

Kim, Y., S., Kotzor, & A. Lahiri. (in press). Is Hanja represented in the Korean mental lexicon?: Encoding cross-script semantic cohorts in the representation of Sino-Korean. *Lingua*.

**Manuscript accepted for publication at *Lingua* (date of acceptance: 15 May 2021)**

© 2021, Elsevier B. V. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: [10.1016/j.lingua.2021.103128](https://doi.org/10.1016/j.lingua.2021.103128).

**Is Hanja represented in the Korean mental lexicon?:  
Encoding cross-script semantic cohorts in the representation of Sino-Korean**

Yoolim Kim<sup>a,b,\*</sup>, Sandra Kotzor<sup>b,c</sup>, & Aditi Lahiri<sup>b</sup>

<sup>a</sup>*Minds and Traditions Research Group, Max Planck Institute for the Science of Human  
History, Jena, Germany*

<sup>b</sup>*Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, United  
Kingdom*

<sup>c</sup>*School of Education, Oxford Brookes University, Oxford, United Kingdom*

Correspondence should be addressed to:

Yoolim Kim

Max Planck Institute for the Science of Human History

Kahlaische Strasse 10

07745 Jena

Germany.

Email: [ykim@shh.mpg.de](mailto:ykim@shh.mpg.de)

## **Abstract**

Korean can be transcribed in two different scripts, one alphabetic (Hangul) and one logographic (Hanja). How does the mental lexicon represent the contributions of multiple scripts? Hangul's highly transparent one-to-one relationship between spellings and sounds creates homophones in spoken Korean that are also homographs in Hangul, which can only be disambiguated through Hanja. We thus tested whether native speakers encoded the semantic contributions of the different Hanja characters sharing the same homographic form in Hangul in their mental representation of Sino-Korean. Is processing modulated by the number of available meanings, that is, the size of the semantic cohort? In two cross-modal lexical decision tasks with semantic priming, participants were presented with auditory primes that were either syllables (Experiment 1) or full Sino-Korean words (Experiment 2), followed by visual Sino-Korean full word targets. In Experiment 1, reaction times were not significantly modulated by the size of the semantic cohort. However, in Experiment 2, we observed significantly faster reaction times for targets preceded by primes with larger semantic cohorts. We discuss these findings in relation to the structure of the mental lexicon for bi-scriptal languages and the representation of semantic cohorts across different scripts.

**Key words:** orthographic processing, bi-scriptal lexicon, cross-script semantic representation, mediated semantic priming, visual word recognition

## 1. Introduction

The knowledge that we as speakers have of the words we know is represented in some way in our mental lexicon, but exactly what lexical information is integrated in our mental representations remains an ongoing question. For languages with multiple writing systems, their mental representations must integrate the contributions of more than one script. How does the mental lexicon represent the relationship between multiple scripts? The Korean language is transcribed in two different scripts, one alphabetic (and native to the language), Hangul, and one logographic (and non-native to the language), Hanja. Hangul is characterized by a highly transparent one-to-one relationship between grapheme and phoneme, which means that it is very often the case that homophones in the spoken language remain homophones in Hangul. That is, homophones are realized as homographs in written form. They can however be disambiguated through Hanja. The relationship between Hangul and Hanja therefore has much to bear on the way in which semantic meaning is accessed.

Exactly how information of multiple orthographic inputs is coded has gained increasing theoretical interest in recent lines of research, although primarily in bilinguals of languages with completely different orthographies (see Hoshino & Kroll, 2008, for Japanese-English; Spivey & Marian, 1999; Marian & Spivey, 2003, for Russian-English; 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). The question remains outstanding for different writing systems belonging to the same language.

Korean is particularly well suited for such an investigation given its use of two different systems: *Hangul* and *Hanja*. Hangul refers to the native alphabet, while Hanja are borrowed Chinese characters incorporated into the Korean language according to native pronunciation. Structurally, the systems vary in how the characters parcel information: Hangul is an alphabetic syllabary, where each symbol represents a sound, while Hanja is a logography,

where symbols represent meaning. The interrelated nature of Hanja and Hangul is most borne out in the Sino-Korean sub-lexicon, which presently comprises more than half of the Korean lexicon. Sino-Korean refers to words of Chinese origin and are privileged relative to pure Korean words. That is, while all words can be written in Hangul, not all can be written in Hanja. Sino-Korean words, on the other hand, can be written in both, whereas native Korean words can only be written in Hangul. For example, ‘picture’ can be written either as **도화** [to.hoa] or **그림** [kw.lim]. The latter is pure Korean, and the former is Sino-Korean, which means that it can also be written in Hanja – **圖書** – the first character corresponding to **도** [to], meaning ‘picture’, and the second to **화** [hoa], meaning ‘drawing’. Therefore, Sino-Korean, unlike pure Korean, maintains more than one orthographic representation.

A further complication arises as a result of Hangul’s high transparency between phonemes and graphemes, characterized by a one-to-one relationship between spellings and sounds. Due to such regularities, there are many instances of homophones in the spoken language remaining as homophones in Hangul, and thus homographs in written form. That Sino-Korean uses only about 440 distinct syllable combinations, out of a possible 2,000 (Taylor & Taylor, 2014)<sup>1</sup>, means repeated use of the same syllable to represent different meanings. This proves rather consequential for Sino-Korean, as the only way homographs can be disambiguated is through the Hanja. Thus, a multitude of different Hanja can potentially be tied to the Hangul representation of a single Sino-Korean form. Figure 1 illustrates one such example where the same syllable is used to express two different meanings. Two possible Hanja are shown for the morpheme **학** [hak], to which each contributes a different meaning, ‘learning’ and ‘harm’. Lexical instantiations then bear semantic relatedness to the Hanja, as also exemplified in Figure 1.

---

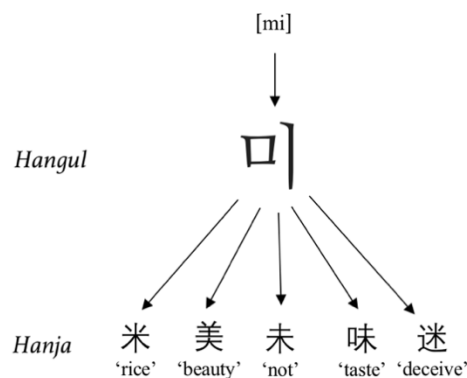
<sup>1</sup> This excludes certain CVCs and all CVCCS; the total number of possible distinct syllables is 11,172.



**Figure 1.** The Sino-Korean morpheme 학 [hak] and its set of two Hanja and the corresponding full-word lexicalizations. The colors indicate the semantic relatedness across the scripts, from the semantics of the Hanja to their realization in full words, via the shared syllable 학. The full-word lexicalization on the left ('school') evokes the Hanja meaning 'learning', while the lexicalization on the right ('abuse') evokes the meaning 'harm'.

Although a different Hanja and its semantics are represented, the form and the phonetic shape of 학 are identical in both lexical instantiations.

The number of Hanja can vary vastly from morpheme to morpheme. Figure 2 illustrates a morpheme for which there are five available Hanja.



**Figure 2.** The Sino-Korean morpheme 미 [mi] and its larger set of Hanja, creating a larger set of meanings attributable to the shared syllable representation 미.

The morphemes 학 [hak] and 미 [mi] underscore the difference in the number of attributable meanings due to the difference in the number of available Hanja which share the Hangul representation. The emphasis in comparing the two is not on the precise number of available

meanings, but rather on the idea that, because the number of Hanja a Sino-Korean morpheme can have can vary widely, so too can the number of available semantic meanings. The collection of Hanja that share and contribute meaning to the Hangeul representation of a Sino-Korean morpheme thus appear to form a semantic cohort of varying size that cuts across the two scripts. The question emerges how exactly this additional layer of 'hidden' complexity between the two orthographies affects semantic processing and the structure of the mental lexicon. We therefore investigate the following: (1) can the automaticity with which phonology activates orthography be extended to multiple orthographies, and if so (2) can the various meanings of available Hanja be co-activated simultaneously, and (3) is lexical processing affected by more or fewer competing meanings?

The properties of a script and their effects on lexical processing also bring into question the assumptions underlying cross-modal priming paradigms. Within these paradigms, participants make a decision on a visual stimulus based on auditory input. This paradigm necessarily assumes a link between phonological and orthographic representations, and the activation of one after the other. That is, when a word is heard, the corresponding orthographic representation is also activated, or at least mediated via the conceptual dimension of the lexical entry. The orthographic and phonological representations then activate the corresponding semantic representation, another key assumption underpinning auditory lexical processing. The observed priming is then taken to indicate lexical activation (Bölte & Coenen, 2002). We now ask how this might hold for languages with more than one script, activating potentially more than one orthographic representation to which semantic representations may be linked.

## **2. Hanja and Hangeul during processing**

Semantic activation from cross-modal priming has so far been shown in a number of different languages, with varying orthographies, from Latinate (Marslen-Wilson, Tyler, Waksler, & Older, 1994) to Semitic (Feldman, Frost, & Pnini, 1994) to Sinitic (Zhou, Marslen-Wilson, Taft, & Shu, 1999), but the corresponding literature on bi-scriptal languages, such as Korean, remains

limited. It is unclear whether for such languages, hearing auditory stimuli will activate just one orthographic representation, or both. The features of written Korean make it an ideal testbed for such an assessment.

Korean's duality of writing systems arises from the language's early use of Chinese logography (known as Hanja) before the innovation of its own native script (Hangul). Use of such characters resulted in numerous Chinese loans into Korean, which were borrowed over the course of a thousand years until the early twentieth century, when Hangul was officially promulgated (Lee, 1983; Kim, 1983)<sup>2</sup>. This means that not only has the number of borrowed forms been vast, but the influence of Chinese loans has also become deeply entrenched in the Korean lexicon. This influence can be directly observed in the current composition of the Korean lexicon, which comprises primarily Sino-Korean words (>60%). Despite the dominance of Hangul, schools still teach some Hanja, and they appear in different printed materials. The experiments presented here focus exclusively on Sino-Korean, and all participants had received a minimum of two years of education for Hanja. Only a limited number of studies have pursued processing differences between pure Korean and Sino-Korean. In a series of experiments involving a naming test and a lexical decision task, Yi (2003) investigated whether the origin of a word, i.e., whether it is of Chinese or of pure Korean origin, affected lexical access. He tested native Korean words, Sino-Korean words and loanwords, and found that of the three, in both paradigms, native Korean words were consistently recognized the fastest, followed by Sino-Korean words. Yi found that Korean readers were able to categorize words written in Hangul as either Sino-Korean or native Korean with 90% accuracy, despite individual differences in levels of proficiency in use of Hanja. These findings suggest that Korean readers are clearly able to intuit whether a word presented in Hangul is of Sino-Korean or pure Korean origin. Additional results from An (2010, 2011)<sup>3</sup> also suggest speakers' general sensitivity to the semantics of individual constituents in their processing of

---

<sup>2</sup> Kim-Renaud (1997) and Shin et al. (1990) both provide a comprehensive historical review on the use of Hanja and the subsequent emergence of Hangul.

<sup>3</sup> An (2011) builds on earlier findings from Song (1992), which offers close examination of the semantic and syllabic structure of Sino-Korean words.



Sino-Korean words.<sup>4</sup> These findings offer some preliminary insights, but exactly what lexical information is enabling these intuitions has yet to be fully elucidated. What is clear, however, is that information which is integral to readers' ability to discern the origin of a word is encoded within the lexical entry of a Korean word, and is clearly activated by orthographic representation alone.

As readers gain proficiency in a new writing system with which they can express the words they know, they form additional orthographic representations as a result (Chang, Monaghan, & Welbourne, 2019; Harm & Seidenberg, 2004; Heuttig & Mishra, 2014). For Sino-Korean, a number of different relationships play a part to connect this information, whether it is between the phonology (e.g., [mi]) and the semantics ('beauty'), the orthography (e.g., 미) and the semantics ('beauty'), or the orthography (미) and the phonology ([mi]). In a study investigating morphological processing in Sino-Korean, Yi (2009) contends that increased exposure to Sino-Korean words, or Hanja, would not only create new links, but also reinforce existing links between Sino-Korean words and their constituent morphemes. Other studies (Yi & Yi, 1999; Yi, 1999a, 1999b) targeting morphological processing in Sino-Korean reveal mixed results for morphological facilitation between morphologically related items. In an earlier study, for example, Yi and Yi (1999) found that lexical entries for Sino-Korean could be activated through constituent morphemes, i.e., through the Hanja. These findings varied according to lag duration (i.e., the number of intervening items), that is, in the long-lag condition, they found clear evidence of morphological facilitation but no orthographic inhibition, while in the short-lag condition, they found orthographic inhibition in the orthographically related condition, but no evidence of morphological facilitation in the morphologically related condition.

What remains a potential confound in these results is the issue of form overlap. Yi and Yi (1999) employed a lexical decision task with three different priming conditions:

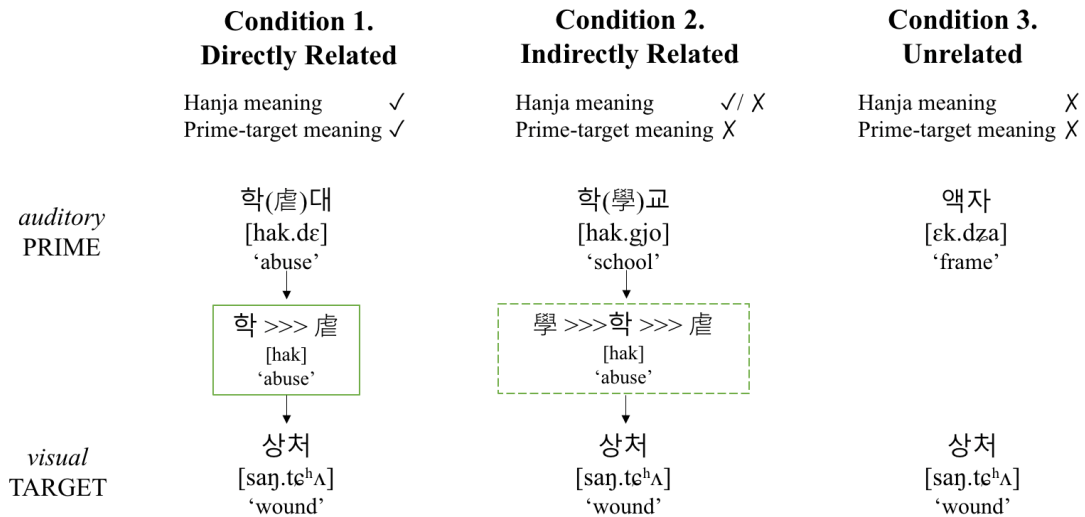
---

<sup>4</sup>An anonymous reviewer aptly pointed out some of the important insights Lee's (1972) examination of boundary phenomena in Korean offer to our discussion. Lee makes note of an interesting type of error children often make in Sino-Korean words, which he attributes to their lack of knowledge of the internal syllabic structure of constituent morphemes in Sino-Korean. He notes that certain Sino-Korean monosyllabic bound morphemes (e.g., 토 [tʰo]) require either being made into a light verb construction using 하 [ha] ('do') (토하다 [tʰo.ha.ta], 'to vomit'), or a noun by adding another Sino-Korean morpheme (구토 [ku.tʰo], 'vomit'). Lee reports that children, unaware of these constraints, mistakenly use 토 [tʰo] on its own.

morphologically related, orthographically related, and unrelated. All three primes were followed by the target 반항 [pan.han], meaning ‘resistance’, its Hanja representation, 反抗. With the exception of the unrelated condition, both morphologically and orthographically related conditions contained within the primes the form 반, creating an overlap in form between prime and target. This raises potential questions regarding whether observed priming effects are the result of morphological or orthographic relatedness, or simply due to form overlap alone. In the morphologically related condition, primes contained a constituent morpheme that shared the same Hanja representation as in the target, e.g., 반칙 [ban.tɕʰik], meaning ‘violation’, contains the morpheme 반, for which the Hanja representation is the same as in 반항 [pan.han] (i.e., 反). In the orthographically related condition, on the other hand, the two morphemes did not share the same Hanja representation, e.g., 반장 [ban.tɕan], meaning ‘captain’, contains the same morpheme 반, but its Hanja representation is different, 班. The sharing of constituent morphemes between prime and target necessarily means orthographic overlap, the processing of which may potentially precede any role that the morphology has to play at the level of Hanja representation, again obscuring proper attribution of results. If there is overlap in orthographic form at the level of Hangul representation, we cannot conclusively attribute observed priming to the role of Hanja, or morphological facilitation (i.e., 반칙 [ban.tɕʰik], 反則 – 반항 [pan.han], 反抗) in the view of Yi and Yi.

In order to properly understand the role of Hanja and the effects emerging across the two scripts, it seems important to ensure that all such confounds are eliminated. We therefore propose the paradigm in Figure 3, which enables us to probe the role of Hanja with no issue of form overlap. We can now properly investigate whether a Sino-Korean word comprising the semantic contribution of one Hanja can activate a word that is semantically related to the contribution of another Hanja sharing the same Hangul representation. Figure 3 shows the paradigm in use for the morpheme 학, which has a number of different Hanja, each with its own meaning. In this example, we exploit the Hanja 學 and 虐, which mean ‘learning’ and

'harm', respectively. Both share the Hangul representation, 학 [hak]. Each of the primes in the two relatedness conditions instantiates one of the two Hanja. 학교 [hak.gjo] ('school') instantiates the Hanja, 學 ('learning'), and 학대 [hak.dɛ] ('abuse') instantiates the Hanja, 虐 ('abuse').



**Figure 3.** The proposed paradigm which involves use of semantic associates to avoid potential confounds of orthographically alike constituent morphemes.

In the Directly Related condition, both prime and target are semantically related, the meaning presented in the target ('wound') matching the Hanja representation of 학 as presented in the prime, 학대 [hak.dɛ] ('abuse'), i.e., 虐 ('abuse'). In the Indirectly Related condition, primes and targets, while not directly related, are related by way of the mediating role of Hanja. This time the meaning presented in the target ('wound') does not match the Hanja representation of 학 as instantiated in the prime, 학교 [hak.gjo], i.e., 學 ('learning'), but the shared representation of 학 may potentially enable mediated access to other Hanja.

In using this design, not only do we avoid the repeated use of 학 [hak], but also eliminate possible confounds between morphological and semantic priming effects. There is no form or semantic overlap between the two words 학교 [hak.gjo] and 상처 [saŋ.ts<sup>h</sup>ʌ]; thus, any resulting facilitation is necessarily governed by activation of the mediating representation of the secondary Hanja character. This paradigm allows us to test the hypothesis that lexical

access is viable from the meaning of one Hanja to another and is mediated by the shared orthographic Hangul representation (i.e., 학), but that effects may be modulated by the number of competing Hanja within a given semantic cohort. We predict that for adult Korean speakers who have some knowledge of Hanja by virtue of being native speakers of Korean, and thus knowledge of both scripts, successful lexical processing of Sino-Korean will rely on co-activating orthographic representations. This notion of co-activation has been shown before by the literature on non-selective access of homophone meanings. These studies have shown that homophones presented in isolation activate a non-selective set of meanings (Swinney, 1979; Swinney, Onifer, Prather, & Hirshkowitz, 1979). Swinney (1979), in a cross-modal associative-priming study, observed that both meanings of homophones like *bug* (e.g., 'insect', 'spying') were activated at the offset of the word, regardless of whether there was any preceding context which specified the interpretation in one way or another. Although both are initially activated, soon after only the one which is context appropriate remains active. Further studies with different experimental approaches have corroborated Swinney's (1979) findings (Lucas, 1987; Oden & Spira, 1983; Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982).

In the following, we report on the findings from two cross-modal lexical decision tasks with semantic priming, one of which is as shown in Figure 3. In Experiment 1, we first investigate whether the Hangul representation of a Sino-Korean syllable, as a fragment, alone can trigger priming of targets semantically related to any one of the available Hanja and its meaning, and whether such effects are modulated by the size of the cohort. Experiment 2 then explores all three conditions, as presented in Figure 3, with full words, rather than syllable fragments, with the same emphasis on the effects of cohort size on lexical access.

### **3. Experiment 1: Cross-modal fragment priming**

#### **3.1. Method**

##### **3.1.1. Participants**

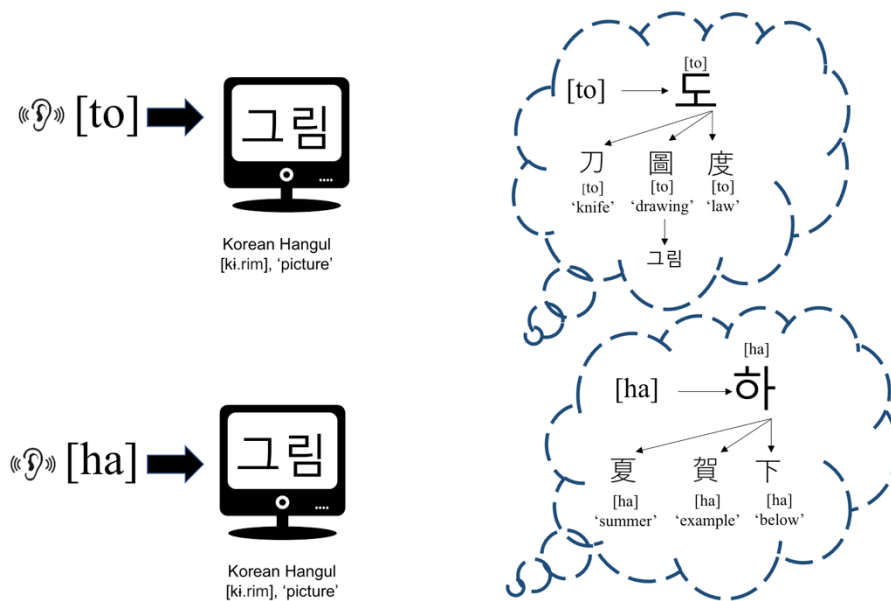
A total of 134 Korean native speakers participated in this experiment, all of whom were students at one of three different universities in Seoul, South Korea (Hansung University, Kyung Hee University and Sookmyung Women's University). All participants had normal or corrected-to-normal vision with no history of neurological disorders, and no reading impairments. They provided consent for participation and were reimbursed accordingly for their time. At the time of testing, all participants had also received a minimum of two years of classes for Hanja in school.

##### **3.1.2. Materials and design**

There currently exists a list of Hanja schools are expected to teach, which is set by the Korean Ministry of Education. These characters are considered to be recognizable to the average native Korean speaker. Using this list, we created 72 prime-target pairs. All prime stimuli were monosyllabic Korean morphemes (i.e., fragments), and all target stimuli were disyllabic Korean words, either pure Korean or Sino-Korean. All primes in the Related condition were Sino-Korean, while primes in the Unrelated condition were either pure Korean or Sino-Korean. It was difficult to create targets that were exclusively pure Korean or Sino-Korean, given that they had to be semantically related to one of the meanings the Hanja they heard evoked, while not containing that very morpheme in the target item itself. Similarly, it was difficult to create semantically unrelated primes which were exclusively pure Korean or Sino-Korean given the constraints of the language.

As Figure 4 shows with two Sino-Korean fragments, one which is related (top) and which is not (bottom), primes were presented auditorily, after which targets were presented visually in Hangul. The idea is that the phonological representation of the fragment evokes the corresponding orthographic representations in both Hangul and Hanja, and that the Hanja representations also facilitate access to their semantics. Of the cohort that is activated upon

hearing the fragment, one will match, or be semantically related to the meaning of the target item, as is demonstrated in the top part of Figure 4. Semantically unrelated primes, on the other hand, would activate a cohort of meanings, none of which would be immediately semantically related to the target item, and thus would not help to facilitate activation, as shown in the bottom part of Figure 4.



**Figure 4.** Illustration of experiment and possible pathway of activation from prime to target.

Each target is preceded by two primes, one which is semantically **Related**, and one which is **Unrelated**. In order to create a balanced design in the number of word and nonword stimuli, 72 additional prime-target pairs with nonword targets were included, resulting in a total of 144 items divided into 12 blocks of six trials. Each participant only saw each target in each condition once, but were exposed to all two conditions (Unrelated and Related).

It should be noted that these stimuli were all derived from the full-word stimuli used in Experiment 2. The auditory stimuli for Experiment 2, i.e., the full words, were recorded in a sound-attenuated room by a female native speaker of Standard Seoul Korean, using the software Audacity with a professional quality USB microphone (Røde NT-USB) at a sampling rate of 44.1 kHz. The words were embedded in a list of unrelated words in a randomized order. The auditory stimuli were then extracted, at the nearest zero crossings, using the acoustic analysis software PRAAT (Boersma & Weenick, 2015). The volume of all items was equalized.

All individual fragments were then excised from these recordings and extracted using the same speech analysis software. The fragment in question was excised from both relatedness conditions, as shown in Figure 3, and the version with the fewest audio clicks was selected for use.

When devising the full-word stimuli, the frequencies of the target items were matched to the greatest extent possible given the constraints of the design's dependence on specific Hangul-to-Hanja relationships to fulfil each of the relatedness conditions. Frequency measurements were taken from a database from the National Institute for Korean Language. Those same targets were then used for Experiment 1. No frequency counts are available for individual Sino-Korean morpho-syllables, nor is there any frequency measure for Hanja as they appear in text. We therefore conducted a separate rating task to ensure appropriate matching for test and control items as well as to confirm our judgements on the semantic relatedness between primes and targets. These were performed by 30 native Korean speakers who did not participate in either of the two experiments and their independent ratings on a 5-point scale confirmed our judgements. These types of judgements have been proven to correlate well with objective measures of frequency (cf. Segui, Mehler, Frauenfelder & Morton, 1982).

### **3.1.3. Procedure**

The experiment took place in quiet rooms located at the different universities in Seoul. The experiment was controlled using Presentation (Neurobehavioral Systems) and displayed on a Dell Latitude 7480 laptop. On each trial, participants were first presented with the auditory stimulus via headphones (Sennheiser PX200-II). Each target was presented visually in font-size 85pt immediately after the prime (ISI = 0ms) for a duration of 300ms. Participants were instructed to respond to every item on the screen by deciding whether it formed a word in Korean, as quickly and accurately as possible, using the button box provided. Button boxes were custom-made with the buttons labelled in Korean with 'YES' and 'NO'. They were given up to 1500ms to make this decision. All 'YES' responses were made using the participant's

dominant thumb. Each experimental session was preceded by a practice session comprising ten practice items to help familiarize the participants with the task.

### 3.2. Results

Three participants were excluded on the grounds of having made a high percentage of errors (>10%), and three target items that elicited high error rates across all participants (>10%), were excluded. Trials with nonwords were discarded, and only correct responses were submitted to further analysis. Reaction times were trimmed to exclude any responses outside  $\pm 2$  standard deviations from the group mean. Further analysis<sup>5</sup> of frequency measures showed two additional target items to be outliers; these were also not submitted to further analysis. Overall, 9.7% 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, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2018). Reaction times were modelled as a function of the main fixed effect, *Prime Type* and *Cohort Size*. *Cohort Size* refers to the size of the cohort of Hanja which share the Hangul representation of the given Sino-Korean fragment. Treatment coding with the reference level Unrelated was applied to *Prime Type* (Directly Related, Unrelated). Goodness of fit was determined by model comparison and normality of residuals. Following Baayen et al. (2008), all t-values greater than 2 or less than -2 were considered significant. *Participant* and *Target* were treated as random factors. The fixed effect of *Cohort Size* was included in the model, and we analyzed its simple effects within condition. *Cohort Size* for a given Sino-Korean morpheme refers to the size of the cohort of the different Hanja which share the Hangul representation of that particular Sino-Korean morpheme.

The optimal model<sup>6</sup> contained *Prime Type* as a significant fixed effect ( $\beta = -6.09$ ,  $t = -4.07$ ,  $p < .001$ ), random slopes and intercepts for *Prime Type* between participants, and

---

<sup>5</sup> In a post-hoc analysis, we established a cut-off (AVG +/- 2\*SD) for rejection of targets based on word frequency, which caused trials involving two target items to be rejected.

<sup>6</sup> RT ~ PRIME TYPE + (1 + PRIME TYPE | PARTICIPANT) + (1 | TARGET).



random intercepts for targets. The R package *lmerTest* was used to estimate  $p$  values using Satterthwaite approximations (Barr, Levy, Scheepers, & Tily, 2013). Through model comparison, it was established that inclusion of *Cohort Size* as a fixed effect did not improve model fit (at  $\chi^2(1) = 0.163$ ,  $p = .686$ ). We report the model coefficients representing simple effects in Table 1.

	$\beta$	SE	df	$t$	$p$
(INTERCEPT)	453.049	5.741	185	79.918	< .001
RELATED	-6.087	1.496	4657	-4.069	< .001

**Table 1.** Results of Linear Mixed Effects analysis, with significant  $p$ -values ( $p < .05$ ) shown bold faced.

Responses were significantly faster in the Related condition than in the Unrelated condition. This finding is taken as evidence for non-selective access of homophone meanings. This means that hearing a syllable fragment alone is sufficient enough to activate a range of meanings, out of which one is ultimately a semantic match for the word presented in the target.

### 3.3. Discussion

The results from Experiment 1 show that hearing a Sino-Korean syllable, which often has many Hanja representations, in addition to its single Hangul representation, is enough to facilitate non-selective semantic access. Semantically related primes were syllable fragments with multiple homophones, all sharing the same orthographic representation in Hangul, but differing only in their Hanja representation. Their shared Hangul representation also means a shared phonological representation in spoken Korean. We observed that semantically related targets elicited faster reaction times than those which were unrelated. We take this to mean that the phonological representation of a Sino-Korean fragment is enough to activate a cohort comprising a non-selective set of homophones, including their corresponding Hanja, from

which a non-selective set of meanings can also be accessed. Recall that our primes were syllables, not full words which would have otherwise restricted the context, thereby targeting a specific Hanja relevant to the meaning of the word. The syllable is underspecified such that other homophones sharing the same phonological representation is accessed, also activating the different meanings, which could have only come about via activation of the Hanja. It is possible that this is facilitated exclusively through phonological representation, without having to activate intermediate orthographic representations of either the Hangul or Hanja. However, it still stands that cross-script relations which underpin Sino-Korean are invoked; otherwise, reactions times would not have been significantly different between the two conditions.

Our finding is also so far consistent with the literature on multiple and exhaustive semantic access (Swinney, 1979; Onifer & Swinney, 1981; Seidenberg et al., 1982; Lucas, 1987), which suggests that at the offset of the word, all possible meanings are simultaneously activated. This literature is primarily through the lens of European languages written exclusively in the Latin alphabet. In other words, unlike Korean, such languages are uni-scriptal. In light of our findings proving consistent with the present literature, this could be taken as evidence of bi-scriptal and uni-scriptal languages patterning similarly. That is, despite the additional lexical information, the extra orthographic representations do not appear to inhibit nor prevent the same non-selective access observed in uni-scriptal systems from being observed in bi-scriptal systems. This is to say that perhaps non-selective semantic access is able to withstand effects potentially caused by differences in structures of writing systems across languages. One could also argue that our findings so far may suggest that additional representations contributed by the Hanja do not appear to enrich the Korean lexicon, but a proper comparison between uni-scriptal and bi-scriptal systems seems necessary to make this assessment, and as that was not the main focus of the study, this seems difficult to substantiate with the present data. Equally, it may also be the case the non-selective access is in fact bolstered by the additional orthographic input, hence the significantly faster reactions times we observed for semantically related items.

What was relatively underexplored in these studies on non-selective access was the degree to which access is influenced by the size of competing meanings. Several studies have explored the effect of number of meanings (NOM) or senses (NOS) 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. In our study, we found that including an index of cohort size as a predictor 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, proving inconsistent with current literature. However, it is important that note that our study involved fragments, while the literature on ambiguity has been primarily through the lens of full words; thus our findings may not be completely incompatible. It may be the case that as there is no neighboring morpheme to further contextualize and thus restrict the range of activated meanings, syllables do not benefit from the same advantages as full words. For example, the fragment 학 [hak] has two available meanings in its semantic cohort: ‘learning’ and ‘abuse’. On its own, 학 [hak] can evoke either meaning; however, in the context of a full word like, 학교 [hak.gjo] (‘school’), the adjacent morpheme 교 [gjo] immediately narrows down the range of activated meanings to the context-appropriate one, which, in this case, would be ‘learning’. When the morpheme appears within a word, the context-appropriate meaning competes for selection with the other members of the semantic cohort, but when presented alone, there is no immediate context to trigger competition between members.

#### **4. Experiment 2: Cross-modal full word priming**

Experiment 1 found evidence of non-selective access of meanings at the fragment level for Sino-Korean. But it remains unclear whether non-selective access is maintained at the word

level as well. In order to determine whether the results of Experiment 1 can be observed in full words, that is, to see whether lexical access is possible from the meaning of one Hanja to another via their shared homophonic representation in Hangul, we conducted an analogous experiment with Sino-Korean full words. To do so, Experiment 2 introduces a second relatedness condition, in which the target is indirectly related to the meaning of the prime (see Figure 3). That is, while the prime and target are not directly related in meaning, they are related insofar as the Hanja underpinning the prime and target share the same orthographic and phonological representations in Hangul. This shared form is posited to serve a mediating role. Therefore, Experiment 2 tests whether semantic priming effects observed at the fragment level can be similarly observed at the lexical level. We also continue our examination of whether the size of the semantic cohort of Hanja has any effects.

#### **4.1. Method**

##### **4.1.1. Participants**

134 Korean native speakers participated in this experiment, all of whom were students at one of the same three universities as in Experiment 1, had normal or corrected-to-normal vision with no history of neurological disorders, and no reading impairments. These participants did not take part in Experiment 1. All participants provided consent for participation and were reimbursed accordingly for their time. It should be noted again that at the time of testing, all participants had learned Hanja in school for a minimum of two years.

##### **4.1.2. Materials and design**

We created 36 prime-target pairs based on the same set of Hanja that was used for stimuli creation for Experiment 1. These characters were considered to be recognizable for the average native Korean speaker. All prime and target stimuli were disyllabic Korean words.<sup>7</sup> Targets preceded by one of three possible primes: (1) **Directly Related**, (2) **Indirectly**

---

<sup>7</sup> The full list of materials for all experiments is provided in the Appendices.

**Related**, and (3) **Unrelated**. Using a Latin Square design to maximize use of all three possible prime-target pairings, every participant saw each prime and target only once and was exposed to all three conditions. In other words, participants were exposed to a target item only once, preceded by only one of the three possible primes. Assignment of an item to the three different prime conditions was counterbalanced across participants. In order to produce a balanced design in the number of word and nonword stimuli, 36 additional prime-target pairs with nonword targets were included, resulting in a total of 72 items divided into 12 blocks of six trials.

All Directly and Indirectly Related primes were Sino-Korean words, both sharing the constituent fragment of interest. The orthographic representation of this syllable fragment in Hangul is shared by multiple different Hanja. The semantic relationship of each relatedness condition is then based on whether the meaning of the Hanja that is evoked in the prime is matched by the meaning presented in the target. Recall Figure 3, which illustrated the use of **학** [hak], whose orthographic and phonological representation of **학** [hak] in Hangul are shared by two different Hanja ('learning', 'harm'), to instantiate the two relatedness conditions. Both conditions share the constituent **학** [hak], but where the meaning of **학** [hak] which is evoked in the prime is semantically related to the meaning of the target in the Directly Related condition, a meaning of **학** [hak] which is different to the meaning of the target is evoked in the Indirectly Related condition. Unrelated primes were either pure Korean or Sino-Korean words and semantically unrelated to targets. All prime-target pairs did not share orthographic or phonological properties.

The syllable position is sometimes varied across the two experimental conditions due to lexical constraints of the language; that is, it was not always possible to find a set of words within the language that instantiated the two meanings (i.e., the equivalent Hanja character) of interest in the same syllable position. 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 primes where the syllable of interest was syllable-initial, though not significantly so ( $p = .109$  for Indirectly Related;  $p = .089$  for Directly Related).

In this particular paradigm, mediated priming avoids possible confounds as a result of shared constituents between prime and target, and enables us to properly attribute whether observed effects are due to speakers accessing the meanings of more than one Hanja, or repeated exposure to the same form (i.e., seeing the constituent 학 [hak] in both prime and target). This method represents a novel approach to understanding semantic processing in bi-scriptal systems where semantic representations may potentially exist across different scripts.

The same constraints present in stimuli creation for Experiment 1 were also present here, with frequency counts unavailable for individual Sino-Korean fragments. Recall that the fragment stimuli from Experiment 1 were created from the full-word stimuli for Experiment 2 (but as Experiment 1 tested fragments, rather than full words, there was only one relatedness condition (Related) instead of two (Directly and Indirectly Related), as this would only be possible with full words). Thus, once target stimuli fulfilling the necessary semantic relationship to the prime, matching the specific meaning evoked by the Hanja, were created, word-frequencies were matched to the greatest extent possible. A one-way ANOVA over the three main experimental conditions confirmed no significant difference in word frequency ( $F(2, 98) = 0.131, p = .877$ ). The auditory stimuli were recorded and prepared in the same manner as in Experiment 1.

#### **4.1.3. Procedure**

The procedure for this experiment was identical to Experiment 1; participants were asked to make a lexical decision on visual targets upon hearing auditory primes.

#### **4.2. Results**

Four participants with more than 10% errors, and three items that elicited high error rates across all participants, were removed. The same trial exclusion criteria was used as in

Experiment 1. Further analysis of frequency measures showed two additional target items to be outliers<sup>8</sup>; these were also excluded from further analyses. Only trials on which the target was correctly classified were submitted to analysis, and trials with nonwords were not included. Reaction times were again trimmed to exclude any response outside  $\pm 2$  standard deviations from the group mean. Overall, 15.90% of trials were excluded from the analysis.

As per Experiment 1, a linear mixed effects analysis was performed for the remaining reaction times using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2018), using the same parameters. That is, reaction times were modelled as a function of the main fixed effect, *Prime Type* and *Cohort Size*. Treatment coding with the reference level Unrelated was applied *Prime Type* (Directly Related, Indirectly Related, Unrelated). All t-values greater than 2 or less than  $-2$  were considered significant (Baayen, Davidson, & Bates, 2008), and model comparison and normality of residuals were used to establish goodness of fit.

The fixed effect of *Cohort Size* was included in the model, and we again analyzed its simple effects within condition. Recall that *Cohort Size* is the size of the cohort of the different Hanja which share the Hangul representation of the particular Sino-Korean morpheme in question. Also note that the size of the cohort is the same in both the Directly and Indirectly Related conditions, as the same morpheme appears in both primes. Unrelated primes would activate Hanja with cohorts that would not include meanings relevant to the target items, and were coded as having a cohort size of zero. *Cohort Size* was therefore nested under *Prime Type* for analysis. The optimal model included *Cohort Size* as a nested factor within the fixed effect *Prime Type*, and random intercepts for participants and targets<sup>9</sup>. By comparing models with and without the relevant nested factor, we found that removing *Cohort Size* nested within *Prime Type* from the analysis affected goodness of fit ( $\chi^2(2) = 8.88, p = .012$ ).

Responses were significantly faster in the Directly Related condition than in the Unrelated condition; we report the model coefficients representing simple effects in Table 2.

---

<sup>8</sup> As per Experiment 1, we established the same cut-off (AVG +/- 2\*SD) for rejection of targets based on word frequency, which caused trials involving two target items to be rejected.

<sup>9</sup> RT ~ PRIME TYPE/COHORT SIZE + (1 | PARTICIPANT) + (1 | TARGET).

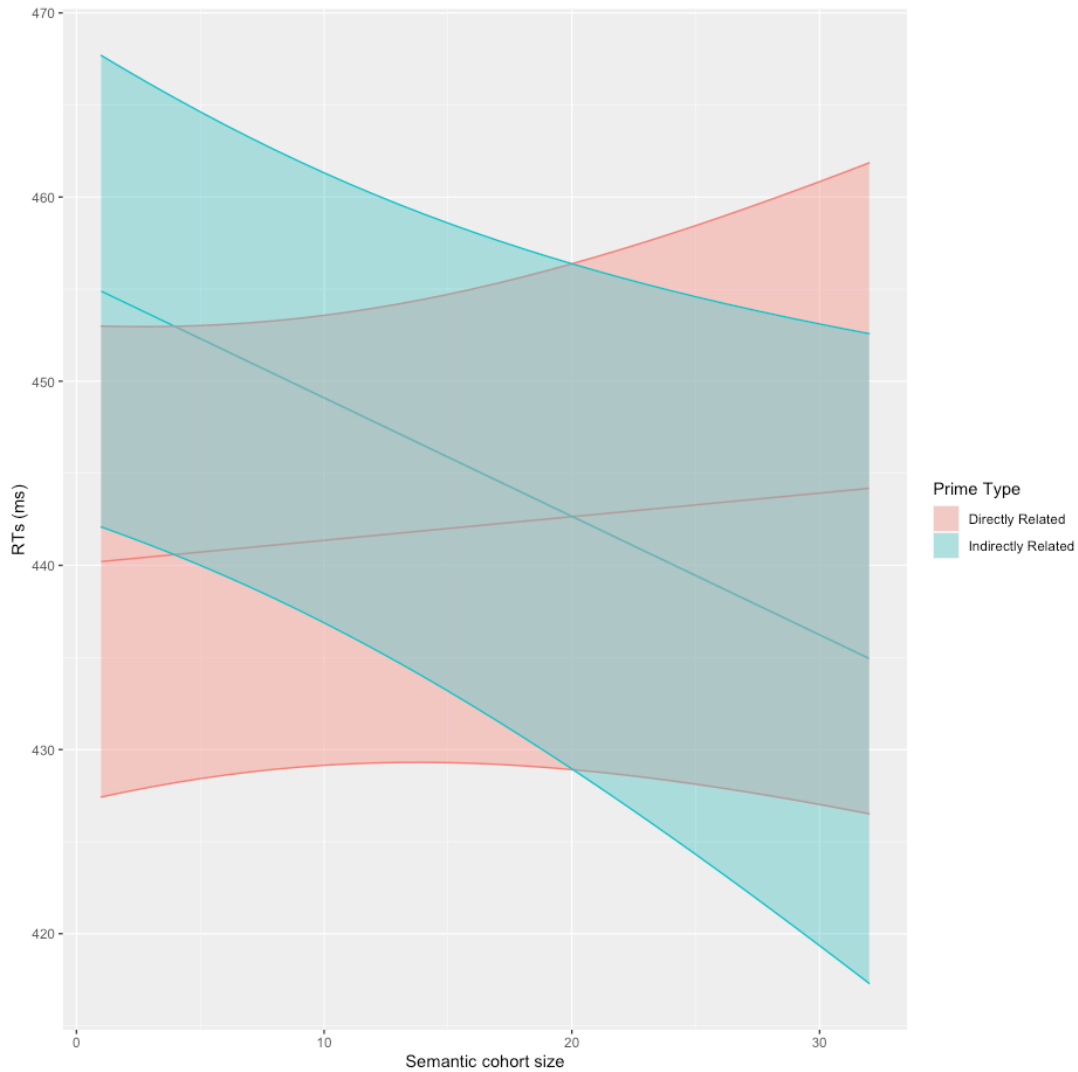
We used the R package *lmerTest* to estimate  $p$  values according to Satterthwaite approximations (Barr et al., 2013).

	$\beta$	SE	df	$t$	$p$
INTERCEPT	453.216	6.218	69	72.883	< .001*
DIRECTLY RELATED	-13.136	3.077	5254	-4.269	< .001*
INDIRECTLY RELATED	2.315	3.088	5237	0.750	.453
<i>Cohort Size</i>					
DIRECTLY RELATED	0.128	0.275	4374	0.467	.641
INDIRECTLY RELATED	-0.643	0.275	4375	-2.342	.019*

**Table 2.** Results of Linear Mixed Effects analysis, with significant  $p$ -values ( $p < 0.05$ ) shown bold faced.

While responses were not significantly faster in the Indirectly Related condition compared to the Unrelated condition, the simple effect of *Cohort Size* in Indirectly Related condition was significant ( $\beta = -0.643$ ,  $t = -2.342$ ,  $p = .019^*$ ). Specifically, an increase in cohort size elicited a significantly greater speed-up in the Indirectly Related condition than in the Unrelated condition. The same was not true in the Directly Related condition.





**Figure 5.** Difference in slopes between the Directly and Indirectly Related conditions in the effect of *Cohort Size*, with decreasing reaction times for increasing cohort size in the Indirectly Related condition.

A post hoc contrast was performed on the difference in slopes between the effect of *Cohort Size* in the Directly Related and Indirectly Related conditions. They showed a significant effect where the Indirectly Related slope (blue) was significantly more negative than the Directly Related slope ( $t[5533] = 2.77, p = .0056$ ), as illustrated in Figure 5.

### 4.3. Discussion

Experiment 2 was conducted to broadly examine the status of Hanja in the Korean mental lexicon. We were interested in testing empirically whether lexical processing of Sino-Korean was affected by the relationship between the two scripts, Hanja and Hangeul. To do this, we

sought to determine whether the directness or indirectness of the semantic relationship between prime and target, as mediated by the shared representation in Hangul, had an impact on target recognition. We additionally examined whether the size of the cohort of Hanja sharing the mediated Hangul representation affected processing. Our results showed a null effect of Prime Type, with reaction times for the Indirectly Related condition not significantly different from those seen in the Unrelated condition. This suggests that participants were just as slow to respond to targets which were indirectly related to primes, as those which are completely unrelated. Reaction times for targets in the Indirectly Related condition were slower than those in the Unrelated condition, although not significantly slower. However, this does still give some suggestion to the possibility that lexical processing is inhibited to some degree by activation of competing Hanja.

The effects of cohort size observed at the lexical level depart from those observed at the fragment level in Experiment 1. The results here showed that within the Indirectly Related condition, the condition of interest, reaction times sped up significantly with increasing size of semantic cohort (i.e., there was, on average, a speedup of 0.64ms for every additional cohort member). The combined findings on cohort size between the two experiments suggest possible effects of lexicality and that access to the semantic cohort is governed by the degree to which a limiting context is available, i.e., a neighboring syllable to instantiate a full word, thus narrowing down the cohort to one context-supported Hanja. While target recognition when preceded by fragment primes was seemingly unaffected by whether the fragment was shared by fewer or more Hanja, such was not similarly nor uniformly observed when targets were preceded by full word primes. This finding of faster reaction times for targets preceded by indirectly related primes with larger cohorts, combined with the overall null effect of Prime Type for the Indirectly Related condition, suggests an interesting interplay. When there is a semantic match between prime and target, both at the lexical level, as well as at the 'hidden' sub-lexical level of Hanja, that is, the Hanja instantiated in the prime matches the semantics of the target, non-selective access of all available Hanja offers no processing benefit. However, when there is no such match, as is the case in the Indirectly Related condition, the benefits of

non-selective access are enacted on in the service of lexical access, as evidenced by faster word recognition.

While the absence of indirect priming does not alone indicate exhaustive access of all Hanja, that reaction times are modulated by cohort size in the Indirectly Related Condition while not in the Directly Related condition offers some evidence of non-selective access. The direction of the effect size, which suggests that greater cohort sizes yield greater semantic facilitation, is striking, given that previous research has the inhibitory effects of a word's neighborhood density during auditory processing (Goldinger, Luce, & Pisoni, 1989; Luce, 1986; Luce, Pisoni, & Goldinger, 1990). One explanation for our results is that the tenuousness of the relationship between primes and targets which are indirectly related presents a higher threshold for successful semantic activation, and that to overcome the initial mismatch between the Hanja instantiated in the prime and the meaning presented in the target, a larger cohort acts to overcompensate in service of lexical access. Further research is needed to further elucidate this discrepancy.

## **5. General discussion**

Lexical processing involves parsing in real-time the form and structure of speech signals. This means that upon hearing a specific signal, a listener activates a number of different representations that are invariably linked to one another and correspond to the lexical information stored within the mental lexicon. Numerous studies have been devoted to understanding the precise nature of these representations, but primarily through the lens of English or languages with Latinate systems (Rumelhart & McClelland, 1982; McClelland & Elman, 1986; Norris, 1994; Marslen-Wilson, Tyler, Waksler, & Older, 1994, McQueen, 2005 for a review). What remains underexplored is how representations are enriched for languages with non-Latinate systems or languages which have more than one script. The literature has expanded to include languages of the former, but the literature remains sparse on those of the latter. Languages which have been sampled that do not use Roman-based alphabets include most notably Chinese (Zhou, Marslen-Wilson, Taft, & Shu, 1999; Liu & McBride-Chang, 2010;

Tsang & Chen, 2013a; 2013b), Arabic (Boudelaa & Marslen-Wilson, 2001; 2005; 2015), and Hebrew (Feldman, Frost, & Pnini, 1994; Frost, Forster, & Deutsch, 1997), however far fewer languages offer the opportunity to study bi-scriptal processing. If different orthographic systems encode lexical information differently, then the ways in which this information is integrated in the mental representation of words may also differ, particularly if the system comprises multiple interacting scripts. It may be the case that orthographic representations reflect critical aspects of writing systems, and that for bi-scriptal lexicons, orthographic representations are enriched by the contributions of multiple scripts.

This study was concerned with the representation of lexical information in bi-scriptal languages, where two different scripts comprise the writing system. We capitalized on Korean's use of both Hangul and Hanja to investigate whether the mental representation of Sino-Korean words encoded the contributions of Hanja. Sino-Korean words can be written in both Hangul and Hanja, and in many cases, can be written using many different Hanja. Sino-Korean's repeated use of the same syllable creates homophones in spoken Korean which are also manifest as homographs in Hangul due to the highly transparent nature of the relationship between sound and spelling. Only through Hanja can these homophones be disambiguated. Take again the example of 학 [hak], a morpho-syllabic fragment the orthographic and phonological shape of which is shared by Hanja characters 學 and 虐, ('learning', 'harm'), which as a result contribute those two meanings to 학 [hak]. The multiplicity in available orthographic representations offers an important opportunity to explore the effects of script on lexical processing. In light of this, our research objective was twofold: (1) to examine whether the semantics of Hanja are represented in the mental lexicon, and (2) to explore whether the relative number of encoded meanings modulates semantic access.

We conducted cross-modal semantic priming experiments examining whether the different semantics of Hanja could be co-activated and accessed by one another, via mediation of their shared orthographic and phonological representation in Hangul. We additionally examined whether target word recognition in Sino-Korean was modulated by the

semantic cohort size of different co-activated meanings of Hanja. We observed that at the fragment level, individual morpho-syllabic fragments facilitated successful word recognition of targets semantically related to primes instantiating one of the possible meanings evoked by the different Hanja sharing the same Hangul representation of the prime. These effects, however, were unaffected by whether the Hangul representation of the prime was shared by fewer or more different Hanja, that is, the cohort size neither bolstered nor inhibited processing. Non-selective semantic access has been observed before in uni-scriptal systems, such as English (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979; Swinney, 1979; Onifer & Swinney, 1981), which suggests that multiple orthographic representations do not inhibit the activation of non-selective sets of meanings.

We then tested whether we would find similar evidence for non-selective access of Hanja at the word level, when primes and targets were indirectly related via mediation of a different Hanja sharing the same Hangul representation. Primes were disyllabic Sino-Korean words containing a constituent, the Hangul representation of which was shared by multiple different Hanja. We reasoned that the orthographic and phonological shape of this constituent would activate a cohort of different meanings as contributed by the different Hanja, one of which is instantiated in the prime itself, and another of which is semantically related to the target. The same Hangul representation between these two meanings of different Hanja would enable a mediated pathway to target word recognition. We also considered in this context as well whether the semantic cohort size would have an effect. This particular paradigm of mediated priming ensured that primes and targets did not overlap in orthographic or phonological form, different to other paradigms where primes and targets did share constituents (Yi & Yi, 1999). Our paradigm also differed from the classic cross-modal semantic priming experiments in one further aspect, in that we employed polymorphemic words within which one constituent (morphemic) fragment is a homophone.

We observed priming between primes and targets which were semantically related at the lexical level, but also at the sub-lexical level of Hanja, as would be expected. We also observed that the cohort of Hanja evoked at the sub-lexical level did not influence the speed

of target word recognition, that is, reaction times remained stable across increasing cohort sizes. While there was no priming was found in cases where primes and targets were only indirectly related, we also found that in these particular instances, reaction times sped up with increasing cohort sizes. In other words, in cases where relatedness between primes and targets relied on the constituent fragment to exhaustively activate the different Hanja meanings, the size of the semantic cohort made a significant difference – the greater the cohort size, the greater the semantic facilitation. That cohort size mattered only for targets indirectly related to primes and not for those which were directly related offers some suggestion of non-selective semantic access at the lexical level.

Overall, across the two studies, we find varied evidence for semantic co-activation of different Hanja during Sino-Korean lexical processing. Broadly these findings are consistent with the findings from non-selective homophone processing reported by Swinney (1979; Onifer & Swinney, 1981), which lends some support for the possibility that uni-scriptal and bi-scriptal systems represent orthographic information in similar ways. This is not to say that the lexical representations of bi-scriptal languages are not enriched by additional scripts, but that additional orthographic information does not prevent non-selective access of homophone meanings. Our findings therefore bolster the existing literature on semantic co-activation, the support for which can now be extended to include bi-scriptal languages, showing further that co-activation can occur across different scripts. This finding appears to be consistent with the literature on embedded word processing, which investigates the processing of constituent words which bear no relevance in meaning to the larger word in which it is embedded (e.g., *bone* within *tromBONE*). Shilcock (1990) found words like *trombone*, when presented in a sentence, to elicit priming of responses to the word *rib*, semantically related in similar fashion to our primes and targets whereby it is associated with the embedded word. The key difference between our study and those similar to Shilcock's (1990) is our use of morphemes (i.e., fragments), rather than full words. Our findings prove not as consistent as those of Shilcock's which used English as its testbed. Such differences in the way in which embedded forms are accessible during processing could reflect possible differences between bi-scriptal and uni-

scriptal languages and the potential for cross-script dynamics to interfere with constituent processing in languages with multiple orthographic inputs.

The findings presented here invite further exploration of the effects of number of meanings (NOM) or senses (NOS). Previous studies have reported 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), wherein words with multiple meanings benefit from this multiplicity, resulting in faster visual word recognition. However, 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)<sup>10</sup>. Different meanings, as contributed by different Hanja, can be more or less closely related to one another, falling on a continuum ranging from homonymous to polysemous forms. 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, 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. The semantic relations between different meanings are therefore not necessarily categorically related. This potentially creates difficulties in fitting the meaning into the mental lexicon, possibly exacerbated if lexical entries are further overloaded by orthographic information.

Our findings also shed new light on some of the existing literature on Sino-Korean processing. A further study by Yi (1999a) was conducted on Sino-Korean compounds to differentiate between effects of morphological and semantic priming, as the latter can often create a confound. The results showed that between morphological and semantic facilitation,

---

<sup>10</sup> 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.

the degree of morphological facilitation was significantly lower than that of semantic facilitation. These findings were taken to suggest that, in Sino-Korean, inter-lexical connections are relatively strong compared to morpho-lexical connections (Yi, 2009). This means that, while lexical processing of Sino-Korean involves representation of Sino-Korean morphemes, it does not take morphological activation as an obligatory step during lexical access. Yi's model assumes that not all contributions of the different Hanja can be activated, since this would require spreading activation across too wide of a network, given the large number of mappings between characters (Hangul) and morphemes (Hanja).

One final consideration is whether the type of cross-script activation observed here could simply be accounted for by phonological priming. Although that may be a possibility, given that participants all had a minimum of two years of learning Hanja, it seems unlikely that this literacy would be completely ignored. The information may not be consciously recruited, but the knowledge in itself is likely to play some role. Once a speaker is equipped with this type 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. In our study, we observed that non-selective semantic access was viable to some extent during visual word recognition, aided by the activation of a semantic cohort, permitting multiple access to all available Hanja. This potentially calls into question a supra-lexical approach in which morphemic constituents are accessible only after activation of the full word, as our results corroborate the view of morphological activation preceding whole word activation. Our results are doubly corroborated by the vast literature consistent with a decompositional approach to visual word recognition wherein morphologically complex words are obligatorily decomposed into their constituent morphemes before lexical retrieval. Reasons for such discrepancies warrants further evidence.



## 6. Conclusions

The objective of our study was to carve out additional yet necessary space for the role that script plays, given the deeply-rooted nature of the scripts' complexities to the Korean lexicon. These findings suggest that the way in which the mental lexicon organizes itself is, in some part, sensitive to the contributions of script and opens up new lines of inquiry for languages with equally intricate writing systems, whether bi-scriptal or not. This is to say that the architectures supporting the mental lexicon for single script languages may also be extended to support those with multiple. 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 towards a more exhaustive understanding of the role of written language in psycholinguistic processing.

Our findings suggest that recognition of Sino-Korean words potentially involve the active processing of the different semantics the multitude of Hanja has to offer. This proved to be modulated by the relative number of possible meanings available for a given Sino-Korean morpheme. These findings are presently specific to native speakers of Korean, so the next step would be to investigate heritage speakers who have varied levels of awareness of Hanja. Further research testing Korean speakers with no previous exposure to Hanja, such as North Korean migrants, or heritage speakers who are unlikely to have had formal education learning Hanja, may offer further insight into Sino-Korean processing. The findings from our study lay important foundation to better understanding the representation of Hanja in native speakers. However, a fuller theory of the Korean mental lexicon requires additional work on speakers with varying levels of understanding of Hanja, including those with limited to no awareness. A point of interest would be how the links between the different levels of representation from Hanja and Hangeul are strengthened or lessened with different levels of literacy depending on the background of the speakers. Likewise, any comprehensive model of the Korean mental lexicon will need to consider those with no formal knowledge of Hanja, yet some awareness through informal exposure to repeated uses of the same orthographic Hangeul representation to express different meanings. Previous research from An (2014) further corroborates the

need to address more carefully individual differences in levels of familiarity with Hanja across different sets of Korean speakers. An's (2014) research highlights yet another crucial consideration for our theorization of the Korean mental lexicon, namely whether analyses should treat Sino-Korean and pure Korean words together or separately. Answering such questions will be key to developing a fuller theory of the Korean mental lexicon, for which a clearer understanding of how Sino-Korean is processed in native speakers is foundationally important. Building on the initial evidence presented here of the intricate nature of cross-script orthographic representations used in Korean word recognition, closer examination of individual differences in literacies between speakers of Korean remains a key question for further research.

### **Acknowledgments**

We would like to thank the members of the Language and Brain Laboratory in Oxford for their assistance and support; in particular, we would like to thank Colin Brooks for his help with the programming of the experiments. We would also like to thank Jongwon Choe, Junyoung Heo, and Yunmo Chung for helping with the organization of the experiments at Sookmyung Women's University, Hansung University, and Kyung Hee University, respectively. We also thank Laurel Brehm and Adam Roberts for their help with statistical analysis and data visualization, and Emily Darley for her helpful suggestions on earlier versions of this manuscript as well as her assistance with statistical modelling. Additionally, we thank Antje Meyer and the members of the Psychology of Language Department for their helpful comments on preliminary findings that formed the basis for the present research. This research was supported by the Clarendon Fund, the Scatcherd European Scholarship, the Basant Kumar and Sarala Birla Graduate Studentship, and the Wolfson College Travel Grant. Aditi Lahiri's research is also supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement number: 695481). The funding sources were not involved in the conduct of the research and the preparation of the article.

## References

- An, S. (2010). A study of how Korean speakers comprehend compositions of disyllabic Sino-Korean words. *Morphology*, 12(2), 201-216.
- An, S. (2011). *A mental lexicon-based study on Sino-Korean words* (Doctoral dissertation). Seoul National University, Seoul, South Korea.
- An, S. (2014). How is the mental lexicon to address questions regarding the morphology of Sino-Korean words? *Korean Linguistics*, 62, 373-394.
- 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. <https://doi.org/10.1006/jmla.1997.2502>
- 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(4), 390-412. <https://doi.org/10.1016/j.jml.2007.12.005>
- 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(3), 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects Models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved October 2015 from <http://www.praat.org/>.
- Bölte, J. & Coenen, E. (2002). Is phonological information mapped onto semantic information in a one-to-one manner? *Brain and Language*, 81(1-3), 384-397. <https://doi.org/10.1006/brln.2001.2532>
- Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 63-85. <https://doi.org/10.1037/0278-7393.22.1.63>
- Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, 81(1), 65-92. [https://doi.org/10.1016/S0010-0277\(01\)00119-6](https://doi.org/10.1016/S0010-0277(01)00119-6)
- Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, 20(1-2), 207-260. <https://doi.org/10.1080/01690960444000106>
- Boudelaa, S. & Marslen-Wilson, W. D. (2015). Structure, form, and meaning in the mental lexicon: Evidence from Arabic. *Language, Cognition, and Neuroscience*, 30(8), 955-992. <https://doi.org/10.1080/23273798.2015.1048258>
- 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. <https://doi.org/10.3758/BF03211829>

- Chang, Y., Monaghan, P., & Welbourne, S. (2019). A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language*, 108, 104025. <https://doi.org/10.1016/j.jml.2019.05.003>
- 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. <https://doi.org/10.3758/s13423-011-0095-8>
- Feldman, L. B., Frost, R., & Pnini, T. (1995). Decomposing words into their constituent morphemes: Evidence from English and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 947-960. <https://doi.org/10.1037/0278-7393.21.4.947>
- Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 829-856. <https://doi.org/10.1037/0278-7393.23.4.829>
- Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28(5), 501–518. [https://doi.org/10.1016/0749-596X\(89\)90009-0](https://doi.org/10.1016/0749-596X(89)90009-0)
- 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. <https://doi.org/10.1037/0278-7393.23.5.1122>
- 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. <https://doi.org/10.1037/0033-295X.111.3.662>
- 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. <http://dx.doi.org/10.1037/0096-1523.22.6.1331>
- Huettig, F., & Mishra, R. K. (2014). How literacy acquisition affects the illiterate mind – A critical examination of theories and evidence. *Language and Linguistics Compass*, 8(10), 401-427. <https://doi.org/10.1111/lnc3.12092>
- Hoshino, N. & Kroll, J. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, 106(1), 501-511. <https://doi.org/10.1016/j.cognition.2007.02.001>
- Jastrzemski, J. E. (1981). Multiple meaning, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, 3, 278-305. [https://doi.org/10.1016/0010-0285\(81\)90011-6](https://doi.org/10.1016/0010-0285(81)90011-6)
- 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. <http://dx.doi.org/10.1037/0096-1523.14.4.601>

- Kim, H. (1983). Chinese characters and the Korean language. In the Korean National Commission for UNESCO (Ed.), *The Korean language* (pp. 121-127). Si-sa-yong-o-sa Publishers.
- Kim, J. & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language*, 49(4), 484-499.  
[https://doi.org/10.1016/S0749-596X\(03\)00093-7](https://doi.org/10.1016/S0749-596X(03)00093-7)
- Kim-Renaud, Y. (1997). *The Korean alphabet: Its history and structure*. University of Hawaii Press.
- Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain Language*, 81, 205-223.  
<https://doi.org/10.1006/brln.2001.2518>
- Lee, K. M. (1983). Foundations of Hunmin Chongum. In the Korean National Commission for UNESCO (Ed.), *The Korean language* (pp. 71-79). Si-sa-yong-o-sa Publishers.
- Lee, C. (1972). Boundary phenomena in Korean revisited. *Paper in Linguistics*, 5(3), 454-474. <https://doi.org/10.1080/08351817209370314>
- Liu, P. D. & McBride-Chang, C. (2010). Morphological processing of Chinese compounds from a grammatical view. *Applied Psycholinguistics*, 31(4), 605-617.  
<https://doi.org/10.1017/S0142716410000159>
- Lucas, M. M. (1987) Frequency effects on the processing of ambiguous words in sentence contexts. *Language and Speech*, 30(1), 25-46.  
<https://doi.org/10.1177/002383098703000103>
- Luce, P. A. (1986). Neighborhoods of words in the mental lexicon (Research on Speech Perception, Tech. Rep. No. 6). Bloomington: Indiana University, Department of Psychology, Speech Research Laboratory.
- Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G.T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). MIT Press.
- Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24(2), 173-193.  
<https://doi.org/10.1017/S0142716403000092>
- 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.  
<https://doi.org/10.1037/0033-295X.101.1.3>
- McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1-86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255-275). Sage Publications.
- Millis, M. L. & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory and Cognition*, 17, 141-147.  
<https://doi.org/10.3758/BF03197064>

- 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. <http://dx.doi.org/10.1080/20445911.2014.953167>
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189-234. [https://doi.org/10.1016/0010-0277\(94\)90043-4](https://doi.org/10.1016/0010-0277(94)90043-4)
- Oden, G. C., & Spira, J. L. (1983) Influence of context on the activation and selection of ambiguous word senses. *Quarterly Journal of Experimental Psychology*, 35A(1), 51-64. <https://doi.org/10.1080/14640748308402116>
- Onifer, W. & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9, 225-236. <https://doi.org/10.3758/BF03196957>
- R Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rodd, J., Gaskell, G. & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245-266. <https://doi.org/10.1006/jmla.2001.2810>
- Rumelhart, D. E. & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89(1), 60-94. <https://doi.org/10.1037/0033-295X.89.1.60>
- Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9, 487-494. [https://doi.org/10.1016/S0022-5371\(70\)80091-3](https://doi.org/10.1016/S0022-5371(70)80091-3).
- Segui J., Mehler, J., Frauenfelder, U., & Morton, J. (1982) The word frequency effect and lexical access. *Neuropsychologia*, 20(6), 615-627. [https://doi.org/10.1016/0028-3932\(82\)90061-6](https://doi.org/10.1016/0028-3932(82)90061-6)
- Seidenberg, M., 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. [https://doi.org/10.1016/0010-0285\(82\)90017-2](https://doi.org/10.1016/0010-0285(82)90017-2)
- Song, K. (1992). Structure of Modern Korean. *Korean Text*, 1, 1-85.
- 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). MIT Press.
- Shin, S., Lee, D., & Lee, H. (1990). *Understanding Hunmin-jong.um*. Hanshin.
- 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. <https://doi.org/10.1111/1467-9280.00151>

- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 645-659. [https://doi.org/10.1016/S0022-5371\(79\)90355-4](https://doi.org/10.1016/S0022-5371(79)90355-4)
- Swinney, D., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition*, 7, 159-165. <https://doi.org/10.3758/BF03197534>
- 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 & Verbal Behavior*, 18(4), 427-440. [https://doi.org/10.1016/S0022-5371\(79\)90237-8](https://doi.org/10.1016/S0022-5371(79)90237-8)
- Taylor, I., & Taylor, M. M. (2014). *Writing and literacy in Chinese, Korean, and Japanese: Revised edition*. John Benjamins. <https://doi.org/10.1075/swll.14>
- Eddington, C. M. & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. *Psychonomic Bulletin & Review*, 22, 13-37. <https://doi.org/10.3758/s13423-014-0665-7>
- 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. <https://doi.org/10.1016/j.jml.2012.11.003>
- 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. <https://doi.org/10.1037/a0033701>
- Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, 35, 938-952. <https://doi.org/10.3758/BF03193467>
- Yi, K. (1999a). Representation and processing of Hanja compound words. *Proceedings of the Annual Conference of Korean Society for Cognitive Science*, Seoul, Korea, 73-79.
- Yi, K. (1999b) Morphological processing in Korean word recognition. *Proceedings of the Annual Conference of Korean Society for Experimental and Cognitive Psychology*, Seoul, Korea, 35-42.
- Yi, K. (2003). The effects of word type on word recognition in Korean. *Korean Journal of Cognitive and Biological Psychology*, 15(4), 479-498. <http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06374352>
- Yi, K. (2009). Morphological representation and processing of Sino-Korean words. In C. Lee, G. Simpson, Y. Kim (Eds.), *The handbook of East Asian psycholinguistics* (pp. 409-417). Cambridge University Press. <https://doi.org/10.1017/CBO9780511596865>
- Yi, K. & Yi, I. (1999). Morphological processing in Korean word recognition. *Korean Journal of Experimental and Cognitive Psychology*, 11(1), 77-91. <http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06374248>
- 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.  
<https://doi.org/10.1080/17470211003718705>

Zhou, X., Marslen-Wilson, W. D., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology in reading Chinese compound words. *Language and Cognitive Processes*, 14(5-6), 525-565. <https://doi.org/10.1080/016909699386185>



## Appendix A

### Stimuli Experiment 1

Target	Related	Unrelated
그림 ('picture')	도	치
눈병 ('eye infection')	안	감
슬픔 ('sadness')	애	합
철회 ('withdrawal')	무	소
꽃밭 ('garden')	화	우
구름 ('cloud')	운	회
아내 ('wife')	처	호
확실 ('certainty')	명	허
높음 ('height')	고	수
울림 ('echo')	향	설
사이 ('relationship')	간	영
벌레 ('bug')	충	거
말씀 ('words')	담	린
소원 ('wish')	희	직
먼저 ('first')	선	단
마당 ('garden')	장	규
맺음 ('conclusion')	결	두
임금 ('monarch')	군	칭
머슴 ('servant')	노	썰
시초 ('root')	원	내
보리 ('barley')	맥	여
엄마 ('mother')	모	행
벼슬 ('gov't post')	관	촌
다툼 ('quarrel')	쟁	응
미리 ('in advance')	예	매
저녁 ('dinner')	석	포
걸음 ('step')	보	후
우측 ('the right-side')	동	덕
고요 ('silence')	적	망
부처 ('Buddha')	불	참
소년 ('boy')	남	속
새싹 ('sprout')	아	완
화합 ('harmony')	협	늑
터짐 ('tear')	폭	스
구슬 ('bead')	옥	마
소리 ('noise')	음	얼
종결 ('end')	졸	투
배려 ('consideration')	정	식
뿌리 ('root')	근	긴
여름 ('summer')	하	필
소금 ('salt')	염	최
종이 ('paper')	지	언
참음 ('patience')	인	과
회상 ('recollection')	기	실

바람 ('wind')	풍	통
곡식 ('grain')	미	토
넓음 ('spaciousness')	광	만
겨레 ('brethren')	민	앵
승리 ('victory')	극	태
행복 ('happiness')	환	툽
홀로 ('alone')	독	차
죽음 ('demise')	사	습
출산 ('childbirth')	생	위
상처 ('wound')	학	액
거짓 ('lie')	가	복
어음 ('IOU')	권	느
여자 ('woman')	매	경
일찍 ('early')	조	의
나비 ('butterfly')	접	꼬
만듦 ('creation')	작	개
노래 ('song')	요	갑
세움 ('standing')	건	교
가난 ('poor')	빈	늘
가슴 ('chest')	흉	번
가을 ('autumn')	추	파
통증 ('pain')	뇌	시
스승 ('teacher')	부	멈
조각 ('piece')	편	함
영웅 ('hero')	승	먹
만남 ('meeting')	봉	골
생선 ('fish')	어	범
나이 ('age')	연	외

## Appendix B

### Nonword Stimuli Experiment 1

<u>Prime</u>	<u>Target</u>
송	우허
제	민텅
흡	당바
증	참강
택	산어
착	시즌
루	잔광
족	가삼
급	유번
창	운타
잘	처텅
료	연탕
능	버하
칭	오머
격	바석
존	거남
잔	사밋
논	손삼
념	하힘
텔	소깊
넘	차망
역	보하
칩	푸호
판	도팜
바	버푸
다	가접
멋	염힌
준	얌어
랑	필처
녀	한버
낭	옥수
략	위송
출	해잔
엽	부누
즉	난양
진	어바
특	짐거
목	궁잠
섯	남마
립	서잔
득	광시
래	잡함
름	거양
양	디보

보  
정  
편  
설  
간  
도  
삼  
안  
변  
미  
박  
간  
분  
위  
병  
성  
섬  
서  
마  
가  
인  
피  
면  
동  
등  
희  
정  
칭  
친  
품  
후  
작  
경  
곤  
상  
주  
잡  
더  
뚝  
단  
몽  
보  
근  
마  
버  
자  
송  
간  
두  
누  
하  
승  
여  
양

방  
국  
콤  
항  
친  
홍  
질  
론  
현  
냥  
타  
온  
껏  
동  
혼  
찰  
늦  
서  
등  
취  
적  
품  
속  
락  
업  
귀  
탁  
퇴

## Appendix C

**Stimuli Experiment 2** (The Sino-Korean syllable of interest is presented in bold.)

Target	Directly Related	Indirectly Related	Unrelated
종결 ('termination')	졸업 ('graduation')	치졸 ('crudity')	질투 ('jealousy')
배려 ('consideration')	애정 ('affection')	정의 (justice)	식품 ('groceries')
뿌리 ('root')	근본 ('foundation')	근대 ('modern time')	긴급 ('urgency')
여름 ('summer')	하계 ('summertime')	하체 ('lower body')	필통 ('pencil box')
소금 ('salt')	염전 ('saltpan')	염색 ('dye')	최대 ('maximum')
종이 ('paper')	백지 ('blank paper')	지구 ('earth')	언니 ('sister')
참음 ('tolerance')	인내 ('patience')	인사 ('greeting')	과립 ('capsule')
회상 ('recollection')	기억 ('memory')	공기 ('atmosphere')	육실 ('classroom')
바람 ('wind')	해풍 ('sea breeze')	중풍 ('stroke')	통계 ('statistics')
곡식 ('grain')	현미 ('brown rice')	미술 ('art')	토굴 ('cave')
넓음 ('spaciousness')	광장 ('plaza')	광택 ('shine')	만화 ('cartoon')
겨레 ('brethren')	민족 ('countrymen')	민감 ('sensitivity')	앵두 ('cherry')
승리 ('victory')	극복 ('conquest')	극심 ('extremity')	황태 ('pollock')
행복 ('happiness')	환희 ('jubilation')	환불 ('refund')	손톱 ('fingernail')
홀로 ('alone')	단독 ('independence')	독서 ('reading')	목차 ('index')
죽음 ('demise')	사망 ('death')	식사 ('meal')	복습 ('review')
출산 ('childbirth')	생일 ('birthday')	희생 ('sacrifice')	거위 ('goose')
상처 ('wound')	학대 ('abuse')	학교 ('school')	액자 ('frame')
거짓 ('lie')	가면 ('mask')	가훈 ('family')	복학 ('return')
어음 ('IOU')	증권 ('stock')	권력 ('authority')	느낌 ('feeling')
여자 ('woman')	자매 ('sisters')	중매 ('matchmaking')	경기 ('game')
일찍 ('early')	조퇴 ('early leave')	조세 ('taxes')	의자 ('chair')
나비 ('butterfly')	점령 ('domination')	접합 ('joining')	꼬리 ('tail')
만듦 ('creation')	작품 ('masterpiece')	수작 ('nonsense')	날개 ('wings')
노래 ('song')	가요 ('music')	요리 ('cooking')	장갑 ('glove')
세움 ('standing')	건국 ('foundation')	두건 ('bandana')	교활 ('artfulness')
가난 ('poor')	빈곤 ('poverty')	귀빈 ('VIP')	하늘 ('sky')
가슴 ('chest')	흉상 ('bust')	흉년 ('poor harvest')	번역 ('translation')
가을 ('autumn')	추분 ('fall equinox')	추억 ('memory')	파도 ('wave')
통증 ('pain')	고뇌 ('anguish')	두뇌 ('brain')	접시 ('plate')
스승 ('teacher')	사부 ('master')	부활 ('revival')	멈춤 ('stall')
조각 ('piece')	과편 ('fragment')	편식 ('picky diet')	포함 ('inclusion')
영웅 ('hero')	승자 ('victor')	승마 ('horse riding')	먹이 ('prey')
만남 ('meeting')	상봉 ('reunion')	봉투 ('envelope')	골절 ('fracture')
생선 ('fish')	어부 ('fisherman')	어뢰 ('torpedo')	범위 ('scope')
나이 ('age')	연세 ('age')	인연 ('destiny')	참외 ('melon')

## Appendix D

### Nonword Stimuli Experiment 2

Prime	Target (nonword)
전송 ('transmission')	우혀
삭제 ('elimination')	민딩
흡수 ('absorption')	당바
검증 ('verification')	참강
선택 ('decision')	산어
착석 ('taking a seat')	시즌
가루 ('powder')	잔팡
만족 ('satisfaction')	가삽
보급 ('dissemination')	유번
창문 ('window')	윤타
잘못 ('fault')	처텅
종료 ('end')	언탕
가능 ('possibility')	버하
명칭 ('title')	오머
합격 ('acceptance')	바석
보존 ('preservation')	거남
잔치 ('feast')	사밋
논문 ('dissertation')	손삼
이념 ('ideology')	하험
호텔 ('hotel')	소깊
넘다 ('climb')	차망
역사 ('history')	보하
첩자 ('spy')	푸호
재판 ('trial')	도팜
바보 ('fool')	버푸
다정 ('tenderness')	가접
멋짐 ('snazziness')	염힌
준비 ('preparation')	얌어
사랑 ('love')	필처
그녀 ('woman')	한버
낭비 ('waste')	옥수
대략 ('approximation')	위송
제출 ('submission')	해잔
엽서 ('postcard')	부누
즉석 ('instant')	난앙
진심 ('sincerity')	어바