects of orthography, however, reveal larger phonological cohort sizes to have a facilitating effect. Although slightly different stimuli were used for different tasks, these findings proved consistent across different paradigms, from lexical decision to naming to semantic categorization, which may suggest the potential confound in exploring the effects of one cohort (phonological) without possible interference from the effects of another (orthographic). Models that do take into consideration such interactions between effects (Andrews, 1997; Coltheart et al., 2001; Grainger & Jacobs, 1996), however, offer evidence to the contrary, that, in fact, larger cohorts inhibit rather than facilitate. Given the mixed results and often conflicting trends, the state of affairs in the literature on cohorts remains inconclusive.

In some ways, the results we have shown so far from Experiment 2 (Chapter 4) and Experiment 4 from the present chapter support such mixed patterns. We found limited evidence of effect of cohort size on priming in visual presentation of Hanja, but more pronounced effects when presented auditorily. However, contrary to the patterns evident in the existing literature, in auditory processing of Sino-Korean, reaction times were faster with increasing cohort size, meaning lexical access was facilitated by denser cohorts. During visual processing of Sino-Korean, however, lexical access was not helped by larger cohorts, and in fact, remained unaffected by increasing cohort size.

Although the trends are sometimes at odds, the overarching trend that the literature presents is the opposite – larger phonological cohorts facilitate lexical access, while larger orthographic cohorts inhibit. Thus, the contributions of Hanja, in some ways, form their own type of cohort, one that is rooted in cross-script semantics. Moving beyond the phonological and orthographic cohorts, the findings presented here demonstrate that Sino-Korean morphemes form cohorts on the basis of semantic contributions derived from another script. Three meanings of different Hanja characters create a network of possible meanings for individual morphemes. Korean's use of different scripts, and subsequently how these scripts contribute meaning to words, is a direction the literature has yet to take fully.

Obviously, the one factor that remains but cannot be easily manipulated nor controlled for in this research is frequency. Unfortunately, the type of frequency counts that are needed and are used in the studies discussed so far are not available for the individual Sino-Korean morphemes, although as we mentioned before, the frequencies of the individual words were accounted for thereafter once the stimuli had been devised. We were confined to the use of words that contained Sino-Korean morphemes which used commonly known Hanja semantics, so it was difficult to account for frequency at any earlier stage.

5.4.3.3 The variable status of Hanja in Sino-Korean

Whether the semantics of Hanja are active in processing is highly dependent on the particular morpheme. Depending on the status of Hanja in a particular representation, facilitation is modulated by the relative density of semantic competition; for every additional member in the semantic cohort, priming must be distributed across a larger semantic network, thereby reducing the strength in activation of the intended meaning. This thesis provides evidence for a new cohort that is built on the basis of script and the semantic information that is encoded across different scripts. The findings in this chapter demonstrate that semantic cohorts can exist across structurally distinct scripts, within one language.

Furthermore, we see that the status of Hanja characters can be quite variable given that their semantic contributions are more or less accessible depending on the specific Sino-Korean morpheme. In spite of this, however, we see that the same structures that are used to account for Sino-Korean morphemes with more salient representations also being used for those with less salient semantic representations. Whether the priming of a morpheme benefits from an increasing cohort size or remains unaffected can be accounted for nonetheless by the same architecture.

These findings also provide insight into the structure of the Korean mental lexicon, and the potential differences in processing between native and Sino-Korean words as result of the multiple semantic links Hanja characters contribute. The processing of these links, however, may reach a point of saturation, whereby after

some number of available meanings, the facilitative effects of having multiple meanings are as strong as they can be, and adding more meanings would make no difference. How this dynamic can then be reflected in a model of the Korean lexicon forms the next step in this research. In the following chapter, we make use of a 'swinging' lexical network to account for observed priming effects between non-categorical relations, as seen in Sino-Korean.

Non-categorical relations pose a unique challenge in the priming literature, especially with regard to how such links may be represented in the organization of the mental lexicon, where two things are more likely to be linked categorically. Korean is of particular interest in this context as non-categorical relations observed in Sino-Korean are rooted in the direct contributions of the Hanja characters and arise primarily because of the idiosyncrasies of the Korean writing system. In a move towards a more comprehensive understanding of the Korean mental lexicon, one which accounts for the role of Hanja, in Chapter 6, we discuss the theoretical implications of our findings. We also perform in the next chapter additional statistical analyses exploiting unexplored by-item variance in our results.

This research has so far been oriented as a body of work specific to the Korean mental lexicon, but it is important to note that underlying the specific building blocks, concepts and structures we use are those which are used in language generally. Tying it all together, the capacity of one architecture to capture the full breadth of linguistic phenomena is the variability we remarked on at the start of this chapter, and is the variability in input we all hope is in fact captured by our model, but lack the supporting empirical data.

6

The state of the Korean mental lexicon

Contents

6.1 Introduction 140
6.2 Investigating by-item variance 141
6.2.1 Visual Sino-Korean full-word priming 143
6.2.2 Auditory Sino-Korean full-word priming 146
6.2.3 Effects of cohort with no semantic priming 148
6.2.4 Point of saturation in cohort size 151
6.3 Cross-script semantic cohorts in the Korean mental lexicon 156

6.1 Introduction

The evidence from our experiments so far suggests that the Korean mental lexicon encodes the semantics of Hanja characters on a conceptual level. We take the effects of cohort size observed in Experiment 4 to indicate that processing of Sino-Korean relies on mental representations of Hanja. The more robust findings from our cross-modal studies (Experiments 3 and 4 in Chapter 5) compared to those of our intra-modal experiments (Experiments 1 and 2 in Chapter 4) corroborate the idea that the semantics of Hanja are conceptually encoded, rather than confined to the visual script or the Hanja characters explicitly. However, our results suggest that although the contributions of Hanja may be conceptually based, these conceptual encodings still rely on some representation of the shared Hangul form, which suggests a 'softer' conceptualization of cross-script representations. The semantics

of Hanja contribute to the meanings of Sino-Korean words, the full semantic access to which is reliant on the Hangul representation that is shared between different Hanja characters that differ in meaning.

Broadly, our studies offer mixed results regarding how exactly Hanja is represented. We demonstrate that the semantics are represented in some way, and our findings from Chapter 5 provide evidence for a more conceptual-level encoding, which is modulated by the relative size of the semantic cohort, which is composed of other competing concepts. In Chapter 5, there was some suggestion of by-item variance worth further investigation. In light of some of the unknowns that remain, in this chapter, we investigate further the cause of such variation by reporting additional post hoc statistical tests in Section 6.2, followed by a theoretical discussion of the collective findings in Section 6.3.

We also propose a possible schema of the Korean mental lexicon in Section 6.3, factoring in the semantic contributions of Hanja characters, in light of the findings presented in the last two chapters. No exhaustive model of the Korean mental lexicon exists, let alone one which teases apart the Sino-Korean sub-lexicon from the pure Korean sub-lexicon. As discussed in Chapter 3, Yi's (2009) model of the Korean mental lexicon does not allow for mental representation of the many meanings of Hanja characters, the absence of which seems theoretically remiss given the high proportion of Sino-Korean words in the whole Korean lexicon. Based on the additional analyses presented here conducted on existing data from Experiments 1-4, we posit the structure of the Sino-Korean lexicon to be divided into two smaller sub-lexicons: one which comprises Sino-Korean morphemes for which the representation of Hanja is salient and accessible, and one which consists of Sino-Korean morphemes for which the representation of Hanja is less so.

6.2 Investigating by-item variance

In Experiments 2 (Chapter 4) and 4 (Chapter 5), a lexical decision task was carried out using mediated semantic priming. In both cases, we tested Sino-Korean full words, rather than individual fragments, and each word contained a specific

Sino-Korean morpheme that had more than one possible meaning. We predicted that upon hearing or seeing the particular Sino-Korean morpheme in question, all possible meanings would be activated. Of course, seeing the word elicits different types of processing compared to hearing it, which potentially modulates the degree to which the full cohort is engaged. We observed that the cohort played a more crucial role in processing during auditory processing of Sino-Korean, compared to visual processing. Recall that we used the following paradigm, with three experimental conditions:

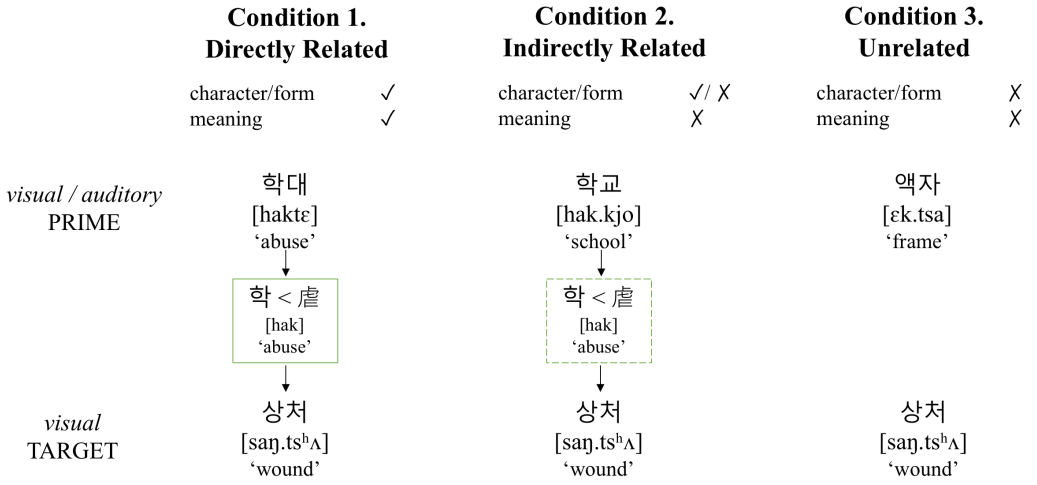


Figure 6.1: Depending on the condition, the prime-target pair involved a direct or indirect semantic relationship that required mediated semantic access via shared homograph (e.g., *hak*).

In both Experiments 2 and 4, we predicted priming in the Directly Related condition, as the prime and target shared a direct semantic relationship without any mediation. We also predicted that this pattern would hold consistently for all targets; that is, when a target is preceded by a directly related prime, the target should be activated, hence a priming effect. This is expected, and there is robust evidence to support this prediction (Posner & Snyder, 1975; Meyer & Schvaneveldt, 1971; McNamara, 2005; Neely, 1991).

However, these predictions were not borne out in Experiments 2 and 4, observing different results. Our results indicated large by-item variance, with only certain targets being facilitated by the prime in the Directly Related condition, and response to the others either unaffected or even inhibited. To understand the reason for

these results, here analyze the relevant subsets of the data separately. We take as our working assumption that Sino-Korean targets should be primed consistently in the Directly Related condition, while we would expect more varied results in the Indirectly Related condition, since the relationship is mediated and thus weaker in strength of activation. For both Experiments 2 and 4, we separated out the subset of target items for which there was active priming – an effect size of 5 ms or more, on average, across all participants (Friedrich, Eulitz, & Lahiri, 2006; Friedrich, Lahiri, & Eulitz, 2008) – in the Directly Related condition. Thus, we created two groups of targets: those which elicited an active priming effect size of 5 ms or above in the Directly Related condition ($n = 18$), and those which failed to elicit such an effect size in the Directly Related condition. We then re-ran our linear mixed effects model on this reduced dataset which only includes those targets for which there was active priming in the Directly Related condition. The syntax of our model remains the same, using the same fixed and random effects. We report the findings here for Experiment 2.

6.2.1 Visual Sino-Korean full-word priming

Recall from Chapter 4 that Experiment 2 involved visual presentation of Sino-Korean primes, followed by visual targets semantically associated with another meaning of a Hanja character that shared the same Hangul representation as one of the morphemes present in the prime (Figure 6.1). The idea is that this morpheme will activate all other possible meanings and thus mediate semantic access to the meaning present in the target. Also recall that for Experiment 2, our best-fitting model[1] included *Participant* and *Target* as random effects, and *Prime Type* as our fixed effect. Not including *Cohort* as a nested factor within *Prime Type* provided the best fit for our data. We find through model comparison that again the simpler model offers better fit with the pared down dataset, so our model structure remains the same. Table 6.1 summarizes the model output.

[1] lmer(RT ~ PRIME TYPE + (1 | PARTICIPANT) + (1 | TARGET)).

Table 6.1: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	492.476	9.291	39.900	53.008	$< .001$*
Directly related	-24.942	3.565	2973.700	-6.997	$< .001$*
Indirectly related	-7.766	3.608	2979	-2.153	.0314*

When we pare down the data to exclude target items for which there is no priming in the Directly Related condition, not only is our priming effect size more pronounced (previously $\sim$ 7 ms) in the Directly Related condition, but we also now observe a significant priming effect in the Indirectly Related condition. These results suggest that priming in the Indirectly Related condition is contingent upon the effects we see in the Directly Related condition.

In fact, when we run the same model on the remaining subset of targets (i.e., those which were not primed in the Directly Related condition), we see the results confirm active inhibition in the Directly Related condition. Targets which are preceded by semantically related primes elicit reaction times that are significantly slower (by $\sim$ 13 ms) than those which are preceded by an unrelated prime. In similar suit, we observe no priming in the Indirectly Related condition. Model comparison again confirms that the best fitting model includes only *Prime Type* as our fixed effect with no nested effect of *Cohort*, with *Participant* and *Target* as random intercepts. These results are shown in Table 6.2:

Table 6.2: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	471.422	8.845	40.100	53.295	$< .001$*
Directly related	12.768	3.781	2653.500	3.377	$< .001$*
Indirectly related	5.439	3.805	2655.300	1.429	.153

Having separated out the data, we are left with smaller sample sizes for individual

sub-sets, thereby reducing the experimental power, but the trends they present merit further discussion. Upon comparing the results from Experiments 2 and 4, where the modality in which the primes were presented was different, we had concluded that visual presentation of Sino-Korean possibly limits complete access to the full semantic network of the individual morphemes, and thus the full range of semantics of the Hanja characters.

The findings observed here may be due to holistic processing of visual forms. The disyllabic, bimorphemic word 학교 'school', for example, comprises two Sino-Korean homographs 학 and 교, which independently encode multiple meanings. We expect that upon hearing 학, all possible meanings, e.g., 'learn', 'harm', etc., are activated, and then upon hearing 교 in immediate succession, one realizes the meaning of the word and subsequently suppresses all other context-inappropriate meanings of 학 [hak], e.g., 'harm'. Visual processing may not be characterized by the same degree of incrementality that is involved in auditory processing, suggesting that not all meanings available to 학 [hak] are pre-activated, or if they are, the ones that contextually inappropriate are deactivated faster and sooner. Thus, activation may not spread to other meanings of 학 [hak], such as 'harm', which means that when a target that is semantically related to the 'harm' interpretation of 학 is presented, it may not necessarily be primed (학교 [hak.kjo] 'school' – 아픔 [a.pʰɨm] 'pain', as per the Indirectly Related condition).

The results presented here suggest that depending on the Sino-Korean morpheme, the link to this larger semantic network of all available meanings is more or less accessible, to the point that even meanings which bear direct relevance with no mediation necessary cannot be activated (학교 [hakkjo] 'school' – 선생 [sʌn.sɛŋ] 'teacher', as per the Directly Related condition). Based on the results, I posit that the presence of active priming in the Directly Related condition could act as a predictor for whether this "semantic link to all" is available. Only when there is priming (i.e, when 학교 [hak.kjo] 'school' primes 선생 [sʌn.sɛŋ] 'teacher') in the Directly Related condition is the link able to spread activation to the larger network of meanings (i.e., 학교 [hak.kjo] 'school' being able to prime 아픔 'pain' [a.pʰɨm]).

This is also consistent with our previous finding in Chapter 4 with the composite dataset that visual presentation of Hanja does not seem to be modulated by cohort size, given that even with the pared down dataset, the model fit did not improve with the inclusion of *Cohort*. If the larger set of available meanings is inaccessible due to weak spreading activation, then it naturally follows from this that variable cohort size would have limited effects on lexical access.

6.2.2 Auditory Sino-Korean full-word priming

We conducted the same analysis as before with the data from Experiment 4, which involved auditory, rather than visual, presentation of Sino-Korean full-word primes. We pare out all targets for which there is no active facilitation in the Directly Related condition, working only with the subset for which there was priming (n = 18). Recall that for this initial analysis in Chapter 5 through model comparison, it was established that the best fitting model included *Cohort* as a nested factor within *Prime Type*, reaching significance at $\chi^2(2) = 8.88$, $p = .012$.[2] Only in the Directly Related condition did the effect of *Prime Type* reach significance; while reaction times were slower in the Indirectly Related condition compared to those in the Unrelated condition, this was not statistically significant. However, reaction times were modulated by cohort size in the Indirectly Related condition, an effect that was not observed in the Directly Related condition.

These results crucially demonstrate that when targets were directly related to the prime, activation was less dependent on the size of the cohort, but when targets were related to the prime via a mediator, activation was bolstered by increasing cohort size. This finding is key and provides new insight not only into the role of semantic cohorts, but also into the ways in which semantically-mediated associations affect processing. Specifically, these results reveal that semantically-mediated relationships are particularly sensitive to cohort size. As further interpretation of these results, we posit that the semantic strength of the link between the directly related prime and the target is unaffected by increasing cohort size, by withstanding the competition

[2] lmer (RT ~ PRIME TYPE / COHORT + (1 | PARTICIPANT) + (1 | TARGET)).

of neighboring meanings also activated in the semantic network surrounding the Sino-Korean morpheme in question. However, in the mediated relationship, the strength of the semantic relationship may be weaker between the prime and the target, in which case activation benefits from more semantic links (i.e., more meanings activated in the network).

With these findings as our baseline, we now turn to see whether again paring down the dataset to only include those targets for which there was active priming in the Directly Related condition causes any changes in the findings from our initial analysis. The same cut-off point as was used in the previous subset analyses (5 ms) was used here. As before, we test two possible models with this new pared-down dataset, one which contains a term for *Cohort*, and one which does not. The simpler model reveals statistically significant priming effects in both the Directly and Indirectly Related conditions; significant priming in the Indirectly Related condition was not observed in the initial analysis presented in Chapter 5. The model output is given below:

Table 6.3: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	465.326	8.069	28.700	57.665	$< .001$*
Directly related	-27.087	2.835	3133.900	-9.555	$< .001$*
Indirectly related	-14.606	2.833	3133.500	-5.156	$< .001$*

As our initial model yielded a significant nested interaction with *Cohort*, we ran the same model with this reduced dataset to see whether the effect is observed again. The model output of the best fitting model containing *Cohort* is presented here:

Table 6.4: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	465.287	8.238	27.700	56.481	$< .001$*
Directly related	-31.937	4.077	3024.500	-7.834	$< .001$*
Indirectly related	-17.412	4.073	3027.300	-4.275	$< .001$*
Directly related × *Cohort*	0.592	0.358	2695.800	1.652	.0986
Indirectly related × *Cohort*	0.343	0.358	2688.900	0.959	.3377

As we observe in Table 6.4, there is no longer a significant effect of increasing cohort size, which we had originally observed in our initial analysis that included all targets exhaustively for the Indirectly Related condition. When we take the subset of targets that were primed in the Directly Related condition, we see that reaction times are no longer modulated by cohort size, in either experimental condition. This appears consistent with our earlier interpretation that the degree of automaticity with which the larger semantic network of meanings can be accessed is not the same for all Sino-Korean morphemes. Whether all meanings are accessed also affects the role the size of the cohort plays in Sino-Korean processing.

6.2.3 Effects of cohort with no semantic priming

Recall that when we analyzed the targets for which there had been no priming in the Directly Related condition for our visual intra-modal study, there was evidence of active inhibition. We found that in the initial analysis presented in Chapter 4, the best-fitting model did not include *Cohort*. This same model structure was also found to provide best model fit for the subset analyses of the intra-modal study. As in seen in Chapter 5, Experiment 4, which used a cross-modal paradigm, revealed that a model including *Cohort* offered the best model fit, showing that auditory processing of Sino-Korean is affected by the multiple semantics of Hanja characters, and that increasing semantic cohort size speeds up recognition. Such a model structure, however, did not offer the same model when only targets that

were actively primed in the Directly Related condition were included in the analysis. In this case, our best fitting model did not contain *Cohort*.

Our findings so far show that targets for which there was active priming in the Directly Related condition were unaffected by increasing semantic cohort size, meaning that a larger semantic cohort did not facilitate processing. This suggests some degree of weightedness in terms of priming effects; that is, the priming effect in the Directly Related condition, in both modalities, is able to withstand, or at least, outweigh the priming effects of Cohort. However, if no priming in the Directly Related condition is observed, then the spreading activation seems to flow to the larger semantic network of meanings that are available for that specific Sino-Korean morpheme, at which point relative size of the cohort becomes important. Given these findings, we predict that the pared out items that remain which represent those which did not actively prime in the Directly Related condition will rely on the cohort size during processing.

With a pared down dataset including only those targets for which there was no active facilitation in the Directly Related condition, we tested again two possible models, one which contains *Cohort*, and one which does not. Model comparison shows that the best fitting model contains *Cohort* as a nested factor within *Prime Type*, with *Participant* and *Target* as random intercepts, reaching significance at $\chi^2(2) = 20.62$, $p < .001$.[3] Table 6.5 summarizes the model output.

Table 6.5: Results of linear mixed effects analysis. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	435.099	8.557	18.700	50.846	$< .001$*
Directly related	13.606	4.645	2122.300	2.929	.003*
Indirectly related	30.728	4.697	2116.600	6.542	$< .001$*
Directly related × *Cohort*	-0.358	0.435	1701.400	-0.823	.411
Indirectly related × *Cohort*	-1.8656	0.435	1707.900	-4.290	$< .001$*

[3] lmer (RT ~ PRIME TYPE / COHORT + (1 | PARTICIPANT) + (1 | TARGET)).

The results in Table 6.5 indicate that reaction times in both the Directly and the Indirectly Related conditions are significantly slower relative to the Unrelated condition, giving clear evidence for inhibition. Furthermore, there is also a significant effect of cohort size, which is consistent with our initial analysis performed on all targets. Crucially, though, these results are contrary to our findings from our previous analysis on the subset of targets for which there was active facilitation in the Directly Related condition.

When the analysis was performed on only these targets, any previously observed effects of cohort size were no longer significant; however, when the remaining subset of targets is taken, for which there was no facilitation, and in fact, rather, active inhibition, the effects of cohort size were significant. These findings demonstrate that spreading activation to different available meanings is only possible when more local activation (i.e., meanings which are directly related) is insufficient. Rather than spreading activation, this seems to be some sort of 'compensatory' activation, whereby the larger semantic network of available meanings is recruited to compensate for the lack of more local activation (i.e., the Directly Related condition). These findings again underscore the possibly variable status of Sino-Korean in the mental lexicon, with certain morphemes more tightly connected to the semantics of Hanja, and others less so. Despite the variable status, the representation of the semantics of Hanja remains clear, as the relative size of the semantic cohort modulates Sino-Korean processing.

Furthermore, the variability in the strength of specific semantic links also suggests that certain meanings are privileged over others, but that the degree to which they are privilege depends on the number of other meanings with which it is competing. That is, whether meanings are embedded within a larger or a smaller cohort influences to what extent those meanings are then privileged. Increased frequencies of or exposure to certain Hanja characters might be one possible reason for semantic privileging.

6.2.4 Point of saturation in cohort size

The results presented so far seem to suggest a facilitative effect of larger semantic cohort sizes in auditory word recognition. Modelling cohort size as a continuous variable, we found in Experiment 4 that during auditory processing, reaction times sped up with increasing cohort size. This invites the question of whether such effects reach a point of saturation; that is, is there a certain point after which increasing cohort size has no further effect? Perhaps effects begin to peter out after a certain point of saturation. To answer this question, we took different subsets based on different cut-off points in terms of the number of meanings (i.e., more than x number of meanings). We take x to be an arbitrary measure, rather than an exact measure, correlating to a cut-off point. These findings will then help to elucidate whether Sino-Korean morphemes benefit from more or less densely concentrated cohort sizes.

6.2.4.1 Semantic cohort size of ten meanings or more

To begin with, we took the subset of all reaction times for targets preceded by primes containing a Sino-Korean morpheme with ten or more possible meanings. This subset was taken for reaction times from Experiments 2, 3, and 4 (i.e., the two full-word priming studies as well as the fragment priming study)[4]. The same model structure[5]. To note, this is not the best fitting model that provides best model fit for all three experiments; recall that only for Experiment 4, did this model structure offer best model fit. For purposes of this exploratory analysis which looks specifically at *Cohort*, we use this model for all three datasets was used for all of the analyses. In this particular case, our results revealed that, from a starting baseline of ten possible meanings, reaction times were not significantly affected by increasing cohort size, suggesting that by ten, the effects of semantic cohort size already begin to diminish.

6.2.4.2 Semantic cohort size of ten meanings or fewer

Given the previous findings, we took the complementary subset for all three experiments, looking only at reaction times for targets preceded by primes containing

[4]Experiment 1 did not measure semantic cohort size.

[5]lmer (RT ~ PRIME TYPE / COHORT + (1 | PARTICIPANT) + (1 | TARGET)).

a Sino-Korean morpheme with ten meanings or fewer. For this, we found that in the cross-modal full-word (Experiment 4) priming study, where Sino-Korean words were presented auditorily, for primes that had anywhere from two to ten possible meanings, there were no significant effects of increasing semantic cohort size.

The lack of any significant trends suggests that semantic saturation is reached at a point that is fewer than ten meanings. This was the case for both experimental conditions (Directly and Indirectly Related). In the intra-modal (Experiment 2) study, where Sino-Korean words were presented visually, there were significant effects of increasing semantic cohort size in the Directly Related condition, but not in the Indirectly Related condition. Similarly, there was also a significant effect of cohort size in the cross-modal fragment (Experiment 3) priming study, where Sino-Korean morphemes were presented in isolation auditorily. In both the intra-modal and the fragment studies, semantic processing appears to be affected by increasing semantic cohort size until the cohort contains ten meanings, at which point, there is no longer a significant effect.

Table 6.6: Results of linear mixed effects analysis for the intra-modal priming experiment (Experiment 2), containing only primes with a semantic cohort size of ten meanings or fewer. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	483.200	7.389	113	65.392	$< .001$*
Directly related	-18.623	6.366	3878	-2.925	.003*
Indirectly related	-12.011	6.357	3907	-1.889	$< .059$
Directly related $\times$ *Cohort*	2.207	1.116	3501	1.978	.048*
Indirectly related $\times$ *Cohort*	1.616	1.121	3544	1.442	.149

Table 6.7: Results of linear mixed effects analysis for the cross-modal fragment priming experiment (Experiment 3), containing only primes with a semantic cohort size of ten meanings or fewer. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	453.127	5.686	188	79.688	$< .001$*
Related	5.410	3.749	6421	1.443	.149
Related × *Cohort*	-2.641	0.683	6119	-3.865	$< .001$*

The intra-modal study revealed reaction times to be significantly slower in the Directly Related condition with every additional member in the cohort (Table 6.6), while reaction times sped up significantly in the fragment study with increasing cohort size (Table 6.7)

6.2.4.3 Semantic cohort size of five meanings or fewer

Following from the previous findings, the same analysis was performed again on a smaller semantic cohort size. We took all reaction times to targets that were preceded by primes containing a Sino-Korean morpheme with five possible meanings or fewer. We found significant effects of semantic cohort size across all three experiments. The findings from the intra-modal (Experiment 2) priming study were consistent with those in Section 6.2.4.2 where increasing semantic cohort size seemed to yield significantly slower reaction times. The same pattern can be observed in Table 6.8, where reaction times in the Directly Related condition are significantly slower ($\sim$ 12 ms) as cohort size increases by one, which is interestingly a larger effect size than what was observed when the cohort size was ten meanings and fewer (Table 6.9).

Table 6.8: Results of linear mixed effects analysis for the intra-modal priming experiment (Experiment 2), containing only primes with a semantic cohort size of ten meanings or fewer. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	483.195	7.550	109.300	64.001	$< .001$*
Directly related	-49.140	14.254	3082.100	-3.447	$< .001$*
Indirectly related	-5.744	14.027	3033	-0.409	.682
Directly related × *Cohort*	12.874	4.035	3058.400	3.191	.001*
Indirectly related × *Cohort*	0.356	3.983	3019.400	0.089	.929

The findings from the cross-modal fragment (Experiment 3) study also yielded significant effects of cohort size. Response times were significantly faster for targets which were preceded by Sino-Korean morphemes with larger cohort sizes, also consistent with the findings in Section 6.2.4.2.

Similar to the intra-modal priming study, the effect size for the cross-modal fragment study was also larger when semantic cohort size was limited to just between two and five meanings, compared to when there could be as many as ten (Section 6.2.4.1).

Table 6.9: Results of linear mixed effects analysis for the cross-modal fragment priming experiment (Experiment 3), containing only primes with a semantic cohort size of five meanings or fewer. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	453.274	5.901	179	76.817	$< .001$*
Related	21.605	6.879	5496	3.140	.002*
Related × *Cohort*	-7.107	1.966	5473	-3.614	$< .001$*

The most interesting results emerge in the analysis for the cross-modal full-word (Experiment 4) priming study, which reveal a significant effect of semantic cohort size, with reaction times becoming significantly slower in both the Directly and Indirectly Related conditions with every additional meaning. This is shown in Table 6.10.

Table 6.10: Results of linear mixed effects analysis for the cross-modal full-word priming experiment (Experiment 4), containing only primes with a semantic cohort size of five meanings or fewer. The asterisk indicates significant results at the 95% significance level.

	β	SE	***df***	***t***	***p***
Intercept	452.419	6.303	63	71.774	$< .001$*
Directly related	-22.714	8.246	3224	-2.754	.006*
Indirectly related	-23.281	8.296	3236	-2.806	.005*
Directly related × *Cohort*	5.516	2.464	3140	2.239	.025*
Indirectly related × *Cohort*	9.767	2.506	3141	3.898	$< .001$*

The results in Table 6.10 indicate that the direction of the effect is different for cross-modal priming of full words in Sino-Korean; it appears that response times become significantly slower in both experimental conditions, as the size of the semantic cohort increases. This may be an indication of different saturation points depending on modality (auditory versus visual) and whether the Sino-Korean morpheme is presented in isolation as a fragment or lexicalized in a full word. Differences in holistic or incremental processing may also affect the point at which semantic saturation is reached.

As a final measure, we also performed the same analysis on reaction times for targets preceded by primes containing a Sino-Korean morpheme with three or fewer meanings. This meant that the Sino-Korean morphemes in this particular subset only had either two or three meanings. The results revealed no significant effect of semantic cohort size on lexical access, and this was found to be the case in all three experiments. These findings taken with our previous findings suggest that the point of semantic saturation may be within the range of three to five meanings, after which point, every additional meaning has an increasingly small effect on semantic access. With these findings in mind, we try to conceptualize a theoretical structure of the Korean mental lexicon which takes in account the semantic representations of Hanja characters and their active role in Sino-Korean processing, reflecting the effects of cross-script cohorts.

6.3 Cross-script semantic cohorts in the Korean mental lexicon

The empirical findings presented in Chapters 4 and 5, as well as the analyses provided in this chapter, offer important considerations for the structure of the Korean mental lexicon. The results shown in this thesis offer evidence for the representation of Hanja characters in the Korean mental lexicon, as observed by the effects of semantic cohort size on the Sino-Korean morphemes. The fact that increasing cohort size significantly modulates reaction times in auditory processing of Sino-Korean suggests multiple semantic access, which is consistent with Verdonschot et al.'s (2013) findings of multiple phonological access of both the ON and KUN readings in Sino-Japanese in Section 3.5.2. The findings presented here show multiple semantic access of Hanja characters in Sino-Korean, which, taken with the findings in Sino-Japanese, suggest that Sino-contributions, however defined, create a rich network of links through which activation can spread to both phonological and semantic nodes. The contributions of Chinese are far-reaching not only in Korean, but also in Japanese, and most likely other languages that have been subject to similar influence.

As introduced in Chapter 3, Joyce (2002) posits a model of the Japanese mental lexicon which assumes a connectionist approach, as done by Taft et al. (1999) in their conceptualization of the Chinese mental lexicon, by positing an abstract layer of lemma units which mediates the connections between both the phonological and the orthographic representations and the semantic representations. In doing so, Joyce (2002) is able to account for all possible script representations, providing a unified account of both native Japanese and Sino-Japanese. This model fits the parameters of the Japanese writing system well, particularly given the possibility for different phonological readings of kanji.

The model, however, is not neat enough for the added parameters of the Korean writing system, which not only uses Hanja characters far less frequently than Japanese, but does not have multiple readings in the same way Japanese does. In fact, the influence of Chinese by way of the Hanja script on Korean is far less overt compared to its influence in Japanese. There is no explicit notion of multiple

activation beyond the immediate context within Joyce's model; his model may be able to account for some degree of cross-script interaction, given the separation created by the lemma level, but it still remains unclear how meaning that is expressed across the different scripts is accessed. Specifically, it is not entirely clear how Joyce's (2002) model would cope with homographs and homophones, each encoding distinct meanings.

Returning to Taft et al.'s (1999) model, their revised model offers the closest fit for addressing the complexities of Korean, given that their model deals explicitly with homographic morphemes, which is resolved at the lemma-level of representation. An orthographic unit representing a character sends activation to a lemma unit that represents an abstract form which has a phonological as well as a semantic representation. This means that activation can spread to lemma units that are semantically-related to the orthographic unit. While this appears to be a fitting approach for the model of the Korean mental lexicon, it is not clear where, within the model, one could posit a level for cross-script representations, especially in Sino-Korean, where the semantics of Chinese is encoded across different scripts. This would suggest that a hybrid between Joyce's (2002) and Taft et al.'s (1999) would perhaps provide the best approach, which the approach that is somewhat taken by Yi (2009). Neither of the models presented here previously is able to account for cohorts that exist across two different scripts, due to the semantics that are encoded across the scripts.

While Yi's (2009) model posits two separate levels of representations for the Hangul and the Hanja, which enables the model to capture cross-script relations to some extent, most crucially, his model is unable to account for the cross-script priming effects observed in the present data. Our findings from Chapter 5 show that processing of Sino-Korean words not only actively relies on the semantics of Hanja characters, but also involves taps into a larger cohort of possible meanings, as observed by significantly faster reaction times for larger semantic cohorts. Recall that while Yi's model acknowledges that Sino-Korean morphemes can map onto

the semantics of different Hanja characters, he does not address this directly in his model; in fact, he does not account for the possibility of multiple semantic access.

Yi's (2009) model predicts that the semantics of Hanja characters are accessed individually, rather than potentially simultaneously, activating a cohort of meanings. More importantly, recall that there was no overlap form between the primes and targets in any of the Experiments presented in Chapter 4 and 5. Yi et al.'s experiments (Jung & Yi, 2004; Yi, 1999; Yi & Yi, 1999) all involved a shared morpheme between the prime and target, which means that his model rests on the actual morpheme itself, rather than the concepts it might encode. While his model takes into some account the contributions of the Hanja characters, they are confined to the script itself, which as the findings in Experiment 4 show, cannot be the case. And by positing an exclusively form-based morphological representation of Hanja characters, he is, in doing so, also neglecting to take into account the interconnectedness of their semantics that underpins the semantic cohort, leading to a non-exhaustive account of their contributions.

To clarify, this is not to say the contributions of Hanja characters cannot be represented at the morphological level within the mental lexicon; rather, in any theorization of how best to represent More crucially, this means that his model cannot account for the priming effects that are observed in the present findings, which show priming between semantic associates.

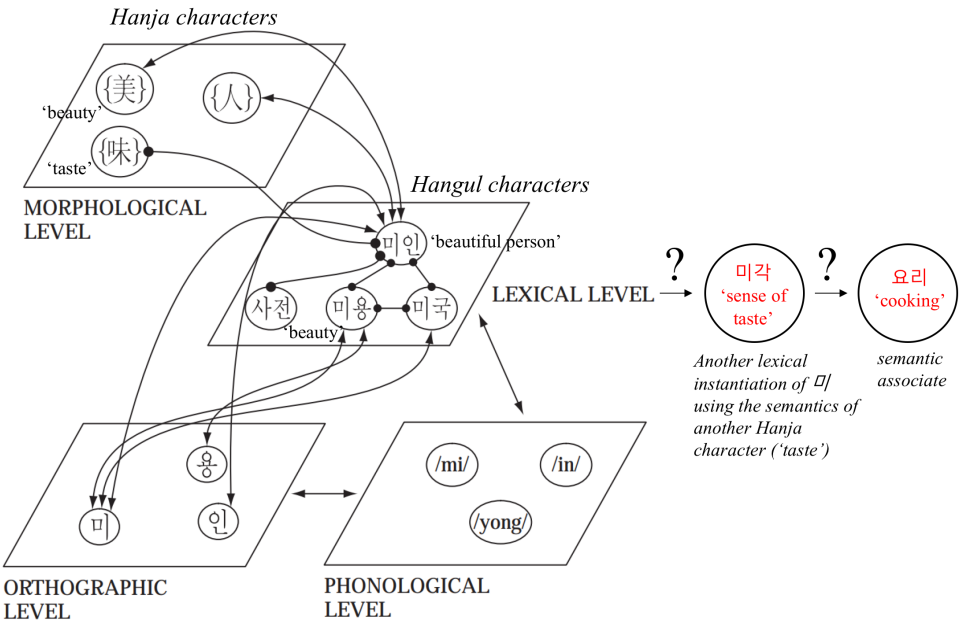


Figure 6.2: Yi's (2009: 407) model does not make clear how items which do not overlap in form are activated, especially as his model specifically posits a morphemic level, and thus it remains unclear how semantically related lexical items would be accounted for, e.g., *cooking*, in his model.

Figure 6.2 shows 미 [mi] linked to the semantics of two different Hanja characters, but whether their semantic representations are activated is not clear, and the only way in which to see that without possible confounds of morphological and orthographic effects would be through semantic associates, but at it stands, the model does not tease apart semantic effects from morpho-orthographic ones. If the links between Yi's (2009) model are form-dependent, then the important aspect of the semantic contributions of Hanja characters is lost. While Yi assumes some conception of links between scripts, such links do not take into account any of the semantic information that is encoded between them. Thus to take into account the cross-script semantics that are activated altogether as a cohort, rather than as individual morphemes, one by one, a revised model of Yi (2009) is presented in Figure 6.3 below.

In the following refinement of Yi (2009), cross-script semantic cohorts are taken into account and represented at the newly developed CONCEPTUAL level of the Korean mental lexicon.

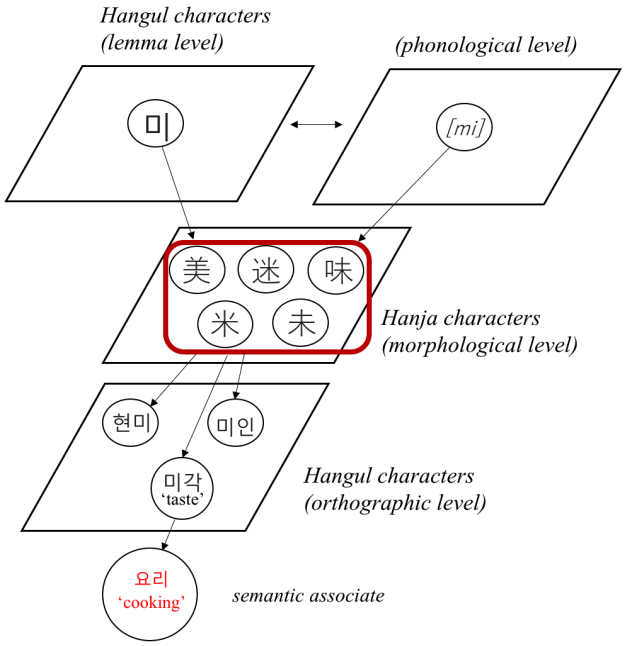


Figure 6.3: A revised model of the Korean mental lexicon for the representation and processing of Sino-Korean words, as proposed by Yi (2009).

Sino-Korean words, by definition, are of Chinese origin, which means that they are derived semantically from Classical Chinese characters. Given that we have shown in this thesis that Sino-Korean processing involves semantic processing of Hanja characters, a model of the Korean mental lexicon should represent in some ways this semantic relationship across scripts, and that more than one meaning can be accessed consistent with earlier literature of multiple access (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979; Onifer & Swinney, 1981). By positing the semantic contributions of Hanja characters as concepts grouped in semantic cohort that is mediated by the shared representation in Hangul, the revised model is able to give some account of how semantic information is activated and accessed in a bi-scriptal lexicon.

The research presented in this thesis indicates that such interactions rest on a foundation of rich semantic connections, networks and cohorts. The findings suggest that the semantic contributions of Hanja characters are encoded at a higher conceptual level, rather than on a morphological level. These findings also highlight the dynamic nature which characterizes the mental representations of Hanja

characters, based on the observed cross-activation of meanings across different scripts. The dynamic nature of conceptual representations has been corroborated by other literature in the field; for example, priming has been found between words which are perceptually related by way of shape (e.g. *ball* and *melon*; Schreuder, Flores D'Arcais, & Glazenborg, 1984; Taylor, 2005), color (Yee, Ahmed, & Thompson-Schill, 2012), or other such properties (e.g., both *piano* and *typewriter* involve use of the hands and thus have been found to prime one another (Myung, Blumstein, & Sedivy, 2006)).

To clarify, theorizing the contributions of Hanja characters as conceptual is not to suggest that an awareness of Hanja characters is meta-linguistic, but rather to differentiate such contributions as purely script-based. Positing a conceptual representation precludes the implication that the status of Hanja characters within the mental lexicon relies solely on literacy, or minimally, a working understanding of the characters themselves. This is particularly important, as ultimately, the architecture with which any model of the Korean mental lexicon is built should work for literates and illiterates alike. However, the exact extent to which the division of labor to organize the morphology of Sino-Korean is split between a morphological level and a conceptual level is open to further study, but given the effects of cohort, as seen in Chapter 5, it seems clear that morphological organization is in some way enriched by the contributions of Hanja characters. Whether Hanja characters alone trigger morphological organization, or serve as a catalyst for existing processes governing morphological structure also remains an open question, but the results presented here suggest a more catalytic role.

Furthermore, Thierry and Wu (2007) found in their study on Chinese-English bilinguals that participants were tapping into implicit orthographic relationships.[6] In their study, although English words were used to devise prime-target pairs that were either semantically related or not, unknown to the participants, the prime-target pairs, when translated into Chinese, actually shared a character. Their results found participants to be significantly affected by such implicit effects, which is consistent with the findings of semantic activation via implict cross-script relations.

[6]Reference provided by N. Kazanina (personal communication, 7 September 2019).

That is, for both Thierry and Wu (2007) and the results presented in this thesis, the implicit contributions of Chinese characters are encoded in mental representations. Building on such work, the evidence presented in this thesis may offer support for perceptual similarity in scripts, but as motivated by the shared semantics. That is, two scripts can be perceived to be 'similar' on semantic grounds, the implications of which may prioritize higher the study of writing systems within the domain of cognition. Writing systems are more than just a visual system, but operate as an important bridge not only between written and spoken language, but also form and meaning.

Recognition of Sino-Korean words involved the active processing of the semantics of Hanja characters, even of the meanings that were not directly relevant. This proved to be modulated by the relative number of possible meanings available for a given Sino-Korean morpheme. Most crucially, Yi's (2009) model fails to capture the multiplicity in meaning at the morphemic level. At the moment, it seems that different Hanja characters and their semantics operate in a vacuum, independent of one another, despite the clear observation that they share the same representation in Korean. Yi's (2009) model captures as much, but neglects to account for how different meanings are being accessed. Thus, we propose a model of the Korean mental lexicon in which the structure of the Sino-Korean lexicon is based on semantic cohorts.

That is, recall the findings from Sections 6.2.1 and 6.2.2 which seemed to indicate that semantic access to the larger cohort of meanings was largely dependent on the specific Sino-Korean morpheme. This was based on the finding that the presence of active facilitation in the Directly Related condition for a given item predicted the presence of active priming for that item in the Indirectly Related condition. These results were taken as evidence to show that multiple semantic access via Hangul is not uniformly available across as Sino-Korean morphemes, suggesting their variable status in the mental lexicon. Thus, we divide the Sino-Korean lexicon into two broad classifications:

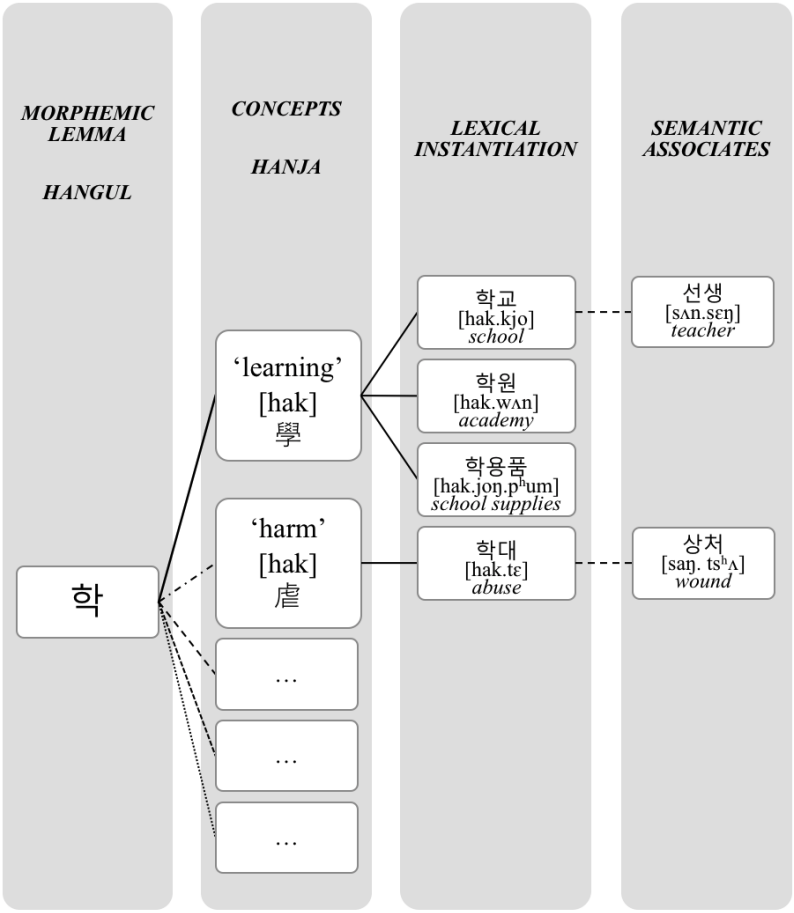


Figure 6.4: Proposed structure of the Korean mental lexicon based on the varying role of the semantic cohort in the Sino-Korean sub-lexicon.

Whether there is immediate access to closely related neighbors for Sino-Korean words will determine whether the constituent Sino-Korean morpheme has access to the larger semantic cohort of meanings. If there is activation of a closely related meaning for the word, then access to other meanings at the morphemic lemma level is blocked; however, in the absence of activation at the lexical level, there is instead cross-script activation at the morphemic lemma level, granting semantic access to other meanings.

This proposed model is akin structurally to the *swinging lexical network* proposal (Abdel Rahman & Melinger, 2007; 2009; 2019), which assumes that conceptual and lexical processing take place in parallel and may have effects which are opposite to those of conceptual priming and lexical competition. The proposal also assumes

that the degree of lexical competition is determined by the co-activation of an inter-related lexical cohort, and is thus sensitive to the number and strength of active competitors. This means that both the number of competitors and the strength of their activation contribute to the magnitude of the competition effect. Accordingly, both the number and strength of active candidates will depend on the spreading activation at the conceptual level.

Representations which are mutually inter-related will strengthen each other's activation level (e.g., *cat* activates *dog*, but also activates *cow*, which in turn also bolsters activation of *dog*.) This resonance within the semantic network and between the semantic and lexical levels is referred to as "swinging" (Abdel Rahman & Melinger, 2009). Equally, representations that are not mutually related will result in a broad distribution of activation, weakly activating many other representations that may not be necessarily directly related to the initial target concept (e.g., *whisker* activates *cat*, which also activates *bristles*, which is unrelated to *cat*). When the network is swinging, a cohort of lexical alternatives becomes strongly activated, which may in turn outweigh conceptual facilitation.

While Abdel Rahman and Melinger (2009; 2019) proposed this model within the context of language production, the underlying principles of an inter-related cohort offer relevant parallels for Sino-Korean processing; the semantic cohort of available meanings for a given Sino-Korean morpheme co-activates a larger cohort of semantic associates. This is visualized in Figure 6.5.

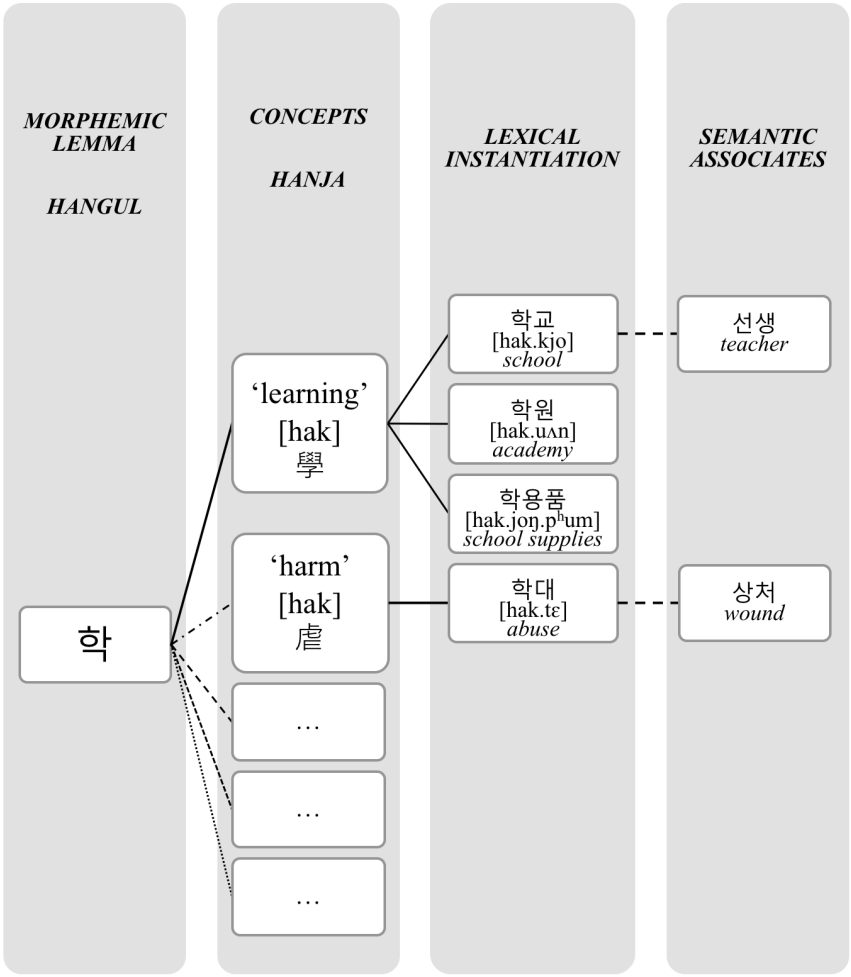


Figure 6.5: The "swinging" within the semantic network between the morphemic lemma level and the conceptual level causes resonance at the conceptual level, the effects of which reverberate into the lexicon and are observed in priming effects between semantic associates.

The priming observed across scripts, from one meaning of a Hanja character to another, rooted in Sino-Korean lexicalizations, can be accounted for by a model that is structured as in Figure 6.5. Each lemma at the morphemic level has conceptual nodes that vary in semantic strength, as denoted by the different dotted lines. The idea is that through resonance at the conceptual level of competing meanings, Sino-Korean morphemes, at the lemma level, are able to activate many other lexical representations that do not have direct semantic relationships with the original target concept (i.e., semantic associates). This creates reverberations in the larger lexical semantic network, whereby we observe the kinds of priming effects seen in the experiments of this thesis, between items that are categorically unrelated

in meaning, but additionally, in the case of Sino-Korean, related through the semantic contributions of scripts.

7 Conclusion

Contents

7.1 Summary of the thesis 168
7.2 Methodological considerations 170
7.3 Cohort activation and asymmetric representations . . 171
7.4 Expanding the revised model of the Korean mental lexicon 176

The work presented in this thesis was conducted to investigate the status of Hanja characters in the Korean mental lexicon. How exactly the various semantic contributions of different Hanja characters were represented, and whether semantic meaning could be activated across two structurally different scripts were the two main areas of interest. This thesis also explored the notion of semantic cohorts based on cross-script relationships where a writing system of a language can encode the semantic contributions of one script into another.

I begin in Section 7.1 by summarizing the key findings of this thesis, and then continue in Section 7.2 to offer some points of consideration regarding these findings from a methodological standpoint. Section 7.3 offers a potential theoretical next step in this research, drawing in aspects of underspecification and the Featurally Underspecified Lexicon (Lahiri & Reetz, 2002). Finally, in Section 7.4, I conclude with suggestion of three directions through which the revised model of the Korean mental lexicon introduced in Chapter 6 (6.3) can be expanded.

7.1 Summary of the thesis

The present research has explored the status of Hanja characters in the Korean mental lexicon, looking at how different models of lexical processing may account for the semantic processing of Hanja characters in both the auditory and visual domain. In doing so, we investigated whether processing of Sino-Korean words involved active processing of the semantics of Hanja characters. This research was in part motivated by a broader interest in the Korean mental lexicon, a satisfactory account of which has yet to be offered. Korean offers a unique window into a bi-scriptal lexicon whereby semantic information is encoded across two different scripts.

In understanding how the semantics of Hanja characters are processed, we not only gain new insights into how a bi-scriptal lexicon copes with semantic processing, but also the discourse surrounding bi-scriptal systems extends beyond the context of bilingualism. The literature presently focuses on bilingual speakers of languages, each of which employs a different script, but studying Korean is an important step towards research on monolingual speakers of different languages comprising two different scripts. To investigate, four behavioral studies were conducted using a lexical decision task with semantic priming, to probe whether Korean speakers are tapping into the semantic contributions of different Hanja characters in Sino-Korean lexical processing. The main findings are as follows:

- The semantic contributions of Hanja characters are represented in the Korean mental lexicon, not as the individual characters themselves, but as the semantic concepts they encode.

- Consistent with earlier findings (Swinney, 1979; Onifer & Swinney, 1981), which show simultaneous access to multiple meanings, the processing of Sino-Korean words, particularly in auditory processing, involves active processing of other available meanings.

- The findings from Experiments 3 and 4, both of which presented primes auditorily, revealed that the size of the semantic cohort (i.e., the number

of available meanings) played a significant role in Sino-Korean processing. Crucially, the results indicated that in auditory Sino-Korean word recognition, processing is significantly faster for those with larger semantic cohorts. That is, as the number of meanings in the cohort size increases, incrementally one-by-one, response times become faster and faster.

- Of the number of models of lexical processing discussed in Chapter 2, the Cohort Model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) offers the best theoretical fit for the present findings. The Cohort Model offers an account of incremental and sequential processing, the nature of which is well-suited to describe the processing of Sino-Korean, particularly in auditory word recognition; as each constituent Sino-Korean morpheme is heard, a cohort of possible meanings is activated, but when the word is presented holistically in the visual domain, access to this cohort is suppressed, or at least, in part weakly activated.

- There was by-item variance which could be attributed to whether there had been active priming in the Directly Related condition. Further analysis showed that in both modalities, semantic access to other possible meanings was dependent on whether primes in the Directly Related condition facilitated activation of the Sino-Korean targets. When targets were split according to priming effects in the Directly Related condition, results showed that semantic cohort size was no longer a significant factor.

- While the semantic contributions of Hanja characters are represented in the mental lexicon, individual access to these contributions were more or less viable for different Sino-Korean morphemes, suggesting the variable status of Sino-Korean. Depending on the particular Sino-Korean morpheme, there was more or less active involvement of the larger semantic cohort; for some morphemes, semantic processing involved the active processing of the other available meanings in the cohort, while for others, semantic processing only engaged the meaning immediately supported by the context and did not

recruit in processing the other meanings available in the cohort. The varying extent to which the size of the cohort affected reaction times suggests there might be some degree of privileging certain meanings over others, thereby making access to those meanings more automatic.

- I proposed a revised model for the Sino-Korean mental lexicon that not only takes into account the processing role of Hanja characters, but also offers an account of the variable status of different Sino-Korean morphemes in the mental lexicon, as observed in the exploratory analyses in Chapter 6. The structure of the Sino-Korean mental lexicon is divided along the lines of whether access to the larger semantic cohort is possible. The existing model of the Korean mental lexicon, as presented by Yi (2009) proved unable to account for cross-script priming effects observed in this thesis, hence the need for the current model to be refined.

7.2 Methodological considerations

In this thesis, semantic cohort size was modelled as a continuous variable, but it may be useful to also model cohort size as a discrete variable, by creating uniform sets of tokens across different cohort sizes during stimuli design. This may offer additional insight into how precisely varying cohort sizes affects processing. Furthermore, this approach may help to elucidate more clearly when a saturation point is reached during semantic processing. The analyses in Chapter 6 made use of coarse-grain measurements, but a more controlled design may allow for finer-grained analyses, resulting in a more definitive way of delimiting a point at which adding any additional members to the cohort has limited effects.

Similarly, while the analyses presented in Chapter 6 did not yield a significant interaction between semantic salience and semantic cohort size, perhaps a study controlling more carefully for these factors may offer different results. Again, a more uniform distribution of tokens across different cross-sections may yield important insights into how, within the semantic cohort, different meanings of varying salience

may compete. Another consideration from a methodological standpoint is controlling for syllable position; in Experiments 2 and 4, which used full words, rather than morphemic fragments, as primes, the position in which the Sino-Korean morpheme that was being tested appeared was not always the same across the two experimental conditions (Directly Related and Indirectly Related).

While this was controlled for to the greatest extent possible during stimuli design, it is possible that some of the observed effects may be attributable to differences in syllable position, particularly given that auditory word recognition involves incremental processing, and the order in which one hears the constituent morphemes may have an effect. This may be less of an issue in visual word recognition, where processing is potentially more holistic, rather than necessarily confined to a linearity in which the stimuli is presented. Additional experiments with stimuli more tightly controlled for syllable position may yield further insights on the processing of Sino-Korean, and whether access to semantic cohorts is more or less robust if the morpheme appears syllable-initial.

7.3 Cohort activation and asymmetric representations

From a theoretical standpoint, one possible next step in this research would be to pursue the idea of semantic underspecification and how this might interact with semantic cohorts. The concept of underspecification has been in use in theoretical linguistics for many decades (Kiparsky, 1982), and has in recent years entered the domain of psycholinguistic research (Lahiri, 1999; 2000; Lahiri & Marslen-Wilson, 1991; 2002; Marslen-Wilson, 1993; Reetz, 1998; 1999). Within phonology, an underspecified representation comprises a proper subset of the phonological features of a segment and their values.

Phonological features or feature values that are predictable are omitted in the underlying form, with features specified depending on both universal and language-specific properties. In light of the findings presented here, and within the theoretical parameters of phonological activation, one could investigate how the size and composition of a cohort changes in the context of asymmetric phonological

representations. For example, one could make the predication that the size of the cohort increases if asymmetric representations are assumed. That is, a sound will activate all possible counterparts, as per Marslen Wilson's (1987) conceptualization of cohort; however, if, within pairs of sounds, one is underspecified, then the size of the cohort will likely change.

To elaborate further, the idea of underspecification may help to explain some of the exploratory findings reported in this thesis, particularly those presented in Chapter 6. While this thesis has shown that the semantic contributions of Hanja characters and their phonological values appear to be represented, how *much* semantic information is in fact represented, and to what extent these representations are specified or even underspecified remains an open question. The analyses in Chapter 6 seem to suggest that broader semantic access to other meanings available in the cohort is modulated by whether there was active priming in the Directly Related condition. This indicates that the degree of semantic information which is represented is not the same for all Sino-Korean morphemes. While the collective findings of this thesis suggest that the various semantics of different Hanja characters are mentally represented in some way, certain specific Sino-Korean morphemes are potentially underspecified for this information, hence the variable status observed in the exploratory analyses in Chapter 6.

A useful model to consider in understanding the results presented in Chapter 6, and more broadly in this thesis, is a model that assumes that not all phonological information is fully specified. This model is known as the Featurally Underspecified Lexicon (FUL) model (Eulitz & Lahiri, 2004; Lahiri & Marslen-Wilson, 1991; 1992; Lahiri & Reetz, 2002; 2010). One of the main questions that FUL addresses is how the perceptual system of a listener is able to resolve the high level of variability and variation in the speech signal. One possible way in which this may be achieved is by way of asymmetric sensitivity to contrast, that is, allowing for greater leeway for the perceptual system to correctly detect, for example, a phoneme such as [d] after an [n], rather than an [n] after a [d] (Cornell, Lahiri, & Eulitz, 2013). The idea is that the overall detection rate across all possible contrasts is higher when

sensitivity to contrasts is kept asymmetric, as compared with a situation where sensitivity to contrast is kept symmetric for all relevant contrasts.

An important assumption underpinning the notion of an asymmetric contrast sensitivity is that not all features that are extractable from the speech signal are also represented in the mental lexicon. In fact, unlike other models which assume that all variants a listener is exposed to are mentally stored (Bybee, 2001; Johnson, 1997; Ranbom & Connine, 2007), FUL supports a more abstractionist approach where not all variants are stored in the mental lexicon. FUL stands in contrast to models that posit fully specified entries, as found in many contemporary connectionist models of spoken word recognition since the TRACE model (McClelland & Elman, 1986). Experimental findings which show an asymmetric sensitivity for only certain contrasts support underspecified representations.

The FUL model, specifically, assumes the underspecification of the place of articulation feature [CORONAL] (i.e., sounds which are articulated with the tip and/or blade of the tongue) as well as of the manner of articulation feature [PLOSIVE] (i.e., the release of air flow after a complete closure made in the vocal tract). Only place representations that do not assimilate are fully specified. For example, in a word such as *rain*, the nasal consonant is underspecified for its place of articulation feature [CORONAL], that is, it is not stored in the mental representation. This is important, as it means that all instantiations of *rain* in different acoustic manifestations, such as [rain], [raim], or [raiŋ] in *raindrop*, *rainbow*, and *raincoat*, respectively, are tolerated as possible variants of the underlying partially specified /n/ in the mental representation of the lexical entry for *rain*.

The mapping of the acoustic signal to specified and unspecified entries results in asymmetric activation patterns based on the extracted features and whether the representation of a particular word has those specific features (Lahiri & Reetz, 2002). A match refers to when the signal and representation contain the same representation. A no-mismatch occurs if a feature that has been extracted from the signal is not represented in the lexicon, which means that elements with [LABIAL] or [DORSAL] place of articulation that are present in the signal do not mismatch with

the underspecified, or empty place slots, of coronals. On the other hand, a mismatch is detected if the signal and the representation contain incompatible features.

Elements with [CORONAL] place of articulation mismatch with representations that have [LABIAL] or [DORSAL] specified. The FUL model predicts that listeners will reject from the cohort, lexical candidates that mismatch, while matching or no-mismatching words remain activated. Thus, the notion of underspecification infers that a lack of certain information in the mental lexicon can result in asymmetric perception, which will only aid listeners in their resolution of speech signals that are highly variable and often presented in noisy environments (Cornell, Lahiri, & Eulitz, 2013). The obvious advantage in positing underspecified representations is the reduced cognitive load that is required when coping with such variability in speech, which has important consequences for lexical search and identification mechanisms. The more that is stored, the more the recognition process must cope with in order to identify the correct word.

The underpinnings of the FUL model can be applied to the results of this thesis. Given that each Sino-Korean morpheme also has a bundle of phonological features, this means that in addition to the semantic cohort, there is also a cohort of sounds that are activated which overlap in phonological features. However, depending on which features are present in the signal, and which are represented, the relative size of lexical candidates with featural overlap will vary. As Figure 7.1 shows, the two Sino-Korean morphemes are not only varied in the relative number of available meanings, but also differ with regards to the size of their respective phonological cohort. Both Sino-Korean morphemes [t^h] and [p^h] share features (e.g., [OBSTRUENT], [PLOSIVE], and [ASPIRATION]), but they also crucially differ in their specification of place of articulation.

The feature [LABIAL] can be extracted from the speech signal in [p^h], which activates its own representation, while for [t^h], the feature [CORONAL] can be extracted from the speech signal, but as it is underspecified in the mental representation, it results in a 'no-mismatch'. This means that while [p^h] can activate a cohort of sounds that also contains [t^h], because [t^h] is underspecified, [t^h] cannot activate a

cohort containing [p^h], because [p^h] is specified for place of articulation, and there is conflict. Thus, despite being underspecified, and in some sense, less restricted, the cohort is much smaller for [t^h] than it is for [p^h], which is, on the other hand, fully specified for place of articulation.

The extent to which phonological featural information is specified or underspecified will have important consequences for the formation of a cohort comprising phonemes that share features. This information in itself also forms a cohort, which has the potential to interact with other cohorts on different levels of representation.

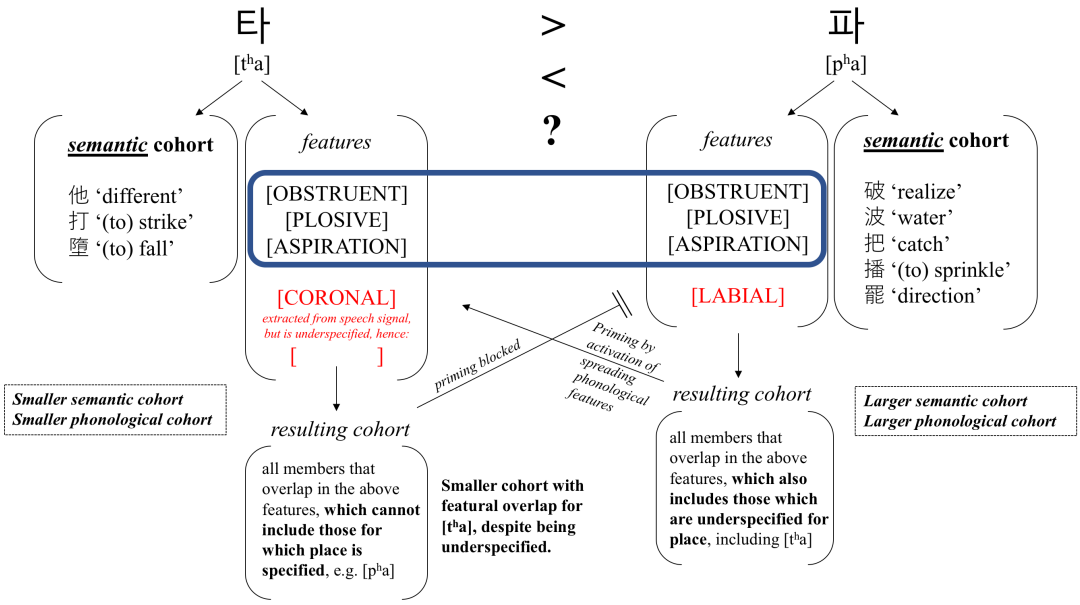


Figure 7.1: Two different Sino-Korean morphemes with different cohort sizes on both the semantic and the featural level.

This direction in research would help to refine our understanding of whether the relative size of a particular cohort outweighs the effects of another cohort, especially in the context of underspecification due to universal default features. The size of the cohort will increase if sensitivity to contrasts is considered asymmetric. Therefore, particular morphemes may result in larger cohorts than others depending on their phonology and not just their semantics.

At first, it may seem that for Sino-Korean there are far more semantic features than there are phonological ones, thus making the relevance of a theoretical parallel to a phonologically underspecified model not immediately apparent. To clarify,

however, the idea is that, given that both sets of features, phonological and semantic, are likely to interact in the lexical representation of Sino-Korean words, positing an underspecified representation may provide a parsimonious architecture to account for both. Further research would be helpful in empirically substantiating these claims. In any case, theorizing notions of semantic underspecification has yet to receive the full attention it deserves, particularly as it relates to semanto-orthographic relations. As has been shown in this thesis, Sino-Korean is especially well-suited for such investigation. Further

7.4 Expanding the revised model of the Korean mental lexicon

I suggest three possible directions through which the presently proposed theoretical model of the Korean mental lexicon can be expanded. So far, this research has presented a mostly autonomous model of the Sino-Korean mental lexicon, that is, it has yet to be built into a larger model which accounts for entire Korean mental lexicon. As such, the next step would be to integrate this model into a larger theoretical proposal positing a more holistic structure of the Korean mental lexicon, drawing together both the Sino-Korean lexicon and the pure Korean lexicon.

Proper theoretical integration of the model presented in Chapter 6 into a larger conceptualization of the Korean mental lexicon is particularly important, as there is a subset of Hangul representations that the two sub-lexicons share between them. For example, the Hangul representation of 유 [ju] has instantiations in both Sino-Korean words as well as pure Korean words. Further research would be necessary to better understand how such forms like 유 [ju] are mentally represented in the larger Korean lexicon, given that it surfaces in both sub-lexicons. For purposes of parsimony, it seems sensible to posit a single morphemic entry for 유 [ju] at a level of representation that is potentially higher than that of the conceptual level given in the revised model in Chapter 6, but this remains open to further investigation.

How this subset of characters is accounted for also has important consequences for semantic cohort formation. Depending on the specific sub-lexicon, morphemes

like 유 [ju] that are shared across the two can potentially have a cohort of many, or a cohort of zero. It may have to be the case that for morphemes like 유 [ju], which can function as either a Sino-Korean or a pure Korean morpheme, the semantic cohort can be conceptualized as the set that contains all possible meanings of different Hanja characters that share the Hangul representation 유, as well as the empty set, to account for instantiations of 유 in pure Korean words, where meanings are not necessarily semantically compositional, nor derived from the semantics of Hanja characters. Another theoretical consideration that arises is whether 유 [ju], as a pure Korean morpheme, can be considered semantically underspecified, possibly in the same way that featural information is sometimes underspecified according to FUL. All of these strands pertaining to how the two sub-lexicons in Korean might interact would prove fruitful lines of further research.

Another direction in which the model could be expanded is how the mental representations of Hanja characters differ for speakers with more or less proficiency in written Korean, particularly in Hanja. The present research tested monolingual native speakers, who are likely to have had significant schooling in and exposure to Hanja characters, but the question emerges whether heritage speakers of Korean maintain any representation of Hanja characters. This is particularly relevant given that the findings here indicated a conceptual representation of the semantics of Hanja, meaning that literacy or any type of proficiency in the script is not necessarily required. Heritage speakers, through informal exposure of Hangul, may begin to develop certain patterns as they observe the same Hangul form occurring in different environments, taking on different lexical instantiations with different meanings. These patterns would also yield important considerations for how then to model the Korean mental lexicon, which should not be specific only to the native speaker of Korean.

Furthermore, this direction of research is not just specific to heritage speakers, but all speakers of Korean who are not in a Korean-dominant environment. Do speakers of varying levels of proficiency in Hanja have vastly different mental representations of the semantics of Hanja characters? Would the cross-script priming effects observed

here be less likely to be observed in a different demographic of Korean speakers, such as heritage speakers in a non-Korean dominant environment? The analyses in Chapter 6 already suggest a degree of variability in the representation of individual Sino-Korean morphemes, but how in flux these representations are remains an unanswered question. Additional scholarship addressing these questions would help to form a more cohesive understanding of the role of Hanja characters in linguistic processing and tie in well with the adjacent literature on language attrition.

A computational model of the semantics of Hanja characters would also provide a complementary direction to the research presented in this thesis, to see whether a model can be trained on not only the larger semantic networks of Sino-Korean, but also on the individual cross-script semantic cohorts. If so, it would be of interest to see whether the results of the computational model align with those of the behavioral results presented here. Future research on the computational modelling of the effects of Hanja characters could apply the "triangle model" (Seidenberg, 2005; Seidenberg & McClelland, 1989; Harm & Seidenberg, 2004; Plaut et al. 1996; see Nayernia, de Vijver & Indefrey, 2019 for a review of studies examining the influence of orthography on processing across different tasks) to Sino-Korean and the reading of Hanja characters, as a potential starting point.

More broadly, the question of how heritage speakers represent the contributions of Hanja characters invites additional lines of inquiry. The assumption stated above suggests that it is unlikely for heritage speakers to have as much competency in recognizing Hanja characters, as Korean-educated native speakers; however, regardless of differences in competency, heritage and native speakers alike are still likely to share the same lexical items in their mental lexicons. The question therefore arises whether heritage speakers can form the same connections between non-categorical meanings, as described in Chapter 6, e.g., $[mi]_1$, 'rice', and $[mi]_2$ 'rice'. Can heritage speakers form connections between the different non-categorically related meanings on sound alone? And if so, to what extent are the orthographic representation needed? This invites one final question which is how the potentially variable status of script during lexical processing affects the state of the Korean

mental lexicon, particularly if no representation of cross-script relations is necessary? What does this mean about the status of Hanja specifically? While this thesis made a case for some representation of script, especially within the context of Korean, it is clear that there still remains much to be explored.

Overall, this series of experiments offers an integral step towards understanding the mental representations found in a bi-scriptal lexicon. Korean is only one exemplar out of many others where the writing system presents a unique lens for exploration. More importantly, this thesis has helped to bridge the gap between written and spoken language, particularly in a field where often the contributions of the writing system are disregarded. Examining the status of written language in cognition can offer key insights, and the findings and the analyses presented here attest to such potential. Understanding the role of script and the contributions of interlinked orthographic representations will only enrich our models of the lexicon and theories of language processing.

Appendices

A Pre-experiment information-gap activity



Figure 2: Copy of the set of pictures used for the information-gap activity (Richards & Schmidt, 2009)

B Stimulus set for Experiment 1 (Chapter 4)

Table 1: The prime-target pairs for Experiment 1. The syllable indicated in boldface is the Sino-Korean morpheme that is of interest. The target is a Sino-Korean word containing a morpheme which can encode the semantic contributions of more than one Hanja character. The test prime is one of the possible Hanja characters whose meaning can be encoded by the Sino-Korean morpheme being tested, but crucially whose meaning is incongruent with the meaning actually presented in the full word. The control prime is a semantically unrelated Hanja character.

Target (Hangul)	*Test prime (Hanja)*	*Control prime (Hanja)*
기억, [ki.ʌk], 'memory'	己, 'body'	罪, 'sin'
미인, [mi.in], 'beautiful person'	米, 'rice'	角, 'horn'
명함, [mjʌŋ.ham], 'business card'	明, 'be bright'	開, 'to open'
희망, [hɨi.maŋ], 'hope'	喜, 'to be happy'	足, 'foot'
하복, [ha.pok], 'summer clothes'	下, 'summer'	大, 'big'
남부, [nam.pu], 'the South'	男, 'male'	果, 'fruit'
여행, [jʌ.hɛŋ], 'travel, journey'	女, 'female'	中, 'middle'
고려, [ko.rjʌ], 'consideration'	高, 'high, tall'	家, 'home'
광장, [koaŋ.tsaŋ], 'plaza'	光, 'light'	骨, 'bone'
문제, [mun.tse], 'problem'	門, 'door'	骨, 'bone'
무용, [mu.joŋ], 'dancing'	無, 'absence'	母, 'Mother'
학대, [hak.tɛ], 'abuse'	學, 'to learn'	金, 'iron'
입**구**, [ip.ku], 'entrance'	九, 'nine'	間, 'in between'
관**상**, [koan.saŋ], 'physiognomy'	上, 'above'	刀, 'knife'
단**독**, [tan.tok], 'one person'	讀, 'to read'	馬, 'horse'
희**생**, [hɨi.saŋ], 'sacrifice'	生, 'birth'	校, 'school'
팔**천**, [pʰal.tsʰʌn], 'eight-thousand'	天, 'sky'	工, 'expert'
곤**충**, [kon.tsʰuŋ], 'insect'	忠, 'loyalty'	前, 'front'
삼**백**, [sam.pɛk], 'three-hundred'	白, 'white'	主, 'owner'
화**산**, [hoa.san], 'volcano'	算, 'calculate'	內, 'inside'
검**색**, [kʌm.sɛk], 'search'	色, 'color'	左, 'left'
상**식**, [saŋ.sik], 'common sense'	食, 'meal'	弓, 'arrow'
우**연**, [u.jʌn], 'coincidence'	年, 'year'	君, 'king'
자**동**, [tsa.toŋ], 'automatic'	東, 'East'	多, 'many'

C Stimulus set for Experiment 2 (Chapter 4)

Table 2: The prime-target pairs for Experiment 2, across the three experimental conditions. The Sino-Korean morpheme of interest appeared in both the *Directly Related* and the *Indirectly Related* conditions, each one lexically instantiating one of the possible meanings encoded by the morpheme. For some, the position in which the Sino-Korean morpheme of interest occured in the Directly Related prime is not necessarily the same position in which it appeared in the Indirectly Related prime. This was controlled to the greatest extent possible, but it was difficult to uphold uniformly across all primes. A post-hoc analysis of reaction times for each condition (Directly and Indirectly Related) revealed that reaction times to targets were significantly faster when preceded by a prime in which the morpheme appeared syllable-initial ($F(1, 31.58) = 8.515$, $p = 0.006$* for Indirectly Related; $F(1, 30.21) = 4.386$, $p = 0.045$* for Directly Related).

Target (Hangul)	*Directly Related*	*Indirectly Related*	*Unrelated*
그림 [ki.ɾim] 'picture'	도화 [to.hoa] 'drawing'	도로 [to.ro] 'road'	치약 [tsʰ.jak] 'toothpaste'
눈병 [nun.pjʌŋ] 'eye infection'	안과 [an.koa] 'ophthalmology'	안정 [an.tsʌŋ] 'stability'	감자 [kam.tsa] 'potato'
슬픔 [sil.pʰim] 'sadness'	애도 [ɛ.to] 'mourning'	애교 [ɛ.kjo] 'loveliness'	합성 [hap.sʌŋ] 'agreement'
철회 [tsʰʌl.hø] 'withdrawal'	무효 [mu.hjo] 'invalidity'	무용 [mu.joŋ] 'dance'	소풍 [so.pʰuŋ] 'outing'
꽃밭 [k*ot.pat] 'flower patch'	화분 [hoa.pun] 'flower pot'	화색 [hoa.sɛk] 'healthy glow'	우유 [u.ju] 'milk'
구름 [ku.ɾim] 'cloud'	운무 [un.mu] 'cloud'	운행 [un.hɛŋ] 'run'	회사 [hø.sa] 'office'
아내 [a.nɛ] 'wife'	처가 [tsʰʌ.ka] 'wife's family home'	출처 [tsʰul.tsʰʌ] 'origin'	호흡 [ho.hip] 'breathing'
확실 [hoak.sil] 'certainty'	명백 [mjʌŋ.pɛk] 'clarity'	명단 [mjʌŋ.tan] 'list'	허파 [hʌ.pʰa] 'lung'
높음 [nopʰ.im] 'height'	고급 [ko.kip] 'high quality'	고려 [ko.ɾjʌ] 'consideration'	수박 [su.pak] 'watermelon'
울림 [ul.rim] 'reverberation'	음향 [im.hjaŋ] 'sound'	향기 [hjaŋ.ki] 'fragrance'	설탕 [sʌl.tʰaŋ] 'sugar'
사이 [sa.i] 'relationship'	중간 [tsuŋ.kan] 'middle'	간편 [kan.pʰjʌn] 'convenience'	영화 [jʌŋ.hoa] 'film'
벌레 [pʌl.re] 'bug'	곤충 [kon.tsʰuŋ] 'insect'	충격 [tsʰŋ.kjʌk] 'shock'	거울 [kʌ.ul] 'mirror'
말씀 [mal.s*im] 'words'	면담 [mjʌn.tam] 'face-to-face talk'	냉담 [nɛŋ.tam] 'apathy'	기린 [ki.rin] 'giraffe'
소원 [so.uʌn] 'wish'	희망 [hii.maŋ] 'jubilation'	희박 [hii.pak] 'rarity'	직각 [tsik.kak] 'right angle'
먼저 [mʌn.tsʌ] 'first'	선배 [sʌn.pɛ] 'senior'	선거 [sʌn.kʌ] 'election'	단어 [tan.ʌ] 'word'
마당 [ma.taŋ] 'yard'	농장 [noŋ.tsaŋ] 'farm'	장벽 [tsaŋ.pjʌk] 'long wall'	규격 [kju.kjʌk] 'standard'
맺음 [mɛts.im] 'closure'	결론 [kjʌl.ron] 'conclusion'	순결 [sun.kjʌl] 'purity'	만두 [man.tu] 'dumpling'
임금 [im.kim] 'monarch'	주군 [tsu.kun] 'master'	육군 [juk.kun] 'army'	방청 [paŋ.tsʰʌŋ] 'attend'
머슴 [mʌ.sim] 'servant'	노비 [no.pi] 'servants'	분노 [pun.no] 'anger'	썰물 [s*ʌl.mul] 'ebb'
시초 [si.tsʰo] 'root'	원조 [uʌn.tso] 'original'	숙원 [suk.uʌn] 'long-cherished wish'	실내 [sil.nɛ] 'interior'
보리 [po.ri] 'barley'	맥주 [mɛk. tsu] 'beer'	산맥 [san.mɛk] 'mountain range'	여분 [jʌŋ.pun] 'remainder'
엄마 [ʌm.ma] 'mother'	모친 [mo.tsʰ in] 'mother'	모방 [mo.paŋ] 'imitation'	통행 [tʰoŋ.hɛŋ] 'passing through'
벼슬 [pjʌ.sil] 'government post'	법관 [pʌp.koan] 'judge'	외관 [ø.koan] 'exterior'	삼촌 [sam.tsʰon] 'uncle'
다툼 [ta.tʰum] 'struggle'	논쟁 [non.tsɛŋ] 'debate'	쟁쟁 [tsɛŋ.tsɛŋ] 'to be oustanding'	응용 [iŋ.joŋ] 'application'
미리 [mi.ri] 'ahead'	예비 [je.pi] 'reserve'	예의 [je.ri] 'etiquette'	겨울 [kjʌ.ul] 'winter'
저녁 [tsʌ.njʌk] 'dinner'	석양 [sʌk.jaŋ] 'sunset'	좌석 [tsoa.sʌk] 'seat'	폭포 [pʰok.pʰo] 'waterfall'
걸음 [kʌl.im] 'step'	보폭 [po.pʰok] 'stride'	보건 [po.kʌn] 'health'	후추 [hu.tsʰu] 'pepper'
우측 [u.tsʰik] 'the right-side'	동해 [toŋ.hɛ] 'East sea'	자동 [tsa.toŋ] 'automatic'	끄덕 [k*i.t*ʌk] 'nod'
고요 [ko.jo] 'silence'	적막 [tsʌk.mak] 'stillness'	유적 [ju.tsʌk] 'historic remains'	피망 [pʰ.maŋ] 'pepper'
부처 [pu.tsʰʌ] 'Buddha'	불상 [pul.saŋ] 'image of Buddha'	불변 [pul.pjʌn] 'inconvenience'	참새 [tsʰa.sɛ] 'sparrow'
소년 [so.njʌn] 'boy'	남성 [nam.sʌŋ] 'man'	남용 [nam.joŋ] 'abuse'	숙직 [suk.tsik] 'night duty'
새싹 [sɛ.s*ak] 'sprout'	발아 [pal.a] 'germination'	아동 [a.toŋ] 'child'	완료 [oan.ɾjo] 'completion'
화합 [hoa.hap] 'harmony'	타협 [tʰa.hjʌp] 'compromise'	협공 [hjʌp.koŋ] 'attach from both sides'	늑대 [nik.dɛ] 'wolf'
터짐 [tʰʌ.tsim] 'burst'	폭죽 [pʰok.tsuk] 'firecracker'	폭설 [pʰk.sʌl] 'heavy snow'	스님 [si.nim] 'monk'
구슬 [ku.sil] 'marble'	옥석 [ok.sʌk] 'jade'	옥토 [ok.tʰo] 'rich soil'	마늘 [ma.nil] 'garlic'
소리 [so.ri] 'sound'	음악 [im.ak] 'music'	음영 [im.jʌŋ] 'shadow'	얼룩 [ʌl.ruk] 'stain'

D Subset of targets primed in Experiment 2

Table 3: The subset of prime-target pairs for which there was active facilitation in the Directly Related condition in Experiment 2. The exploratory analyses presented in Chapter 6 are based on the reaction times for this particular subset of targets.

Target (Hangul)	*Directly Related*	*Indirectly Related*	*Unrelated*
눈병 [nun.pjʌŋ] 'eye infection'	안과 [an.koa] 'ophthalmology'	안정 [an.tsʌŋ] 'stability'	감자 [kam.tsa] 'potato'
슬픔 [sil.pʰim] 'sadness'	애도 [ɛ.to] 'mourning'	애교 [ɛ.kjo] 'loveliness'	합성 [hap.sʌŋ] 'agreement'
꽃밭 [k*ot.pat] 'flower patch'	화분 [hoa.pun] 'flower pot'	화색 [hoa.sɛk] 'healthy glow'	우유 [u.ju] 'milk'
아내 [a.nɛ] 'wife'	처가 [tsʰʌ.ka] 'wife's family home'	출처 [tsʰul.tsʰʌ] 'origin'	호흡 [ho.hip] 'breathing'
확실 [hoak.sil] 'certainty'	명백 [mjʌŋ.pɛk] 'clarity'	명단 [mjʌŋ.tan] 'list'	허파 [hʌ.pʰa] 'lung'
높음 [nopʰ.im] 'height'	고급 [ko.kip] 'high quality'	고려 [ko.rjʌ] 'consideration'	수박 [su.pak] 'watermelon'
벌레 [pʌl.re] 'bug'	곤충 [kon.tsʰuŋ] 'insect'	충격 [tsʰŋ.kjʌk] 'shock'	거울 [kʌ.ul] 'mirror'
소원 [so.uʌn] 'wish'	희망 [hii.maŋ] 'jubilation'	희박 [hii.pak] 'rarity'	직각 [tsik.kak] 'right angle'
먼저 [mʌn.tsʌ] 'first'	선배 [sʌn.pɛ] 'senior'	선거 [sʌn.kʌ] 'election'	단어 [tan.ʌ] 'word'
임금 [im.kim] 'monarch'	주군 [tsu.kun] 'master'	육군 [juk.kun] 'army'	방청 [paŋ.tsʰʌŋ] 'attend'
머슴 [mʌ.sim] 'servant'	노비 [no.pi] 'servants'	분노 [pun.no] 'anger'	썰물 [s*ʌl.mul] 'ebb'
시초 [si.tsʰo] 'root'	원조 [uʌn.tso] 'original'	숙원 [suk.uʌn] 'long-cherished wish'	실내 [sil.nɛ] 'interior'
보리 [po.ri] 'barley'	맥주 [mɛk. tsu] 'beer'	산맥 [san.mɛk] 'mountain range'	여분 [jʌŋ.pun] 'remainder'
다툼 [ta.tʰum] 'struggle'	논쟁 [non.tsɛŋ] 'debate'	쟁쟁 [tsɛŋ.tsɛŋ] 'to be oustanding'	응용 [iŋ.joŋ] 'application'
저녁 [tsʌ.njʌk] 'dinner'	석양 [sʌk.jaŋ] 'sunset'	좌석 [tsoa.sʌk] 'seat'	폭포 [pʰok.pʰo] 'waterfall'
고요 [ko.jo] 'silence'	적막 [tsʌk.mak] 'stillness'	유적 [ju.tsʌk] 'historic remains'	피망 [pʰ.maŋ] 'pepper'
부처 [pu.tsʰʌ] 'Buddha'	불상 [pul.saŋ] 'image of Buddha'	불변 [pul.pjʌn] 'inconvenience'	참새 [tsʰa.sɛ] 'sparrow'
화합 [hoa.hap] 'harmony'	타협 [tʰa.hjʌp] 'compromise'	협공 [hjʌp.koŋ] 'attach from both sides'	늑대 [nik.dɛ] 'wolf'

E Stimulus set for Experiment 4 (Chapter 5)

Table 4: The prime-target pairs for Experiment 4, across the three experimental conditions. The Sino-Korean morpheme of interest appeared in both the *Directly Related* and the *Indirectly Related* conditions, each one lexically instantiating one of the possible meanings encoded by the morpheme. For some, the position in which the Sino-Korean morpheme of interest occured in the Directly Related prime is not necessarily the same position in which it appeared in the Indirectly Related prime. This was controlled to the greatest extent possible, but it was difficult to uphold uniformly across all primes. A post-hoc analysis of reaction times for each condition (Directly and Indirectly Related) revealed that reaction times to targets were faster when preceded by a prime in which the morpheme appeared syllable-initial, though this was not significant ($F(1, 28.34) = 2.727$, $p = 0.109$ for Indirectly Related; $F(1, 28.47) = 3.105$, $p = 0.089$ for Directly Related).

Target (Hangul)	*Directly Related*	*Indirectly Related*	*Unrelated*
종결 [tsoŋ.kjʌl] 'termination'	졸업 [tsol.ʌp] 'graduation'	치졸 [tsʰi.tsol] 'crudity'	질투 [tsil.tʰu] 'jealousy'
배려 [pɛ.ɾjʌ] 'consideration'	애정 [ɛ.tsʌŋ] 'affection'	정의 [tsʌŋ.ii] 'justice'	식품 [sik.pʰum] 'groceries'
뿌리 [p*u.ri] 'root'	근본 [kin.pon] 'foundation'	근대 [kin.tɛ] 'modern time'	긴급 [kin.kip] 'emergency'
여름 [jʌ.ɾim] 'summer'	하계 [ha.kje] 'summer season'	하체 [ha.tsʰe] 'lower body'	필통 [pʰil.toŋ] 'pencil case'
소금 [so.kim] 'salt'	염전 [jm.tsʌn] 'saltpan'	염색 [jʌm.sɛk] 'dye'	최대 [tsʰue.tɛ] 'maxiumum'
종이 [tsoŋ.i] 'paper'	백지 [pɛk.tsi] 'blank paper'	지구 [tsi.ku] 'earth'	언니 [ʌn.ni] 'sister'
참음 [tsʰam.im] 'tolerance'	인내 [in.nɛ] 'patience'	인사 [in.sa] 'greeting'	과립 [kua.iip] 'capsule'
회상 [hø.saŋ] 'recollection'	기억 [ki.ʌk] 'memory'	공기 [koŋ.ki] 'atmosphere'	욕실 [jok.sil] 'classroom'
바람 [pa.ɾam] 'wind'	해풍 [hɛ.pʰuŋ] 'sea breeze'	중풍 [tsuŋ.pʰuŋ] '(medical) stroke'	통계 [tʰoŋ.kje] 'statistics'
곡식 [kok.sik] 'grain'	현미 [hjʌn.mi] 'brown rice'	미술 [mi.sul] 'art'	토굴 [tʰo.kul] 'cave'
넓음 [nʌlp.im] 'vastness'	광장 [koaŋ.tsaŋ] 'plaza'	광택 [koaŋ.tʰɛk] 'shine'	만화 [man.hoa] 'cartoon'
겨레 [kjʌ.ɾe] 'brethren'	민족 [min.tsok] 'nation's people'	민감 [min.kam] 'sensitivity'	앵두 [ɛŋ.tu] 'cherry'
승리 [siŋ.ɾi] 'victory'	극복 [kik.pok] 'conquest'	극심 [kik.sim] 'extremity'	황태 [hoaŋ.tʰɛ] 'pollock'
행복 [hɛŋ.pok] 'happiness'	환희 [hoan.hii] 'jubilation'	환불 [hoan.pul] 'refund'	손톱 [son.tʰop] 'fingernail'
홀로 [hol.ɾo] 'alone'	단독 [tan.tok] 'independence'	독서 [tok.sʌ] 'reading'	목차 [mok.tsʰa] 'table of contents'
죽음 [tsuk.im] 'demise'	사망 [sa.maŋ] 'death'	식사 [sik.sa] 'meal'	복습 [pok.sip] 'review'
출산 [tsʰul.san] 'childbirth'	생일 [saŋ.il] 'birthday'	희생 [hii.saŋ] 'sacrifice'	거위 [kʌ.ui] 'goose'
상처 [saŋ.tsʰʌ] 'wound'	학대 [hak.tɛ] 'abuse'	학교 [hak.kjo] 'school'	액자 [ɛk.tsa] 'frame'
거짓 [kʌ.tsit] 'lie'	가면 [ka.mjʌn] 'mask'	가훈 [ka.hun] 'family'	복학 [pok.hak] 'return to school'
어음[ʌ.im] 'promissory note'	증권 [tsiŋ.kuʌn] 'stock'	권력[kuʌn.ɾjʌk] 'authority'	느낌 [ni.k*im] 'feeling'
여자 [jʌ.tsa] 'woman'	자매 [tsa.mɛ] 'sisters'	중매 [tsuŋ.mɛ] 'matchmaking'	경기 [kjʌŋ.ki] 'game'
일찍 [il.ts*ik] 'early'	조퇴 [tso.tø] 'early departure'	조세 [tso.se] 'taxes'	의자 [ii.tsa] 'chair'
나비 [na.pi] 'butterfly'	접영 [tsʌp.jʌŋ] 'domination'	접합 [tsʌp.hap] 'joining'	꼬리 [k*o.ri] 'tail'
만듦 [man.dim] 'creation'	작품 [tsak.pum] 'masterpience'	수작 [su.tsak] 'nonsense'	날개 [nal.kɛ] 'wings'
노래 [no.ɾɛ] 'song'	가요 [ka.jo] 'music'	요리 [jo.ri] 'cooking'	장갑 [tsaŋ.kap] 'glove'
세움 [se.um] 'standing'	건국 [kʌn.kuk] 'foundation'	두건 [tu.kʌn] 'bandana'	교활 [kjo.hoal] 'artfulness'
가난 [ka.nan] 'impoverishment'	빈곤 [pin.kon] 'poverty'	귀빈 [kui.pin] 'VIP'	하늘 [ha.nil] 'sky'
가슴 [ka.sim] 'chest'	흉상 [hjuŋ.saŋ] 'bust'	흉년 [hjuŋ.njʌn] 'year of famine'	번역 [bʌn.jʌk] 'translation'
가을 [ka.il] 'autumn'	추분 [tsʰu.pun] 'autumn equinox'	추억 [tsʰu.ʌk] 'memory'	파도 [pa.to] 'wave'
통증 [tʰoŋ.tsiŋ] 'pain'	고뇌 [ko.nø] 'anguish'	두뇌 [tu.nø] 'brain'	접시 [tsʌp.si] 'plate'
스승 [si.siŋ] 'teacher'	사부 [sa.pu] 'master'	부활 [pu.hoal] 'revival'	멈춤 [mʌm.tsʰum] 'stall'
조각 [tso.kak] 'piece'	파편 [pa.pjʌn] 'fragment'	편식 [pjʌn.sik] 'unbalanced diet'	포함 [po.ham] 'inclusion'
영웅 [jʌŋ.uŋ] 'hero'	승자 [siŋ.tsa] 'victor'	승마 [siŋ.ma] 'horseback riding'	먹이 [mʌk.i] 'prey'
만남 [man.nam] 'meeting'	상봉 [saŋ.boŋ] 'reunion'	봉투 [poŋ.tʰu] 'envelope'	골절 [kol.tʌl] 'fracture'
생선 [sɛŋ.sʌn] 'fish'	어부 [ʌ.pu] 'fisherman'	어뢰 [ʌ.ɾø] 'torpedo'	범위 [pʌm.ui] 'scope'
나이 [na.i] 'age'	연세 [jʌn.se] 'age'	인연 [in.jʌn] 'destiny'	참외 [tsʰam.ø] 'melon'

F Subset of targets primed in Experiment 4

Table 5: The subset of prime-target pairs for which there was active facilitation in the Directly Related condition in Experiment 4. The exploratory analyses presented in Chapter 6 are based on the reaction times for this particular subset of targets.

Target (Hangul)	*Directly Related*	*Indirectly Related*	*Unrelated*
종결 [tsoŋ.kjʌl] 'termination'	졸업 [tsol.ʌp] 'graduation'	치졸 [ts^{h}i.tsol] 'crudity'	질투 [tsil.t^{h}u] 'jealousy'
뿌리 [p*u.ɾi] 'root'	근본 [kin.pon] 'foundation'	근대 [kin.tɛ] 'modern time'	긴급 [kin.kip] 'emergency'
여름 [jʌ.ɾim] 'summer'	하계 [ha.kje] 'summer season'	하체 [ha.ts^{h}e] 'lower body'	필통 [p^{h}il.toŋ] 'pencil case'
소금 [so.kim] 'salt'	염전 [jm.tsʌn] 'saltpan'	염색 [jʌm.sɛk] 'dye'	최대 [ts^{h}ue.tɛ] 'maxiumum'
종이 [tsoŋ.i] 'paper'	백지 [pɛk.tsi] 'blank paper'	지구 [tsi.ku] 'earth'	언니 [ʌn.ni] 'sister'
참음 [ts^{h}am.im] 'tolerance'	인내 [in.nɛ] 'patience'	인사 [in.sa] 'greeting'	과립 [kua.iip] 'capsule'
곡식 [kok.sik] 'grain'	현미 [hjʌn.mi] 'brown rice'	미술 [mi.sul] 'art'	토굴 [t^{h}o.kul] 'cave'
넓음 [nʌlp.im] 'vastness'	광장 [koaŋ.tsaŋ] 'plaza'	광택 [koaŋ.t^{h}ɛk] 'shine'	만화 [man.hoa] 'cartoon'
행복 [hɛŋ.pok] 'happiness'	환희 [hoan.hii] 'jubilation'	환불 [hoan.pul] 'refund'	손톱 [son.t^{h}op] 'fingernail'
죽음 [tsuk.im] 'demise'	사망 [sa.maŋ] 'death'	식사 [sik.sa] 'meal'	복습 [pok.sip] 'review'
세움 [se.um] 'standing'	건국 [kʌn.kuk] 'foundation'	두건 [tu.kʌn] 'bandana'	교활 [kjo.hoal] 'artfulness'
가난 [ka.nan] 'impoverishment'	빈곤 [pin.kon] 'poverty'	귀빈 [kui.pin] 'VIP'	하늘 [ha.nil] 'sky'
가슴 [ka.sim] 'chest'	흉상 [hjuŋ.saŋ] 'bust'	흉년 [hjuŋ.njʌn] 'year of famine'	번역 [bʌn.jʌk] 'translation'
통증 [t^{h}oŋ.tsiŋ] 'pain'	고뇌 [ko.nø] 'anguish'	두뇌 [tu.nø] 'brain'	접시 [tsʌp.si] 'plate'
스승 [si.siŋ] 'teacher'	사부 [sa.pu] 'master'	부활 [pu.hoal] 'revival'	멈춤 [mʌm.ts^{h}um] 'stall'
영웅 [jʌŋ.uŋ] 'hero'	승자 [siŋ.tsa] 'victor'	승마 [siŋ.ma] 'horseback riding'	먹이 [mʌk.i] 'prey'
만남 [man.nam] 'meeting'	상봉 [saŋ.boŋ] 'reunion'	봉투 [poŋ.t^{h}u] 'envelope'	골절 [kol.tʌl] 'fracture'
나이 [na.i] 'age'	연세 [jʌn.se] 'age'	인연 [in.jʌn] 'destiny'	참외 [ts^{h}am.ø] 'melon'

Bibliography

Abdel Rahman, R. & Melinger, A. (2007). When bess hamper the production of honey: Lexical interference from associates in speech produciton. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 604-614.

Abdel Rahman, R. & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes, 24*(5), 713-734.

Abdel Rahman, R. & Melinger, A. (2019). Semantic processing during language production: An update of the swinging lexical network. *Language, Cognition, and Neuroscience, 34*(9), 1176-1192.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814-823.

Ahn, S.-C. (2009). *Korean phonology*. London, England: Saffron Books.

Akmaijian, A., Demers, R. A., Farmer, A. K., & Harnish, R. M. (2001). *Linguistics: An introduction to language and communication (5th Ed.)*. Cambridge, MA: The MIT Press.

Allen, M. & Badecker, W. (2002). Stem homographs and lemma level representations. *Brain and Language, 81*, 79-88.

Almabruk, A. A., Paterson, K. B., McGowan, V., & Jordan, T. R. (2011). Evaluating effects of divided hemispheric processing on word recognition in foveal and extrafoveal displays: The evidence from Arabic. *PLoS ONE, 6*(4), e18131.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*(4), 439–461.

Asfaha, Y. M., Kurvers, J., & Kroon, S. (2009). Grain size in script and teaching: Literacy acquisition in Ge'ez and Latin. *Applied Psycholinguistics, 30*(4), 709-724.

Aro, M. (2006). Learning to read: The effect of orthography. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 531-550). London: Lawrence Erlbaum Associates.

Azuma, T. & Van Orden, G. C. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language, 36*, 484-504.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 153*(2), 283-316.

Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (2006). Single- versus dual-process models of lexical decision performance: Insights from response time distributional analysis. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1324–1344.

Balota, D. A. & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 336-345.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*(1), 32–55.

Baluch, B. & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 644-652.

Baron, J. (1973). Phonemic stage not necessary for reading: Lexical access and/or production? *Quarterly Journal of Experimental Psychology, 25*, 241-246.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 225-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects Models using lme4. *Journal of Statistical Software, 67*, 1-48.

Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child Development Perspectives, 9*(2), 117-121.

Binder, K. S. & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*(5), 1186-1196.

Binder, K. S. & Morris, R. K. (2011). An eye movement analysis of ambiguity resolution: Beyond meaning access. *Discourse Processes, 48*, 305-330.

Borowsky, R. & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal Experimental Psychology: Learning, Memory, and Cognition, 22*, 63-85.

Boot, I. & Pecher, D. (2008). Word recognition is affected by the meaning of orthographic neighbors: Evidence from semantic decision tasks. *Language and Cognitive Processes, 23*, 375-393.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*(3), B45-54.

Bowers, J. S. & Michita, Y. (1998). An investigation into the structure and acquisition of orthographic knowledge: Evidence from cross-script kanji–hiragana priming. *Psychonomic Bulletin & Review, 5*(2), 259–264.

Bowers, J.S., Mimouni, Z. & Arguin, M. (2000). Orthography plays a critical role in cognate priming: Evidence from French/English and Arabic/French cognates. *Memory & Cognition, 28*, 1289-1296.

Buchanan, L., Westbury, C. F., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8*(3), 531-544.

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 567-575.

Burton, M. W. & Blumstein, S. E. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 1230-1235.

Bybee, J. L. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 119-141). New York, NY: Academic Press.

Bybee, J. L. (2001). *Phonology and language use*. Cambridge, England: Cambridge University Press.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 857–871.

Chang, S.-J. (1996). *Korean*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Chen, H-C., Vaid, J., Boas, D. A., & Bortfeld, H. (2011). Examining the phonological neighborhood density effect using near infrared spectroscopy. *Human Brain Mapping, 32*(9), 1363-1370.

Chen, H.-C., Flores D'Arcais, G. B., & Cheung, S.-L. (1995). Orthographic and phonological activation in recognizing Chinese characters. *Psychological Research, 58*(2), 144–153.

Cho, C.-H., Sohn, Y.-J., & Yang, H. (2002). *Korean reader for Chinese characters*. Honolulu, HI: University of Hawaii Press.

Cho, J.-R. (2009). The role of phonology in word recognition of Korean Hangul and Hanja. In C. Lee, G. Simpson, & Y. Kim (Eds.), *The handbook of East Asian psycholinguistics* (pp. 409-417). Cambridge, England: Cambridge University Press.

Cho, J.-R. & Chen, H.-C. (1999). Orthographic and phonological activation in the semantic processing of Korean Hanja and Hangul. *Language and Cognitive Processes, 14*(5-6), 481-502.

Cho, J.-R., McBride-Chang, C., & Park, S.-G. (2008). Phonological awareness and morphological awareness: Differential associations to regular and irregular word recognition in early Korean Hangul readers. *Reading and Writing: An Interdisciplinary Journal, 21*(3), 255–274.

Cho, J.-R. & Chen, H.-C. (2005). Semantic and phonological processing in reading Korean Hangul and Hanja words. *Journal of Psycholinguistic Research, 34*(4), 401-414.

Cho, T., Jun, S., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*(2), 193-228.

Coderre, E. L., Filippi, C. G., Newhouse, P. A., & Dumas, J. A. (2008). The Stroop effect in kana and kanji scripts in native Japanese speakers: An fMRI study. *Brain & Research, 107*(2), 124-132.

Cole, R. A. & Jakimik, J. (1980). How are syllables used to recognize words. *The Journal of the Acoustical Society of America, 67*(3), 965-970.

Collins, A. M. & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407-428.

Collins, A. M. & Quillian, M. R. (1970). Facilitating retrieval from semantic memory: The effect of repeating part of an inference.. *Acta Psychologica, 33*, 304-314.

Collins, A. M. & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. Gregg (Ed.), *Cognition and Learning*. New York, NY: Wiley.

Colombo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 226-234.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27*(3), 183-228.

Coltheart, M. (1998). Reading phonological recoding and deep analysis. In M. Coltheart, K. Patterson, & J. C. Marshal (Eds.), *Deep dyslexia* (pp. 197-226). London, England: Routledge & Kegan Paul.

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.),*The science of reading: A handbook* (pp. 6–23). Oxford, England: Blackwell Publishing.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 589–608.

Coltheart, M., Davelaar, E., Jonassen, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-556). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 757-772.

Coulmas, F. (1996). *The Blackwell encyclopedia of writing systems*. Cambridge, MA: Blackwell Publishing.

Daneman, M. & Reingold, E. (1993). What eye fixations tell us about phonological recoding during reading. *Canadian Journal of Experimental Psychology, 47*, 153-178.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition.

Journal of Experimental Psychology: Human Perception and Performance, 28(1), 218–244.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems.* Honolulu, HI: University of Hawaii Press.

Dijkstra, T., Koji, M., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language, 62*(3), 284-301.

Dimitropoulou, M., Duñabeitia, J.A. & Carreiras, M. (2011). Transliteration and transcription effects in biscriptal readers: The case of Greeklish. *Psychonomic Bulletin & Review, 18*(4), 729-735.

Doctor, E. A. & Coltheart, M. (1980). Children's use of phonological encoding when reading for meaning. *Memory & Cognition, 8*(3), 195-209.

Duchek, J. M. & Neely, J. H. (1989). A dissociative word-frequency×levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition, 17*(2), 148–162.

Duffy, S. A., Kambe, G., & Rayner, K. (2001). The effect of prior disambiguating context on the comprehension of ambiguous words: Evidence from eye movements. In D. S. Gorfein (Ed.), *Decade of behavior. On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 27-43). American Psychological Association.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*(4), 429-446.

Dufour, S. & Frauenfelder, U. H. (2010). Phonological neighbourhood effects in French spoken-word recognition. *Quarterly Journal of Experimental Psychology, 63*(2), 226-238.

Duñabeitia, J. A., Carreiras, M. & Perea, M. (2008). Are *coffee* and *toffee* served in a cup? Ortho-phonologically mediated associative priming. *The Quarterly Journal of Experimental Psychology, 61*(12), 1861-1872.

Eddington, C. M. & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. *Journal of Verbal Learning and Verbal Behavior, 22*, 417-436.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., van Daal, V. H. P., Polyzoe, N., Tsipa, M., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39*(4), 438-468.

Eulitz, C. & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16*(4), 577-583.

Fan, K. Y., Gao, J. L., & Ao, X. P. (1984). Pronunciation principles of Chinese characters and alphabetic script. *Chinese Character Reform, 3*, 23-27.

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.

Foss, D. J. (1970). Some effects of ambiguity upon sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 9*, 699-706.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales, & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam, Netherlands: 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: The MIT Press.

Forster, K. I. & Bednall, E. S. (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition, 4*(1), 53–61.

Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 498-514.

Frauenfelder, U. H. & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition, 25*(1-2), 1-20.

Friedrich, C. K., Eulitz, C. & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions, 2*(36).

Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1545-1559.

Frost, R. (1998). Towards a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*, 71-99.

Frost, R. (2005). Orthographic systems and skilled word recognition process in reading. In M. J. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272-295). Oxford, England: Blackwell Publishing.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 104–115.

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.

Gahl, S. & Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language, 89*, 162-178.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*(4), 789-806.

Georgiou, G. K., Parrilla, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100*(3), 566–580.

Gernsbacher, M. A. & St. John, M. F. (2001). Modeling suppression in lexical access. In D. S. Gorfein (Ed.), *Decade of behavior. On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 47-64). American Psychological Association.

Geva, E., & Wang, M. (2001). The development of basic reading skills in children: A cross-language perspective. *Annual Review of Applied Linguistics, 21*, 182–204.

Giraudo, H. & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes, 15*, 421-444.

Glucksberg, S., Kreuz, R. J., & Rho, S. H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 323-335.

Gollan, T., Forster, K., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*(5), 1122-1139.

Grainger, J. (1992). Orthographic neighborhoods and visual word recognition. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 131-146). Amsterdam, Netherlands: Elsevier.

Grainger, J. & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Trends in Neuroscience, 7*, 218-233.

Grainger, J. & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes, 13*(6), 601-623.

Grainger, J., Diependaele, K., Spineli, E., Ferrand, L., & Farioli, F. (2003). Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology, 29*, 1256-1269.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*(3), 518–565.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58A*(6), 981–998.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Seguí, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics, 51*(1), 49–56.

Bybee, J. L. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 119-141). New York, NY: Academic Press.

Grant, B. K. (1982). *A guide to Korean characters: Reading and writing Hangul and Hanja*. Elizabeth, NJ: Hollym International.

Green, D. W. (2011). Language control in different contexts: The behavioral ecology of bilingual speakers. *Frontiers in Psychology, 2*, 1-4P.

Grosjean, F. (1985). The bilingual as a competent but specific speaker–hearer. *Journal of Multilingual and Multicultural Development, 6*, 467-477.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. M. de Groot, & J. F. Kroll (Eds.), *Tutorials in bilingualism* (pp. 225-254). Mahwah, NJ: Lawrence Erlbaum Associates.

de Groot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22*, 417-436.

Gwilliams, L., Poeppel, D., Marantz, A., & Linzen, T. (2018). Phonological (un)certainty weights lexical activation. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics* (pp. 29-34). Salt Lake City, Utah.

Hammond, E. J., & Green, D. W. (1982). Detecting targets in letter and non-letter arrays. *Canadian Journal of Psychology/Revue canadienne de psychologie, 36*(1), 67–82.

Hansen, G. F. (2008). Visual word recognition in Arabic: Towards a language specific reading model. *Centre for Contemporary Middle East Studies' Working Paper Series No. 13*. Odense: University of Southern Denmark.

Harm, M. W. & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662–720.

Hasenäcker, J, Solaja, O., & Crepaldi, D. (2019). Food in the corner and money in the cashews: Semantic activation of embedded stems in the presence or absence of a morphological structure. *Psychonomic Bulletin & Review, 27*(1), 155-161.

Henderson, L. (1982). *Orthography and word recognition in reading*. London, England: Academic Press.

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development, 84*(5), 1668-1685.

Hino, Y. & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1331-1356.

Hino, Y., Kusunose, Y., Miyamura, S., & Lupker, S. J. (2017). Phonological-orthographic consistency for Japanese words and its impact on visual and auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 43*(1), 126–146.

Hirose, T. (1992). Recognition of Japanese kana words in priming tasks. *Perceptual and Motor Skills, 75*(3), 907-913.

Hoshino, N. & Kroll, J. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106*(1), 501-511.

Humphreys, G. W., Besner, D., & Quinlan, P. T. (1988). Event perception and the word repetition effect. *Journal of Experimental Psychology: General, 117*(1), 51–67.

Humphreys, G. W. & Rumiati, R. I. (1998). Agnosia without prosopagnosia or alexia: Evidence for stored visual memories specific to objects. *Cognitive Neuropsychology, 15*, 243-277.

Hutchinson, K. A. (2003). Is semantic priming due to association strength or feature overlap? A *micro*analytic review. *Psychonomic Bulletin & Review, 10*(4), 785-813.

Iwata, M. (1984). Kanji versus kana: Neuropsychological correlates of the Japanese writing system. *Trends in Neuroscience, 7*, 290-293.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120*(4), 358–394.

Jastrzembski, J. E. (1981). Multiple meaning, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 3*, 278-305.

Johnson, K. (1997). Speech perception without speaker normalizations: An exemplar model. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 144-166). San Diego, CA: Academic Press.

Johnson, N. & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26*, 240-346.

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* (pp. 511-514). Tokyo, Japan.

Joyce, T. (2002). Constituent-morpheme priming: Implications from the morphology of two-kanji compound words. *Japanese Psychological Research, 44*(2), 79-90.

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.

Jung. J & Yi. K. (2004). Morphological representation and processing of Korean words. In *Proceedings of the Annual Conference of Korean Society for Experimental Pscyhology* (pp. 89-95). Seoul, South Korea.

Kahn, D. (1976). *Syllable-based generalizations in English phonology.* New York, NY: Garland.

Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory & Cognition, 29*(2), 363-372.

Kang, H. & Simpson, G.B. (1996). Development of semantic and phonological priming in a shallow orthography. *Developmental Psychology, 32*, 860–866.

Kang, H. & Simpson, G. B. (2001). Local strategic control of information in visual word recognition. *Memory & Cognition, 29*, 648-655.

Katz, L & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In L. Katz & R. Frost (Eds.), *Orthography, phonology, morphology and meaning* (pp. 67-84). Amsterdam, Netherlands: Elsevier.

Kayamoto, Y., Yamada, J., & Takashima, H. (1998). The consistency of multiple-pronunciation effects in reading: The case of Japanese logographs. *Journal of Psycholinguistic Research, 27*, 619-637.

Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the timecourse of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 601-609.

Kess, J. F. & Miyamoto, T. (1999). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing.* Amsterdam, Netherlands: John Benjamins Publishing Company.

Kim, C.-H. (1983). The making of the Korean language. In the Korean National Commission for UNESCO (Ed.), *The Korean language* (pp. 121-127). Seoul, Korea: The Si-sa-yong-o-sa Publishers, Inc.

Kim, H.-G. (1983). Chinese characters and the Korean Language. In the Korean National Commission for UNESCO (Ed.), *The Korean language* (pp. 121-127). Seoul, Korea: The Si-sa-yong-o-sa Publishers, Inc.

Kim, J.-H. (2015) *Phonological development in Korean-English bilingual children* (Doctoral dissertation). The University of Auckland, Auckland, New Zealand.

Kim, H. & Kang, B. (1997). *Frequency analysis of Hangul usage.* Seoul, Korea: Korean Cultural Resarch Center, Korea University.

Kim, J. & Davis, C. (2001). Loss of rapid phonological recoding in reading Hanja, the logographic script of Korean. *Psychonomic Bulletin & Review, 8*(4), 785-790.

Kim, J. & Davis, C. (2002). Using Korean to investigate phonological priming effects without the influence of orthography. *Language and Cognitive Processes, 17*(6), 569-591.

Kim, J. & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language, 49*(4), 484-499.

Kim, J., Taft, M., & Davis, C. (2004). Orthographic-phonological links in the lexicon: When lexical and sublexical information conflict. *Reading and Writing: An Interdisciplinary Journal, 17*, 187-218.

Kim, S. Y., Yap, M. J., & Goh, W. D. (2018). The role of semantic transparency in visual word recognition of compound words: A megastudy approach. *Behavior Research Methods*, 1-11.

Kim-Renaud, Y.-K. (1997). *The Korean alphabet: Its history and structure.* Honolulu, HI: University of Hawaii Press.

Kim-Renaud, Y.-K. (2009). *Korean: An essential grammar.* New York, NY: Routledge.

Kiparsky, P. (1982). From cyclic to lexical phonology. In H. van der Hulst (Ed.), *The structure of phonological representation* (pp. 165-225). Dordrecht, Netherlands: Foris.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain Language, 81*, 205-223.

Kuperberg, G. R., Lakshmanan, B. M., Greve, D. N., & West, W. C. (2008). Task and semantic relationship influence both the polarity and localization of hemodynamic modulation during lexico-semantic processing. *Human Brain Mapping, 29*, 544-561.

Lahiri, A. (1999). Speech recognition with phonological features. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 715-718). San Francisco, California.

Lahiri, A. (2000). Phonology: Structure, representation and process. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 165-225). London, England: Psychology Press.

Lahiri, A. & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*, 245-294.

Lahiri, A. & Marslen-Wilson, W. D. (1992). Lexical processing and phonological representation. In R. D. Ladd & G. Docherty (Eds.), *Papers in Laboratory Phonology II: Gesture, segment, prosody* (pp. 229-254). Cambridge, England: Cambridge University Press.

Lahiri, A. & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 637-677). Berlin, Germany: Mouton de Gruyter.

Lahiri, A. & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics, 38*, 44-59.

Landerl, K. (2006). Reading acquisition in different orthographies: Evidence from direct comparison. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 513–530). London: Lawrence Erlbaum Associates.

Leinenger, M. & Rayner, K. (2013). Eye movements while reading biased homographs: Effects of prior encounter and biasing context on reducing the subordinate bias effect. *Journal of Cognitive Psychology, 25*(6), 665-681.

Ledyard, G. K. (1998). *The Korean language reform of 1446.* Seoul, Korea: Sin'gu Munhwasa.

Lee, H.-M. (1990). The letter formation principles of Hunmin-jông.ûm. In S.-S. Shin, D.-J. Lee, & H.-M. Lee (Eds.), *Understanding Humin-jŏng.ŭm* (pp. 201-220). Seoul, Korea: Hanshin Publishing Company.

Lee, K.-M. (1983). Foundations of Hunmin Chongum. In the Korean National Commission for UNESCO (Ed.), *The Korean language* (pp. 121-127). Seoul, Korea: The Si-sa-yong-o-sa Publishers, Inc.

Lee, I. & Ramsey, S. R. (2000). *The Korean language*. Albany, NY: State University of New York Press.

Lee, C., Lee, Y., & Kim, T. H. (2016). The exploration of the effects of word frequency and word length on Korean word recognition. *Journal of the Korea Academia-Industrial Cooperation Society, 17*(1), 54–61.

Leong, C. K. & Tamaoka, K. (1998). Cognitive processing of Chinese characters, words, sentences and Japanese kanji and kana: An introduction. *Reading and Writing: An Interdisciplinary Journal, 10*, 155-164.

Lesch, M. F. & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 285-294.

Levelt, W.J.M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: The MIT Press.

Lin, C. Y., Wang, M., & Ko, I. Y. (2018). The time course of activation of semantic and orthographic information in morphological decomposition by Korean adults and developing readers. *Frontiers in Communication, 3*(51), 1-18.

Liow, S. J. R., Green, D., & Tam, M. M. L. (1999). The development of visual search strategies in biscriptal readers. *International Journal of Bilingualism, 3*(4), 333-349.

Liu, Y. & Peng, D.-L. 1997. Meaning activation of Chinese compounds and its time course. In H.-C. Chen (Ed.), *Cognitive processing of Chinese and related Asian languages* (pp. 219-232). Hong Kong: The Chinese University of Hong Kong Press.

Locker, L., Simpson, G. B., & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Memory & Cognition, 31*(4), 505-515.

Longtin, C.-M., Segui, J., Hallé, P. A. (2003). Morphological priming in morphological relationship. *Language and Cognitive Processes, 18*(3), 313-334.

Lucas, M. (1987). Frequency effects on the processing of ambiguous words in sentence contexts. *Language and Speech, 30*(1), 25-46.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7*, 618-630.

Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1-36.

Luk, G. & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605-621.

Lukatela, G. & Turvey, M. T. (1991). Phonological access of the lexicon: Evidence from associative priming with pseudohomophones. *Journal of Experimental Psychology: Human Perception & Performance, 17*, 951-966.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676-703.

MacKay, D. G. (1966). To end ambiguous sentences. *Perception & Psychophysics, 1*, 426-436.

Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173-193.

Marslen-Wilson, W. D. (1984). Function and processing in spoken word recognition – a tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processing*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. D. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second sperlonga meeting* (pp. 187-210). Hove, England: Lawrence Earlbaum Associates.

Marslen-Wilson, W. D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1-71.

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*(1), 3-33.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Martin, S. E. (2006). *A reference grammar of Korean: A complete guide to the grammar and the history of the Korean language*. Tokyo, Japan: Tuttle Publishing.

Mason, R. A. & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research, 1146*(1), 115-127.

Masuda, H. & Joyce, T. (2018). Constituent-priming investigations of the morphological activation of Japanese compoud words. In H. K. Pae (Ed.), *Writing systems, reading processes, and cross-linguistic infleunces: Reflections from the Chinese, Japanese, and Korean languages* (pp. 221-244). Amsterdam, Netherlands: John Benjamins Publishing Company.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 375-407.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*, 375-407.

McClelland, J. L. & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114*, 159-188.

McDonald, S. A. & Shillcock, R. C (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing.*Language and Speech, 44*(3), 295-322.

McNamara, T. P. & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language, 27*, 545-559.

McNamara, T. P. (2005). *Essays in cognitive psychology.Semantic priming: Perspectives from memory and word recognition.* New York, NY: Psychology Press.

McQueen, J. M. & Cutler, A. (2001). Spoken word access processes an introduction. *Language and Cognitive Processes, 16*(5-6), 469-490.

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes, 10*(3-4), 309–331.

McRae, K. & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 558-572.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*, 99-130.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*(3), 283-312.

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.

Meyer, D. E. & Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. *Science, 192*, 27-33.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt, & S. Dornic (Eds.), *Attention and performance V* (pp. 98-118). New York, NY: Academic Press.

Millis, M. L. & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition, 17*, 141-147.

Mirman, D. & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 65-79.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language, 33*(2), 175-202.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*(2), 165–178.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 863-883.

Müller, O., Duñabeitia, J. A., & Carreiras, M. (2010). Orthographic and associative neighborhood density effects: What is shared, what is different?*Psychophysiology, 47*(3), 455–466.

Munson, B. & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*(5), 1048-1058.

Myung, J., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition, 98*(3), 223–243.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7-22.

Nakayama, M., Verdonschot, R., Sears, C., & Lupker, S. (2014). The masked cognate translation priming effect for different-script bilinguals is modulated

by the phonological similarity of cognate words: Further support for the phonological account. *Journal of Cognitive Psychology, 26*(7), 714-724.

Nam, K. (1995). *Korean word recognition: Comparison of Hangul and Hanja word recognition* Paper presented at the Conference on the Korean Experimental and Cognitive Psychological Association.

Nayernia, L., van de Vijver, R., & Indefrey, P. (2019). The influence of orthography on phonemic knowledge: An experimental investigation on German and Persian. *Journal of Psycholinguistic Research, 48*,(6) 1391-1406.

The National Institute of the Korean Language. (2008). *알기 쉽게 풀어 쓴 훈민정음: Written plainly so as to be understood by everyone Hunmin Jeongeum.* Seoul, Korea: Kungnip Kogŏwŏn

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 Inc.

Nelson, D. L., Bennett, D. J., Gee, N. R., Schreiber, T. A., & McKinney, V. M. (1993). Implicit memory: Effects of network size and interconnectivity on cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 747–764.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Norris, D. & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395.

Okano, K., Grainger, J., & Holcomb, P. J. (2013). An ERP investigation of visual word recognition in syllabary scripts. *Cognitive, Affective, & Behavioral Neuroscience, 13*(2), 390-404.

Onifer, W. & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9*(3), 225-236.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Park, C. (2009). Visual processing of Hangul, the Korean script. In C. Lee, G. Simpson, & Y. Kim (Eds.), *The handbook of East Asian psycholinguistics* (pp. 379-389). Cambridge, England: Cambridge University Press.

Park, C. (2015). *Hangeul: The Korean alphabet*. Seoul, Korea: Ewha Womans University Press.

Park, J. (1990). The Chinese phonological background of Humin-jŏng.ŭm. In P. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (pp. 67-82). New York, NY: Plenum Press.

Park, K.-S. (1995). Visual processes involved in *hangul* word recognition. *Korean Journal of Experimental and Cognitive Psychology, 7*, 61-78.

Park, K.-S. (1996). The role of phonology in *hangul* word recognition. *Korean Journal of Experimental and Cognitive Psychology, 8*, 25-44.

Park, K.-S. (1999). Is phonology obligatory in visual access to word meaning? Negative evidence for associative homophone priming in Korean word naming task. *Korean Journal of Experimental and Cognitive Psychology, 11*, 17-28.

Park, T. J. (1990b). Phonological encoding when reading Chinese-character-words in Korean. *Korean Journal of Experimental and Cognitive Psychology, 2*, 90-102.

Park, K. & Iyotsna, V. (1995). Lexical representation of script variation: Evidence from Korean biscriptals. In I. Taylor, & D. R. Olson (Eds.), *Scripts and literacy: Reading and learning to read alphabets, syllabaries and characters* (pp. 327-339). Dordrecht, Netherlands: Kluwer Academic.

Pechar, D., Zeelenberg, R., & Wagenmakers, E.-J. (2005). Enemies and friends in the neighborhood: Orthographic similarity effects in semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 121-128.

Peeters, D., Dijkstra, & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and EPR effects. *Journal of Memory and Language, 68*, 315-352.

Perfetti, C. A. & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first and second language literacy development* (pp. 13-38). London, England: Routledge.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 24–33.

Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). Transposed-letter priming effects for close versus distant transpositions. *Experimental Psychology, 55*(6), 384–393.

Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition, 62*(215), 223-240.

Perea, M., Marcet, A., Vergara-Martinez, M., & Gomez, P. (2016). On the dissociation of word/nonword: An evidence accumulation account. *Frontiers in Psychology, 7*(215), 1-9.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review, 15*(1), 161-167.

Pexman, P. M., Holyk, G. G., & Monfils, M. (2003). Number-of-features effects and semantic processing. *Memory & Cognition, 31*(6), 842-855.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review, 9*(3), 542-549.

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*(1), 56–115.

Prior, A. & Gollan, T. H. (2011) Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society, 17*(4), 682-691.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012) Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 71-99.

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*(4), 1142–1158.

Posner, M. I. & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pulleyblank, E. (1991). *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin.* Vancouver, Canada: UBC Press.

R Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition, 17*(5), 525-535.

Ramsey, S. R. (1997). The invention of the alphabet and the history of the Korean language. In Y.-K. Kim-Renaud (Ed.), *The Korean alphabet: Its history and structure* (pp. 131-143). Honolulu, HI: University of Hawaii Press.

Ranbom, L. R. & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*, 273-298.

Rayner, K. & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*, 191-201.

Rayner, K. & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*(5), 779-790.

Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language, 33*(4), 527-544.

Reetz, H. (1998). Automatic speech recognition with features. *Habilitationsschrift, Universität des Saarlandes.*

Reetz, H. (1999). Converting speech signals to phonological features. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1733-1736). San Francisco, California.

Reetz, H. & Kleinmann, A. (2008). Multi-subject hardware for experiment control and precise reaction time measurement. In M. J. Solè, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1489-1492). Barcelona, Spain.

Rhodes, S. & Donaldson, D. I. (2008). Association and not semantic relationships elicit the N400 effect: Electrophysiological evidence from an explicit language comprehension task. *Psychophysiology, 45*, 50-59.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*(1), 1-20.

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience, 15*(8), 1160-1175.

Rodd, J. (2004). When do leotards get their spots? Semantic activation of lexical neighbors in visual word recognition. *Psychonomic Bulletin & Review, 11*(3), 434-439.

Rodd, J., Gaskell, G. & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*, 245-266.

Rodd, J., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68*(2), 180-198.

Rossell, S. L. & Nobre, A. C. (2004). Semantic priming in different affective categories. *Emotion, 4*, 354-363.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*, 487-494.

Rusted, J. (1989). Developmental differences in the influence of orthographic and phonological information in visual word recognition. *British Journal of Developmental Psychology, 7*(1), 73-81.

Sakurai, Y. Yagishita, A., Goto, Y., Ohtsu, H., & Mannen, T. (1991). Fusiform type alexia: Pure alexia for words in contrast to posterior occipital type pure alexia for letters. *Journal of Neurological Science, 247*, 81-92.

Salasoo, A. & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24*(2), 210-231.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*(1), 51–89.

Sampson, G. (1985). *Writing systems: A linguistic introduction.* London, England: Hutchinson.

Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes, 9*(3), 227-269.

Schreuder, R., & Baayen, R. H. (1995). Modelling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131-154). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Schreuder, R., Flores D'Arcais, G. B., & Glazenborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychological Research, 45*(4), 339–354.

Schriefers, H., Zwitserlood, P., & Roelofs, A. (1991). Morphological decomposition vs. left-to-right matching. *Journal of Memory and Language, 30*, 26-47.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14*(5), 238-242.

Seidenberg M. S. & Gonnerman L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4*(9), 353-361.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523–568.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*(4), 489-537.

Sereno, S. C., O'Donnell, P. J., & Rayner, K. (1986). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception & Performance, 32*(2), 335-350.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.

Shelton, J. R. & Martin, R. C. (1992). How semantic is automatic semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(6), 1191-1210.

Shen, D. & Forster, K. I. (1999). Masked phonological priming in reading Chinese words depends on the task. *Language and Cognitive Processes, 14*(5/6), 429-459.

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24-49). Cambridge, MA: The MIT Press.

Shin, J., Kiaer, J., & Cha, J. (2013). *The sounds of Korean.* Cambridge, England: Cambridge University Press.

Shin, S.-S., Lee, D.-J., & Lee, H.-M. (1990). *Understanding Humin-jŏng.ŭm.* Seoul, Korea: Hanshin Publishing Company.

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning & Verbal Behavior, 20*(1), 120-136.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 359-374). New York, NY: Academic Press.

Simpson, G. B. & Kang, H. (1994). The flexible use of phonological information in word recognition in Korean. *Journal of Memory and Language, 33*, 319-331.

Simpson, G. B. & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language, 30*(6), 627-643.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 8*(3), 214-241.

Sohn, H.-M. (1994). *Korean: Descriptive grammars*. London, England: Routledge.

Sohn, H.-M. (1999). *The Korean language*. Cambridge, England: Cambridge University Press.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281-284.

Spruyt, A., Hermans, D., De Houwer, J., Vandromme, H., & Eelen, P. (2007). On the nature of the affective priming effect: Effects of stimulus onset asynchrony and congruency proportion in naming and evaluative categorization. *Memory & Cognition, 35*(1), 95-106.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-659.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27*(3), 324-340.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research, 49*(2-3), 161-167.

Tabossi, P. & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing sytems: Italian and English. *Memory & Cognition, 20*, 303-313.

Tabossi, P. & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language, 32*(3), 359-372.

Taft, M. (1985). The decoding of words in lexical access: A review of the morphographic approach. In D. Besner, T. Waller, & G. Mackinnon (Eds.), *Reading research: Advances in theory and practice (volume 5)*. London, England: Academic Press.

Taft, M. (1991). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9*(3), 271-294.

Taft, M. (1994). *Reading and the mental lexicon.* Hove, England: Lawrence Earlbaum Associates.

Taft, M. & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 638-647.

Taft, M., Liu, Y., & Zhu, X. (1999). Morphemic processing in reading Chinese. In J. Wang, A. W. Inhoff, & H.-C., Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 91-113). Mahwah, NJ: Lawrence Erlbaum Associates.

Taft, M. & Nguyen-Hoan, M. (2010). A sticky stick? The locus morphological representation in the lexicon. *Language and Cognitive Processes, 25*(2), 277-296.

Taft, M. & Zhu, X. (1995). The representation of bound morphemes in the lexicon: A Chinese study. In L. Feldman (Ed.), *Morphological aspects of language* (pp. 293-316). Hillsdale, NJ: Lawrence Erlbaum Associates.

Taft, M. & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 761-775.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 427-440.

Tan, L. H., Hoosain, R., & Peng, D.-l. (1995). Role of early presemantic phonological code in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 43–54.

Tavassoli, N. T. & Han, J. K. (2001). Scripted thought: Processing Korean Hancha and Hangul in a multimedia context. *Journal of Consumer Research, 28*(3), 482-493.

Taylor, I. (1997). Psycholinguistic reasons for keeping Chinese characters in Korean and Japanese. In H.-C. Chen (Ed.), *Cognitive processing of Chinese and related Asian languages* (pp. 299-319). Hong Kong: The Chinese University of Hong Kong Press.

Taylor, I. (1980). The Korean writing system: An alphabet? A syllabary? A logography? In P. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (pp. 67-82). New York, NY: Plenum Press.

Taylor, I. & Olson, D. R. (1995). *Scripts and literacy: Reading and learning to read alphabets, syllabaries and characters.* Norwell, MA: Kluwer Academic Publishers.

Taylor, I. & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean, and Japanese*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Taylor, J. R. (2005). On the perceptual basis of semantic memory: Representation, process, and attentional control revealed by behavior and event-related brain potentials. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(5-B), 2850.

The Korea Foundation (2010). *Hangeul: Korea's unique alphabet*. Seoul, Korea: Seoul Selection.

Thierry, G. & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *PNAS, 104*(30), 12530-12535.

Traxler, M. J. (2012). *Introduction to psycholinguistics: Understanding language science*. Chichester, England: Wiley-Blackwell.

Tsang, Y. & Chen, H. (2010). Morphemic ambiguity resolution in Chinese: Activation of the subordinate meaning with a prior dominant-biased context. *Psychonomic Bulletin & Review, 17*(6), 875-881.

Tsang, Y. & Chen, H. (2013a). Early morphological processing is sensitive to morphemic meanings: Evidence from processing ambiguous morphemes. *Journal of Memory and Language, 68*(3), 223-239.

Tsang, Y. & Chen, H. (2013b). Morpho-semantic processing in word recognition: Evidence from balanced and biased ambiguous morphemes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1990-2001.

Twilley, L. C. & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin & Review, 7*(1), 49-82.

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics, 35*(5), 417-427.

Tyler, L. K., Marslen-Wilson, W. D., Rentoul, J., & Hanney, P. (1988). Continuous and discontinuous access in spoken word-recognition: The role of derivational prefixes. *Journal of Memory and Language, 27*(4), 368-381.

Verdonschot, R. G., La Heij, W., & Schiller, N. O. (2010). Semantic context effects when naming Japanese kanji, but not Chinese hànzì. *Cognition, 115*, 512-518.

Verdonscot, R. G. (2011). *Word processes in languages using non-alphabetic scripts: The cases of Japanese and Chinese* (Doctoral dissertation, Utrecht University, Utrecht, Netherlands.

Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W., & Schiller, N. O. (2011). The functional unit of Japanese word naming: Evidence from masking priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1458-1473.

Verdonschot, R. G., La Heij, W., Paolieri, D., Zhang, Q. & Schiller, N. O. (2011). Homophonic context effects when naming Japanese kanji: Evidence for processing costs? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 64*, 1836-1849.

Verdonschot, R. G., La Heij, W., Tamaoka, W., Kiyama, S., You, W. P., & Schiller, N. O. (2013). The multiple pronunciations of Japanese kanji: A masked priming investigation. *The Quarterly Journal of Experimental Psychology, 66*(10), 2023-2038.

Vitevitch, M. S. & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*(4), 325-329.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*(1-2), 306–311.

Vitevitch, M. S. & Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders, 3*(1), 64-73.

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition, 35*, 938-952.

Vu, H., Kellas, G. & Paul, S. T. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Memory & Cognition, 26*(5), 979-1001.

Wagenmakers, E.-J., Van der Mass, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3-22.

Wagner, A. D., Desmond, J. E., Demb, J. B., Glover, G. H., & Gabrieli, J. D. E. (1997). Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. *Journal of Cognitive Neuroscience, 9*(6), 714-726.

Weber, A. & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science,3*(3), 1-15.

Wheeldon, L. & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language, 90*, 401-412.

Woong, H. (1983). Development of the Korean language. In the Korean National Commission for UNESCO (Ed.), *The Korean language* (pp. 121-127). Seoul, Korea: The Si-sa-yong-o-sa Publishers, Inc.

Wurm, L. H. (1997). Auditory processing of prefixed English words is both continuous and decompositional. *Journal of Memory and Language,37*(3), 438-461.

Wydell, T. N., Butterworth, B. L., & Patterson, K. E. (1995). The inconsistency of consistency effects in reading: Are there consistency effects in Kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1155-1168.

Yamada, J. (1992). Why are Kana words named faster than Kanji words? *Brain and Language, 43*, 682-693.

Yang, S., Yang, H., & Hartanto, A. (2019). The effects of script variation, literacy skills, and immersion experience on executive attention: A comparison of matched monoscriptal and biscriptal bilinguals. *Bilingualism: Language and Cognition, 22*(1), 142-156.

Yap, M. J., Daragh, E. S., Balota, D., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 597-613.

Yates, M., Locker, L., Jr., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review, 11*(3), 452–457.

Yamada, J., Mitarai, Y., & Yoshida, T. (1991). Kanji words are easier to identify than katakana words. *Psychological Research, 53*(2), 136-141.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (CAN) prime furiously. *Psychological Science, 23*(4), 364–369.

Yi, K. (1993). Effect of orthography on mental lexicon structure and search during the Korean word identification. *Korean Journal of Experimental and Cognitive Psychology, 5*, 26-39.

Yi, K. (1996). Phonological rules in oral reading of Korean. *Korean Journal of Experimental and Cognitive Psychology, 8*, 1-23.

Yi, K. (1999). Representation and processing of Hanja compound words In *Proceedings of the Annual Conference of Korean Society for Experimental Pscyhology* (pp. 73-79). Seoul, South Korea.

Yi, K. (2009). Morphological representation and processing of Sino-Korean words. In C. Lee, Y. Kim, & P. Li (Eds.), *The handbook of East Asian*

psycholinguistics: Volume III Korean (pp. 398-408). Cambridge, England: Cambridge University Press.

Yi, K. & Yi, I. (1999). Morphological processing in Korean word recognition. *Korean Journal of Experimental Cognitive Psychology, 11*, 77-91.

Yoon, J. (2017, January 8). 초등 교과서 한자 병기 필요할까. *학국교육신문*, Retrieved from http://www.hangyo.com/news/article.html?no=79543.

Yoon H. W. & Lim J. (2011) Differential brain activity in reading Hangul and Hanja in Korean language. In T. Kim, H. Adeli, R. J. Robles, M. Balitanas (Eds.), *Ubiquitous computing and multimedia applications. UCMA 2011. Communications in computer and information science, Vol 151* (pp. 52-59). Heidelberg, Germany: Springer.

Zhou, H., Chen, B., Yang, M., & Dunlap, S. (2010). Language nonselective access to phonological representations: Evidence from Chinese-English bilinguals. *The Quarterly Journal of Experimental Psychology, 63*(10), 2051-2066.

Zhu, X. & Taft, M. (1994). The influence of perceptual experience on Chinese character processing. In H. W. Chang, J. T. Huang, C. W. Hue & O. J. L. Tzeng (Eds.), *Advances in the study of Chinese language processing* (pp. 85-99). Taipei, Taiwan: Department of Psychology, National Taiwan University.

Ziegler, J.C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21*(4), 551-559.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 547–559.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*(1), 25-64.