

# Korean *Wh*-Island Sentences: Individual Differences in Acceptability Judgments and Working Memory

Hye-ryeong Hahn<sup>†</sup>

Seowon University

---

## ABSTRACT

While it has been claimed that scrambling in East Asian languages is not subject to *wh*-island effects, studies into theoretical and experimental syntax have suggested that native Korean speakers vary in their acceptance of island constructions. This study addresses the individual differences in the acceptability judgments of Korean *wh*-island sentences in terms of working memory. In order to determine if the varied judgments of acceptability are attributable to differences in individuals' working memory, an acceptability judgment test on various Korean *wh*-island constructions, and two memory span tests—a reading span test and a digit span test—were administered to 66 adult Korean native speakers. Pearson correlation analysis and a factor analysis were then performed on the participants' acceptability judgment ratings and memory spans. The results revealed that the individuals' acceptability judgments for *wh*-island sentences were not correlated with their reading span or digit span, although the group's mean acceptability of the constructions gradually decreased as their expected processing complexity increased. These findings indicate that differences in individuals' working memory are not the source of the observed acceptability variation.

**Keywords:** *wh*-island, Korean, acceptability judgment, individual differences, working memory

---

## 1. Introduction

One of the long-held assumptions about native speakers' knowledge of language—or grammar—has been that it is rule-based: a sentence is either well-formed or ill-formed depending on whether or not it is constrained by the rules of grammar. Another assumption about native speakers' language competence has been that the rules of grammar in a language are shared by its native speakers (Chomsky 1965). However, there are cases where acceptability judgments diverge among native speakers. A typical case concerns judgments of sentences involving long-distance

---

<sup>†</sup> Corresponding author: [hyerhahn@seowon.ac.kr](mailto:hyerhahn@seowon.ac.kr)



scrambling by native speakers of East Asian languages such as Korean and Japanese. For some speakers of these languages, island constructions containing an element extracted out of a *wh*-clause such as the Japanese sentence in (1) are quite acceptable, whereas for others, they are completely unacceptable (DH Chung 2005; J-I Han 1992; S-H Hong 2004; Y-S Kang 1986; R-H-Y Kim 2003; H-S Lee 1982; H-M Sohn 1980; C Suh 1987).

- (1) Nani-o<sub>i</sub> John-ga [<sub>WH-ISD</sub> Taroo-ga t<sub>i</sub> katta-ka] siritagatteiru. (Saito 1989)  
what-ACC John-NOM Taro-NOM \_ bought-Q want-to-know  
'John wants to know what Taro bought.'

Many syntacticians and psycholinguists who have argued that scrambling in Korean or Japanese is not subject to *wh*-island effects have delved into the properties of this long-distance movement as a key to the syntactic nature of these languages (DH Chung 1996; Y-S Lee 1993; Miyagawa 2005; H-K Park 2010; Saito 1989; Ueyama 2008; Yoshimura 1992). Meanwhile, experimental studies based on acceptability judgments generally indicate that native speakers tend to show varying degrees of rejection to sentences with such long-distance scrambling (H-r Hahn 2015a, 2015b; B Kim and Goodall 2014; J-M Yoon 2013). B Kim and Goodall (2014), for example, found that scrambling out of the embedded *wh*-clause resulted in negative acceptability ratings around -0.4 on a scale from -1 to 1. J-M Yoon (2013) reported that native speakers' ratings for *wh*-island sentences in reading and listening contexts ranged from 0.33 to 0.6 out of 1. In Hahn (2015b), native speakers' ratings for various *wh*-island constructions ranged from 2.0 to 2.46 on a 6-point Likert scale (where 1 was 'very bad' and 6 was 'very good'). This being the case, we have reason to ask why these sentences are acceptable to only a subset of native speakers, and further, whether the disagreement regarding acceptability is due to individuals' different grammars or to some other cognitive factors such as their variable working memory (WM) resources (Belletti and Rizzi 2013; Hofmeister and Sag 2010; Hofmeister et al. 2013; O'Grady 2010; Sprouse, Wagers and Phillips 2012a, 2012b).

The numerous acceptability judgment studies that have involved native speakers have mostly focused on their overall judgment patterns. The role of native speakers in these studies has been by and large restricted to that of the norm group. More recently, however, individual variation in acceptability and factors underlying the variation are increasingly gaining attention in the field of experimental syntax and psycholinguistics (Hofmeister and Sag 2010; Hofmeister et al. 2012a, 2012b, 2013;

Sprouse et al. 2012a, 2012b; Staum Casasanto et al., 2010). Hofmeister and colleagues, among others, attempted to account for the disagreement in sentence acceptability among native speakers in terms of their varying memory spans (Hofmeister, Staum Casasanto and Sag, 2013; Staum Casasanto, Hofmeister and Sag, 2010). Their special focus was on individuals' judgments for complex and hard-to-process sentences, which are expected to tax high processing costs. Hofmeister et al. (2013) found that complex but grammatical sentences tended to receive lower ratings from English native speakers with low memory span than from those with high memory span. In Staum Casasanto et al. (2010), low-span comprehenders were found to judge ungrammatical sentences as more acceptable than high-span comprehenders did.

Given that island sentences with a long-distance scrambling are complex to process, it is quite presumable that individuals' perceived acceptability of Korean *wh*-island constructions is affected by their WM capacity. However, variation across individuals, especially in relation to their WM, has not yet been the subject of systematic investigation in the Korean language context. While there exist some studies on the relationship between L2 learners' acceptability judgments and their WM (for example, E Kim and M-K Park's (2016) study on Korean learners of English), studies that address individual variation among "Koreans as native speakers" are hard to find. The current paper explores native speaker variation in acceptability judgments, with a special focus on *wh*-island constructions in Korean, whose acceptability is highly controversial among its native speakers. To be specific, it addresses the variations in Korean native speakers' acceptability judgments for various *wh*-island constructions from the perspective of individual WM differences.

## 2. Theoretical Background

### 2.1. Working memory, sentence processing, and acceptability

WM consists of the attentional resources used for storing and manipulating information, and so is a workspace that temporarily holds incoming data and performs computational operations while the data are available (Baddeley 2003; Cowan 2010; Daneman and Carpenter 1980; King and Just 1991). As WM is limited in its capacity, a heavy task requirement beyond an individual's WM capacity has to lead to the deterioration of performance quality. In terms of sentence processing,

long and/or syntactically complex sentences require excessive memory resources, causing poorer performance such as lowered accuracy in comprehension and slower processing time (Grodner and Gibson 2005; King and Just 1991; B-t Lee, K-J Kim and M-h Zoh 1996). This limited nature of WM thus forces the parser to seek resource-saving processing options among multiple alternatives and to avoid complex, resource-consuming options (Aoshima, Philips and Weinberg 2004; Gibson 1990; H-r Hahn 2015c; Kimball 1973; King and Just 1991; Miyamoto and Takahashi 2003).

Given that our WM limitations constrain our language processing, it is likely that processing limitations in turn constrain how complex a sentence can be, or more specifically, how far an element can be displaced from its argument position, and still be communicated and thus accepted as part of human grammar. If a deeply embedded phrase is moved too far from its original position, the displaced phrase (i.e., the filler) will decay before the parser reaches the gap site, as memory resources will run out while processing the intervening materials (Grodner and Gibson 2005; O'Grady 2008; Roberts and Gibson 2002). In other words, human grammar will generally resist structures that charge a processing cost above some threshold (Hofmeister and Sag 2010; Hofmeister et al. 2013; O'Grady 2008).

Hofmeister and Sag (2010) investigated whether such limitations in cognitive resources can account for the *wh*-island effects in English. As noted by Hofmeister and Sag, processing *wh*-island sentences requires not only holding the filler in mind while parsing intervening discourse referents between the filler and its gap, but also crossing a clausal boundary and computing interrogative clauses, which are all costly (Frazier and Clifton 1989; Ginzburg and Sag 2000; Hofmeister and Sag 2010; Kluender 1998). Processing complexity might thus provide a partial explanation for why English grammar rules out *wh*-island constructions.

Importantly, Hofmeister and Sag (2010) noted that the island effects are not equally strong across all *wh*-island constructions. According to studies in theoretical and experimental linguistics, some sentences with an island violation are better accepted than others by English native speakers (Belletti and Rizzi 2013; Chomsky 1962; Kluender 1991, 1992, 1998; O'Grady 2010; Pesetsky 1987, 2000). The acceptability contrast shown in (2a-b), is a good example.

- (2) a. \* What do you wonder who would write?
- b. ?? Which book do you wonder who would write?

Following Pesetsky (1987), the *which*-N sequence like *which book* is D(iscourse)-linked and is perceived as more natural than a bare *wh*-pronoun. One of the syntactic accounts for the observed acceptability contrast centers on the intervention effects (Belletti and Rizzi, 2013; Rizzi, 2001). Belletti and Rizzi (2013) employed the notion of Relativized Minimality, which states, informally, that in A' dependency, a filler cannot be associated with its gap when there intervenes an element that “fully matches the relevant feature specification” of the filler (p. 296). In (2a), the filler *what* and the intervener *who* have the identical feature specification [+Q], and thus the sentence violates Relativized Minimality. By contrast, (2b) does not as severely violate the Relativized Minimality, because the filler *which book* has the feature specification [+Q, +NP] while the intervener *who* is only specified by [+Q]: the match is incomplete.

From a WM point of view, these intervention effects can be easily translated into processing difficulty. Gordon, Hendrick and Johnson (2001, 2004) proposed that “a dependency is hard to compute when an element intervenes which is similar to the target of the dependency” (cited in Belletti and Rizzi 2013, p. 299). As both *wh*-phrases in (2a) are bare *wh*-pronouns whereas the filler *wh*-phrase in (2b) contains additional lexical material other than the *wh*-word itself, the effect of the intervener *who* is expected to be stronger in (2a) than in (2b). Besides, the more semantically-rich *which*-N is likely to better resist decay due to the initial deeper processing, and therefore, to be better retrieved later at the gap site than the bare *wh*-pronoun (Diaconescu and Goodluck 2004; Hofmeister and Sag 2010; Hofmeister et al. 2013; Kluender 1998).

In order to verify whether the graded acceptability of island sentences is indeed related to their graded processibility, Hofmeister and Sag (2010) constructed two types of *wh*-island items that varied in terms of the complexity of the filler phrase (bare *wh*-pronoun vs. *which*-N) as exemplified in (3a) and (3b) (p. 394).

- (3) a. Who did Albert learn whether they dismissed after the annual performance review? (*who/whether*)  
b. Which employee did Albert learn whether they dismissed after the annual performance review? (*which*-N/*whether*)

The researchers then measured native English speakers' reading times at different regions of the sentences as well as their acceptability ratings. The reading time analysis showed a significant difference in reading times at the critical region (i.e.,

the embedded verb *dismissed*) between these conditions, with shorter reading time in the *which-N/whether* condition. In their acceptability judgment test as well, the *which-N* constructions received a higher rating than the *who* construction. The authors took the parallel between processing times and acceptability judgments as evidence that acceptability decreases with the increase of processing burden. Given that processing difficulty comes from limited memory resources, the above study essentially suggests that the unacceptability of some complex sentences is in fact a WM effect.

## 2.2. Individual WM and acceptability judgment

Importantly, individuals have different WM capacity. Therefore, their differences in WM are likely to affect various aspects of linguistic performance (Daneman and Carpenter 1980; Just and Carpenter 1987; King and Just 1991; Roberts and Gibson 2002). King and Just (1991) investigated whether individual WM interacts with the processing complexity of a sentence. In their study of adult native speakers' comprehension of relative clauses, King and Just asked native English speakers to read sentences containing two types of center-embedded relative clauses (RCs): subject RCs and object RCs. They hypothesized that processing object RCs would require more WM resources than processing subject RCs, as the former requires more complex processing (such as associating two noun phrases with their respective predicates and assigning conflicting thematic roles to the head noun). They predicted that this kind of heavy demand on WM would separate high-span readers from low-span readers in both comprehension accuracy and reading speed. As predicted, low-span readers were found to comprehend object RC constructions more slowly and less accurately than high-span readers.

Similar findings have been reported for Korean native speakers' processing of relative constructions. B-t Lee et al. (1996) found that a difference between high-span and low-span readers emerged with an increase of processing complexity. In their self-paced reading experiment, native Korean-speaking adults were asked to read sentences containing different types of center-embedded RCs—OSV/SO and SOV/OO constructions. While no significant difference was found in reading times between the high- and low-span groups for the less demanding SOV/OO constructions, a significant difference was found between the two groups for the more demanding OSV/SO constructions. The findings lend additional support to King and Just's (1991) claim that the effect of individual memory differences on sentence processing

emerges more often in the processing of complex sentences.

If processibility is essentially a matter of WM, and if acceptability judgments of English island sentences can vary for processing reasons, then we have reason to ask if some cognitively demanding sentences are more processible, and thus more acceptable, to high-span readers, but less processible and less acceptable to low-span readers.

Recent experimental studies indeed have addressed whether individuals' WM differences influence their acceptability judgments (Hofmeister et al. 2012a, 2012b, 2013; Sprouse et al. 2012a, 2012b; Staum Casasanto et al. 2010). Sprouse et al.'s (2012a) experiments on WM and island constructions are of particular relevance. They reasoned that memory-based accounts of acceptability judgment would predict a correlation between individuals' WM and their perceived island effects. In order to measure the island effects, the researchers asked their English-speaking participants to rate the acceptability of four types of sentences, with filler-gap distance (i.e., extraction from either a matrix or an embedded clause) and embedded clause structure (i.e., island or non-island) as variables. Individuals' perceived island effects were then calculated based on their acceptability ratings for the four types of constructions (See Sprouse et al. 2012a, p. 92 for details). In addition, the researchers measured the participants' individual memory spans using a serial recall task and an n-back task. Their regression analyses of the two sets of data (i.e., the participants' acceptability ratings and their memory spans) led them to conclude that the effect of individual WM on acceptability judgments for island-violating sentences is statistically insignificant, and further, that island effects in English are the reflection of mental grammar rather than limited processing resources.

Sprouse et al.'s findings were followed by a series of arguments and counterarguments. Hofmeister et al. (2012a, 2012b, 2013) raised questions as to Sprouse et al.'s (2012a) research method and data interpretation. WM measurement was one of the issues. Hofmeister et al. were skeptical about the adequacy of serial recall and n-back used as verbal WM measurements by Sprouse et al. (2010). They claimed that serial recall and n-back tasks measure short-term memory rather than verbal WM, and further, argued for more valid measurements such as reading span. Sprouse et al. (2012b) contended that their memory tasks were as valid as reading span because serial recall and n-back show strong correlations with other memory tasks including reading span, citing (Conway et al. 2005). In reply, Hofmeister et al. (2012b) claimed that no such correlations had been reported in Cowan (2015) between reading span and the memory tasks used in Sprouse et al. (2012a). The

above dispute on how to measure WM suggests that caution is needed in choosing proper verbal WM measurements when examining any possible involvement of WM in acceptability judgment.

Another issue raised by Hofmeister et al. (2012a, 2012b) centers on the difficulty of the stimuli sentences. To be specific, Hofmeister et al. suggested that the lack of correlation between acceptability judgment and WM can be attributable to the extreme difficulty of the stimuli sentences used by Sprouse et al. (2012a). There is indeed some evidence that extremely complex sentences cause processing breakdown even for high-span readers, in which case, acceptability ratings do not simply increase or decrease in proportion to WM capacity (Hofmeister et al. 2012a, 2012b, 2013). Hofmeister et al. (2013) manipulated the processing complexity of grammatical *wh*-questions in terms of the length of the dependency of the fronted *wh*-phrase (short vs. long dependency) and the intervening relative clause type (subject RC vs. object RC). While neither of these constructions violated any island constraints, high-span participants performed better only when judging the “reasonably difficult” short-dependency constructions; in the extremely complex long-dependency condition, no relationship was found between WM estimates and acceptability judgments. Given that there is a certain range of processing difficulty within which memory differences emerge in the form of acceptability judgment differences, as suggested by Hofmeister and colleagues, acceptability judgment test items need to be carefully constructed so as to ensure that the items are not too complex to process even for high-span participants.

Also important to note is that the effect of individual memory span can differ depending on whether the sentence is grammatical or ungrammatical. Staum Casasanto et al. (2010) examined how people with different WM capacity judged ungrammatical sentences on one hand and complex but grammatical sentences on the other. The results showed that subjects’ WM affected their acceptability judgments differently depending on the sentence type: when judging ungrammatical sentences, high-span readers judged them as worse than low-span readers did, but when rating difficult-to-process but grammatical sentences, high-span readers judged them as more acceptable than low-span readers did. Their findings allow us to make different predictions regarding Korean *wh*-island constructions, which is our main concern. That is, Korean native speakers will show different acceptability judgment patterns, depending on whether the target sentence is grammatical or not. If Korean is not subject to *wh*-island effects, its *wh*-island constructions, which are complex to process but grammatical, will be judged as more acceptable by high-span readers

than by low-span readers. Conversely, if *wh*-island effects are at work in Korean, the same *wh*-island constructions, which are now ungrammatical, will be judged as worse by high-span readers than by low-span readers.

To summarize, researchers are beginning to query whether acceptability is a WM phenomenon, and by extension, whether acceptability judgment is affected by individual WM differences. However, little research thus far has investigated this matter in the East Asian language context. Further, considering the lack of consensus among native speakers on the grammatical status of Korean/Japanese *wh*-island sentences, as we will discuss in 2.4, the involvement of individuals' WM in their acceptability contrast is indeed an interesting research question.

### 2.3. Korean *wh*-island constructions

Scrambling in East Asian languages such as Korean and Japanese is known to be different in nature from the syntactic movement observed in European languages such as English, as the former allows much more flexible word order than the latter. Further, it has been argued that these East Asian languages allow long-distance scrambling that is not subject to some island effects (DH Chung 1996, 2005; Y-S Kang 1986; R-H-Y Kim, 2003; Y-S Lee 1993; Miyagawa 2005; H-K Park 2010; Saito 1989; H-M Sohn 1980; C Suh 1987).

Saito (1989), for example, proposed that scrambling in Japanese is an optional, semantically vacuous operation and that island sentences such as (1) above are acceptable in Japanese although a *wh*-phrase has been extracted from an embedded *wh*-clause.

Likewise, the corresponding Korean sentences such as (4) are considered to be acceptable by many syntacticians (DH Chung 1996, 2005; Y-S Kang 1986; H-K Park 2010; H-M Sohn 1980; C Suh 1987).

- (4) Muwes-ul<sub>i</sub> John-i [<sub>WH-ISD</sub> Mary-ka t<sub>i</sub> sass-nun-ci] alkosipehanta.  
what-ACC John-NOM Mary-NOM \_ bought-Q want-to-know  
'John wants to know what Mary bought.'

Note, however, that sentences like (1) and (4) do not actually violate Subadjacency per se, given that the *wh*-phrase moves to the sentence initial position via an intermediate step, as suggested by Miyagawa (2005). A true case of *wh*-island violation then occurs when the embedded *wh*-phrase (or a lexical NP) is forced to be fronted

directly to the sentence-initial position, crossing another *wh*-phrase as in (5).

(5) What<sub>i</sub> does John wonder [<sub>CP</sub> where<sub>j</sub> Mary bought t<sub>i</sub> t<sub>j</sub>]?

A corresponding sentence in Korean would look like (6), which contains a *wh*-phrase moved over another *wh*-phrase:

(6) *Muwes-ul*<sub>i</sub> John-i [*eti-eyse*<sub>j</sub> Mary-ka t<sub>j</sub> t<sub>i</sub> sassnun-ci] alkosipehani?  
what-ACC John-NOM [where Mary-NOM \_ \_ bought-Q] want-to-know-Q  
'?What does John want to know where Mary bought?'

Yet, even if sentences like (6) have a linear ordering parallel to (5), it is still arguable whether the embedded clause constitutes a *wh*-island in Korean as it does in English. Indeed, there has been a long-standing dispute as to the syntactic properties of Korean (clause-internal and clause-external) scrambling over the subject (J-H Cho 1994; R-H-Y Kim 2003; H Ko 2018; Y-S Lee 1993; Miyagawa 2005; H-K Park 2010; Saito 1989; Ueyama 2008). In her overview of previous findings on Korean scrambling, H Ko (2018) proposed that clause-internal and external scrambling cannot be characterized in terms of standard A/A' distinction. Further, clause-internal and external scrambling across the subject has often been analyzed as a case of adjunction rather than a regular *wh*-movement to [Spec, CP] (J-H Cho 1994; Y-S Lee, 1993).

If the scrambling of the *wh*-phrase *eti-eyse* over the embedded subject in (6) is an instance of adjunction, as suggested by Y-S Lee (1993) or J-H Cho (1994), then it does not target the [Spec, CP] of the embedded clause as a landing site, and accordingly, it does not block another *wh*-phrase *muwes-ul* from moving via this intermediate [Spec, CP] to the sentence-initial position. The fronted *muwes-ul*, in turn, may have been scrambled over the matrix subject via adjunction to IP and then to CP (or TopicP), or in some way other than typical A'-movement. In short, the movement of *muwes-ul* does not necessarily lead to Subjacency violation, nor does the embedded clause led by a *wh*-phrase as in (6) necessarily constitute a *wh*-island *per se*. While acknowledging such differences between Korean *wh*-scrambling and English *wh*-movement, for ease of discussion, we will continue using the term “*wh*-island” in referring to an embedded *wh*-clause in Korean from which an element is extracted.

## 2.4. Differences in acceptability judgments of Korean *wh*-island sentences

A related problem, which is the main focus of the present paper, is that these so-called *wh*-island sentences do not meet with universal acceptance. Saito (1989) accepted sentences containing a single *wh*-phrase as in (1) only with some reservations, marking it with a question mark, as Miyagawa (2005) notes. Miyagawa himself judged the same sentence as “quite acceptable” (p. 195) in that the syntactic movement in (1) does not violate Subjacency *per se*. Others have claimed that dependency cannot enter into an embedded *wh*-island, arguing for *wh*-island effects in Japanese (Nishigauchi 1990; Watanabe 1992).

The same dispute is found among scholars of Korean syntax. While many have suggested that scrambling in Korean as in (4) is not subject to *wh*-island effects (DH Chung 2005; Y-S Kang 1986; H-M Sohn 1980; C Suh 1987), others have argued that scrambling out of a *wh*-island is unacceptable (J-I Han 1992; S-H Hong 2004; H-S Lee 1982). In the midst of the ongoing dispute among linguists, several recent experimental studies have attempted to explore how *wh*-island constructions in Korean are perceived by native-speakers with no training in formal syntax (H-r Hahn 2015a, 2015b; B Kim and Goodall 2014; J-M Yoon 2013). Based on their participants' accessibility ratings, these studies reported Korean native speakers' general tendency to give low ratings to *wh*-island sentences (near -0.4 on a scale from -1 to 1 in B Kim and Goodall (2014) and 2.83 on a 6-point Likert scale in H-r Hahn (2015a)).

H-r Hahn (2015b) examined how Korean native speakers judge the acceptability of *wh*-island sentences with an intervening *wh*-phrase as in (7), as opposed to those with no intervening *wh*-phrase as in (4).

- (7) *Muwes-ul<sub>i</sub> John-i [eti-eyse<sub>j</sub> Mary-ka t<sub>j</sub> t<sub>i</sub> sassnun-ci]* alkosipehanta.  
what-ACC John-NOM [where Mary-NOM \_ \_ bought-Q] want-to-know  
'John wants to know what Mary bought where.'

She found that native speakers judged sentences with scrambling from *wh*-fronted embedded clause like (7) as significantly worse than their counterparts without an intervening *wh*-phrase like (4). H-r Hahn's (2015b) findings demonstrate that the acceptance of *wh*-island sentences might vary depending on what type of *wh*-island sentence is being measured.<sup>1)</sup> By extension, it is also presumable that people may

1) As a reviewer noted, different degrees of acceptability may also ensue from different syntactic functions of the island, e.g., whether the extraction is from an clausal complement or a clausal

accept some island constructions within their memory threshold, and reject others that are beyond their processing capacity. To see if any interaction between construction type and WM emerges, the present study included different types of *wh*-island constructions in its experimental design.

To summarize, native speakers show individual differences in their judgments on long-distance scrambling from a *wh*-clause. The picture is further complicated due to different *wh*-island types examined. Our question is whether the lack of consensus among native speakers reflects differences in individual WM. A pure memory-based account of the acceptability variation would predict that high-span readers would better accept island sentences, as they have sufficient memory resources for the processing. However, a reverse pattern can be borne out if the Korean grammar prohibits the movement out of a *wh*-island, in which case, high-span readers will better detect the violation, and as a consequence, they will better reject the construction. Still another possibility is that the variation is attributable to some other factors than WM, in which case, individuals' acceptability judgments will not be correlated with their memory spans. Finally, the participants' acceptability judgments can show free variation if the construction itself is extremely difficult to process. The present study will examine each of these possibilities in answering our research question: Are individuals' differences in WM responsible for their differences in acceptability judgements of Korean *wh*-island sentences?

### 3. Method

An acceptability judgment test and two memory span tests were designed to see if Korean native speakers' varied acceptability judgments on *wh*-island sentences are related to individual native speakers' WM differences. The acceptability judgment test included *wh*-island sentences that varied in their expected processing difficulty. Reading span task was used as the main WM task. In addition, a digit span test was administered to see if the storage component of WM alone can also contribute to acceptability judgment.

Sixty-six adult native Korean speakers living in South Korea participated in the experiments. All of them were undergraduate students in their twenties. They each

---

adjunct. For the complement/adjunct asymmetry issues, see Y-S Lee (1993), H Ko (2018), Sprouse et al. (2012), and B Kim & Goodall (2014).

completed the three tasks in three separate sessions. They participated voluntarily, and received cash and gifts for participating.

### 3.1. Acceptability judgment task

#### 3.1.1. Materials

All test items contained a *wh*-phrase originating from an interrogative embedded clause with a Q-particle, *-ci*. The items included three types of sentences presented in either the scrambled or the in-situ condition. Type 1 sentences in the scrambled condition contained two *wh*-phrases (WH1 & WH2) originating from the embedded interrogative clause, and had a structural pattern identical to (6). One of the *wh*-phrases (WH1) was placed in the left periphery of the embedded clause, and the other (WH2) was scrambled from the embedded object position to the sentence-initial position, crossing WH1 (WH/WH scrambling, henceforth).

WH/WH scrambling:       $WH2_i \cdots [_{CP} WH1_j \cdots t_j t_i \cdots Q] \cdots$

Type 2 scrambled constructions contained an accusative NP extracted from its base-generated embedded object position to the sentence-initial position, crossing a *wh*-phrase that leads the embedded clause (NP/WH scrambling, henceforth).

NP/WH scrambling:       $NP_i \cdots [_{CP} WH_j \cdots t_j t_i \cdots Q] \cdots$

While Type 2 also involves movement across an intervening *wh*-phrase, the intervening *wh*-phrase does not match the features of the NP filler as they do not share [+Q]. Therefore, the intervention effect in Type 2 was expected to be smaller than in Type 1.

The Type 3 scrambled construction had a structural pattern identical to Saito's (1989) sentence in (1), containing a single *wh*-phrase, which has been scrambled to the sentence-initial position out of its base-generated object position in the embedded interrogative clause, but not crossing an intervening *wh*-phrase (WH/∅ scrambling, henceforth).

WH/∅ scrambling:                       $WH_i \cdots [_{CP} \emptyset \cdots t_i \cdots Q] \cdots$

As no *wh*-phrase appears at the left edge of the embedded clause in Type 3, the processing costs in terms of intervention is expected to be the smallest among the three types. While WH/Ø scrambling does not contain a *wh*-intervenor), the experimental design included it because the embedded clause has a [+Q] feature marked by the question particle *-ci* and thus is differentiated from non-island sentences in English such as “What did you say that he saw?” Moreover, its grammaticality is highly controversial (Aoshima, Philips and Weinberg 2004; H-r Hahn 2015a, 2015b; H-r Hahn and S Hong 2014; Miyagawa 2005; Saito 1989, 2004; J-M Yoon 2013), and may present an interesting case of graded acceptability.

The fronted element in all three types of scrambling was restricted to an accusative argument to ensure that the participants would not interpret the fronted phrase as the argument of the matrix verb but as that of the embedded verb. (See Aoshima et al. 2004, H-r Hahn 2015b, and Miyamoto and Takahashi 2003 for details.) Fourteen island-related items—six items in Type 1, four in Type 2, and four in Type 3—were constructed. In addition to the island constructions in the scrambled condition, matching constructions in the in-situ condition were also constructed. Sample test items in the scrambled condition are presented in (8–10), alongside their in-situ counterparts.

(8) Type 1

Scrambled Condition (WH/WH)

*Muwsun os-ul sahoeca-nun [eti-eyse kasuw-ka ipul-ci] kuwngkumhayss-ta.*  
 which dress-ACC show.host-TOP [where-LOC singer-NOM wear-Q]  
 wondered-DECL

In-situ Condition

*Sahoeca-nun [kasuw-ka eti-eyse muwsun os-ul ipul-ci] kuwngkumhayss-ta.*  
 show.host-TOP [singer-NOM where-LOC which dress-ACC wear-Q] wondered-  
 DECL

‘The show host wondered where the singer would wear which dress.’

(9) Type 2

Scrambled Condition (NP/WH)

*Dress-nul sahoeca-nun [eti-ese kasuw-ka ipul-ci] kuwngkumhayss-ta.*  
 gown-ACC show.host-TOP [where-LOC singer-NOM wear-Q] wondered-  
 DECL

In-situ Condition

Sahoyca-nun [kasuw-ka *eti-eyse dress-rul* ipul-ci] kuwngkumhayss-ta.  
Show.host-TOP [singer-NOM where-LOC gown-ACC wear-Q] wondered-DECL

‘The show host wondered where the singer would wear the gown.’

(10) Type 3

Scrambled Condition (WH/Ø)

*Muwsun os-ul* sahoeyca-nun [kasuw-ka pammuwatay-eyse ipul-ci] kuwngkumhayss-ta.  
which dress-ACC show.host-TOP [singer-NOM night.show-LOC wear-Q]  
wondered-DECL

In-situ Condition

Sahoyca-nun [kasuw-ka pammuwatay-eyse *muwsun os-ul* ipul-ci] kuwngkumhayss-ta.  
Show.host-TOP [singer-NOM night.show-LOC which dress-ACC wear-Q]  
wondered-DECL

‘The show host wondered which dress the singer would wear at the late night show.’

The items in the two conditions were distributed in two lists using a Latin-square design to ensure that each participant encountered each item in only one of the two conditions (scrambled vs. in-situ).

In addition to the three types of island sentences, three completely ungrammatical sentences were included in the test items. These sentences erroneously contained two accusative arguments associated with a single embedded predicate, one fronted and the other in situ, as in (11).

- (11) \*Myengphuwmpak-ul kokayk-un cemwon-i muwes-ul cinyelhayssnun-ci  
muwlepoass-ta.  
designer.bag-ACC customer-TOP salesclerk-NOM what-ACC displayed-Q  
asked-DECL  
‘\*The customer asked what the salesclerk displayed the designer bag.’

We call this type of sentence a “filled-gap” sentence, as the gap site for the fronted filler phrase is illegally occupied by an overt argument, violating the Theta theory as well as the Case theory. As this construction is structurally similar to the scrambled *wh*-island constructions except that the embedded gap position is filled,

it was expected that the constructions would involve a similar degree of processing difficulty. Moreover, as a clearly ungrammatical sentence type, the filled-gap sentences were predicted to be judged as worse by high-span readers than low-span readers, as long as (i) individual WM makes difference in acceptability judgments, and (ii) the construction was not extremely difficult to process. The seventeen test items (fourteen island-related sentences and three filled-gap sentences) were interspersed with twenty-six filler items, and the presentation order was randomized.

### 3.1.2. Procedure

Each sentence was presented as a single line projected on a computer screen. A new item was introduced by a short clicking sound, stayed for twelve seconds, and then disappeared as the next sentence appeared. The participants were asked to read the sentence on the screen and rate its acceptability on a 6-point Likert scale (1=very bad, 2=bad, 3=somewhat bad, 4=somewhat good, 5=good, 6=very good) while the sentence remained on the screen.<sup>2)</sup>

## 3.2. Reading span test

### 3.2.1. Materials

The study employed reading span as a major tool to assess the participants' WM capacity. As discussed in Section 2.2, Reading span has been employed in many language processing studies and is considered highly predictive of various aspects of language performance, especially on-line and off-line language processing (Daneman and Carpenter 1980; Farmer et al. 2017; Hofmeister et al. 2012b, 2013; King and Just 1991; B-t Lee 2002). This study employed a Korean reading span test using items selected from the reading span test developed by B-t Lee (2002),

---

2) Some potential problems of using of an n-point numerical scale include whether subjects actually perceive the size of the difference between 2 and 3 as equal to that between 5 and 6, or whether it yields sufficient distinctions (i.e., unavailability of gradient rating such as 4.5 (Y-h Lee & Y Park 2015). However, recent experimental studies have demonstrated that numerical scale yields as reliable results as other recent alternatives such as a magnitude estimation (ME) task (Bader and Häußler 2010; Fukuda et al. 2012). In Fukuda et al.'s (2012) experimental study on different types of acceptability rating (Yes/No vs. n-point scale vs. ME), n-point ratings yielded surprisingly similar results to ME. As ME can be unfamiliar and difficult for subjects to understand and requires "some mathematical sophistication" (Fukuda et al. 2012: 328), and as Likert scale measurement is found as effective, the present study chose Likert scale for its acceptability measurement.

which is a Korean equivalent of Daneman and Carpenter's (1980) reading span test. Following B-t Lee, the current test contained five sets each of two, three, four, five, and six unrelated sentences, together constituting a total of 25 items. Each sentence contained seven to eight content words on average (plus case markers and suffixes), and mostly ended with a noun suffixed with a declarative marker, as in *yuwchaykkot-ita* (canola.flower-Decl). The sentence-final words (the target words to be recalled) within a single item set were all different and were not semantically associated with one another. A sample three-sentence item (S1-S2-S3) used in the test and the correct response are presented in (12).

(12) [S1] Oman-kwa toksen-un chengsonyen-ul sentoha-nun saramtul-ui kocilcekin pyengphye-ita.  
 pride-and self-righteousness-TOP youth-ACC guide-that people-POSS chronic ill-DECL  
 'Pride and self-righteousness are the chronic ill observed in the people who guide the youth.'

[S2] Cecuwdo-rul sayngkakhmyen kacang mence tteoru-nun ces-i yuwchaykkot-ita.  
 Jeju.Island-ACC think-when the first occur-that thing-NOM canola.flowers-DECL  
 'The first thing that comes to mind when you think of Jeju Island is canola flowers.'

[S3] Ipancekici anhta-nun ces-i saramtul-i say ukyen-ey pantayha-nun iyuw-ita.  
 common NEG that-NOM people-NOM new idea-with disagree-that reason-DECL  
 'That it is not common is the reason why people disagree with a new idea.'

Correct Response: *pyengphye-ita, yuwchaykkot-ita, iyuw-ita*

### 3.2.2. Procedure

The reading span test items were presented on a computer screen. The participants were instructed to read aloud a series of sentences, and then to recall all the last words of the sentences they had just read. They were given two practice trials (a two-sentence item and a three-sentence item) before the actual test. In the case that the participant did not clearly understand the procedure, the practice trials were repeated. The test was administered individually in a quiet place.

Each sentence appeared in a single line across the center of the screen, one sentence at a time, and the participants proceeded to the next sentence by pressing the space bar. Upon reading a set of sentences, the next key press led the participants to the instruction to recall all the last words of the read sentences in the item. A cross signal appeared on the screen before each trial. The participants started with two-sentence items and moved to increasingly larger sets. The experiment terminated when an individual failed in three trials out of five at a given level.

### 3.3. Digit span test

In addition to the reading span test, a forward digit span test was administered to see if the storage component of working memory (or short-term memory) alone also affects acceptability judgments. The test included three sets each of 4-to-10 digits. The participants saw series of digits presented on the computer screen, beginning with four-digit items, and moving to increasingly longer digit strings, up to ten digits. The numbers were presented one digit per second, and each digit was replaced by the upcoming digit. The participants were asked to read them in a small voice as they appear, and to recall them in their presented order after each digit string.

### 3.4. Data analyses

The participants' acceptability judgment scores for the different types of island constructions were compared to see if their judgments were sensitive to island sentence types. They were also compared with the acceptability judgments for the filled-gap constructions to see if *wh*-island sentences are differentiated from similar but clearly ungrammatical sentences.

Reading span was scored in two ways. First, span was defined as the largest number of sentences all of whose final words were recalled correctly on at least

three out of five trials (rounded Rspan, henceforth), following B-t Lee (2002) and King and Just (1991). For example, if a participant was correct on four trials out of five at the two-sentence level, on three trials at the three-sentence level, and on two trials at the four-sentence level, the individual's rounded Rspan was 3, as the three-sentence level was the highest level at which the participant successfully provided the correct responses on at least three trials.

To compensate for the rounded Rspan's insensitivity to differences between those who correctly recalled the final words on all five trials and those who managed to recall them on only three or four out of five trials at a given level, a second measure of reading span was used. This second measure counted the total number of items (i.e., trials) in which the last words of all the sentences in the set were correctly recalled (item-based Rspan, henceforth). For example, if a participant was correct on four trials at the two-sentence level, three trials at the three-sentence level, and two trials at the four-sentence level, the individual's item-based Rspan was 9 (4+3+2).

Digit span was operationalized as the longest string of digits recalled correctly on two of the three trials. For example, if a person correctly recalls all three given strings of digits in the six-digit items, two strings in the seven-digit items, and one string in the eight-digit items, the person's digit span was "7."

A Pearson's correlation analysis was conducted with the participants' acceptability judgment ratings and the reading span and digit span estimates in order to see if individuals' acceptability judgment is related to their memory span. To further verify whether WM is related to or separate from individuals' acceptability judgment, a factor analysis was run.

## 4. Results

### 4.1. Acceptability judgment and individual variation

Overall, the participants were found to give generally low ratings to the island sentences in the scrambled conditions, although they also displayed a wide range of individual variation. As can be seen in Table 1, all three types of *wh*-island constructions in the scrambled condition were rated below 2.5 on average on the 6-point Likert scale, where the neutral value was 3.5. The generally negative judgments on these island constructions replicates the previous findings from native

speaker acceptability judgment experiments on Korean *wh*-island constructions (H-r Hahn 2015a, 2015b; H-r Hahn and S Hong 2014; B Kim and Goodall 2014; J-M Yoon 2013). The mean scores in the three scrambled conditions showed a stark contrast with those in the in-situ conditions, which all received high ratings above 5. The difference between the scrambled conditions and the in-situ conditions was highly significant in all three types ( $t=27.294$ ,  $p<.001$  in Type 1;  $t=27.144$ ,  $p<.001$  in Type 2;  $t=23.594$ ,  $p<.001$  in Type 3).

**Table 1.** Acceptability judgment ratings on *wh*-island and filled-gap constructions

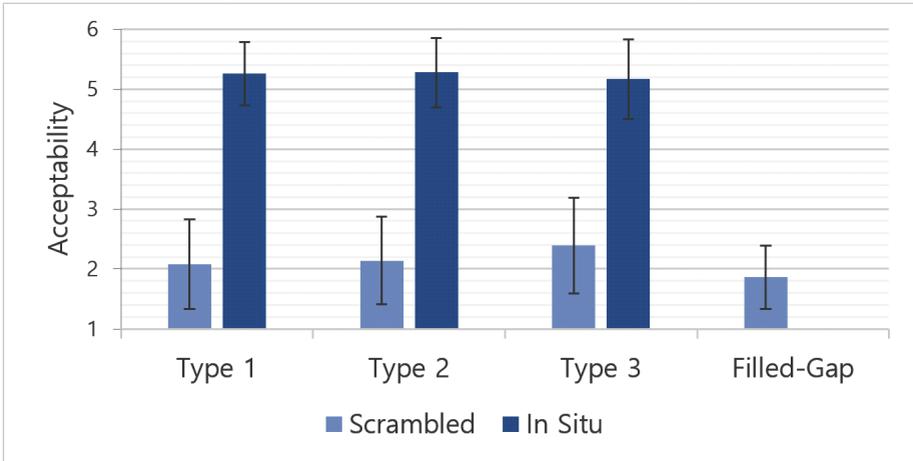
		Mean	SD	N
Type 1	Scrambled (WH/WH)	2.08	0.75	66
	In-situ	5.26	0.53	66
Type 2	Scrambled (NP/WH)	2.14	0.73	66
	In-situ	5.28	0.58	66
Type 3	Scrambled (WH/Ø)	2.39	0.80	66
	In-situ	5.17	0.66	66
Filled-Gap		1.86	0.53	66

In the scrambled conditions, Type 3 (WH/Ø) received a higher acceptability rating than Type 1 and Type 2 constructions. The difference among the three scrambled conditions was significant ( $F=5.059$ ,  $p<.01$ ). Pair-wise comparisons revealed that the difference was significant between Type 1 and Type 3 ( $p<.01$ ), but not between Type 1 and Type 2 or between Type 2 and Type 3, suggesting a graded differentiation among them.

While the three types of scrambled constructions received overall low ratings, they were still judged as more acceptable than the clearly ungrammatical filled-gap constructions, whose two accusative arguments were associated with a single embedded verb. The ratings for the filled-gap sentences were significantly lower than those for Type 1 (WH/WH) ( $t=2.270$ ,  $p<.05$ ), Type 2 (NP/WH) ( $t=2.943$ ,  $p<.01$ ), and Type 3 (WH/Ø) ( $t=4.839$ ,  $p<.001$ ), suggesting that long-distance extraction out of *wh*-island in Korean does not cause as serious degradedness as Case violation.

Turning now to individual differences, the participants varied in their ratings for the *wh*-island constructions. Island sentences were judged as completely unacceptable (i.e., rated “1”) by a subset of the participants (4 participants in Type 1; 8

participants in Type 2; 6 participants in Type 3). The same island sentences were judged as rather acceptable (i.e., rated higher than 3.5) by other participants (2 participants for Type 1; 5 participants for Type 2; 7 participants for Type 3). The participants' judgments in the scrambled conditions showed a larger variation than their judgments in the in-situ and filled-gap conditions, as also displayed by the error bar sizes in Figure 1 (The error bars in the figure represent one standard deviation).



**Figure 1.** Acceptability ratings: means and standard deviations.

#### 4.2. WM and acceptability judgment

Participants' reading span scores also showed a wide range of individual variation. In terms of the maximum number of sentences whose final words were all correctly recalled in more than three trials out of five (i.e., rounded Rspan), the participants' performance was just below the 3-sentence level on average ( $M=2.85$ ,  $SD=.899$ ), varying between 2 ( $n=16$ ) and 6 ( $n=1$ ). In terms of the number of items whose sentence-final words were all correctly recalled (i.e., item-based Rspan), the participants recalled all the final words correctly on 8.41 trials on average, with a minimum of 3 ( $n=2$ ) and a maximum of 21 ( $n=1$ ).

The participants digit span was 8.16 digits on average, ranging from 5 digits minimum ( $n=1$ ) to 10 maximum ( $n=10$ ). The participants' mean reading spans and digit span are presented in Table 2.

**Table 2.** Participants' reading spans and digit span

	Mean	SD	N
Rounded Rspan	2.85	.89	66
Item-based Rspan	8.41	3.63	66
Digit Span	8.16	1.22	63

Pearson correlation coefficients among memory estimates and acceptability ratings for different types of constructions are presented in Table 3.

**Table 3.** Correlations among memory spans and acceptability ratings

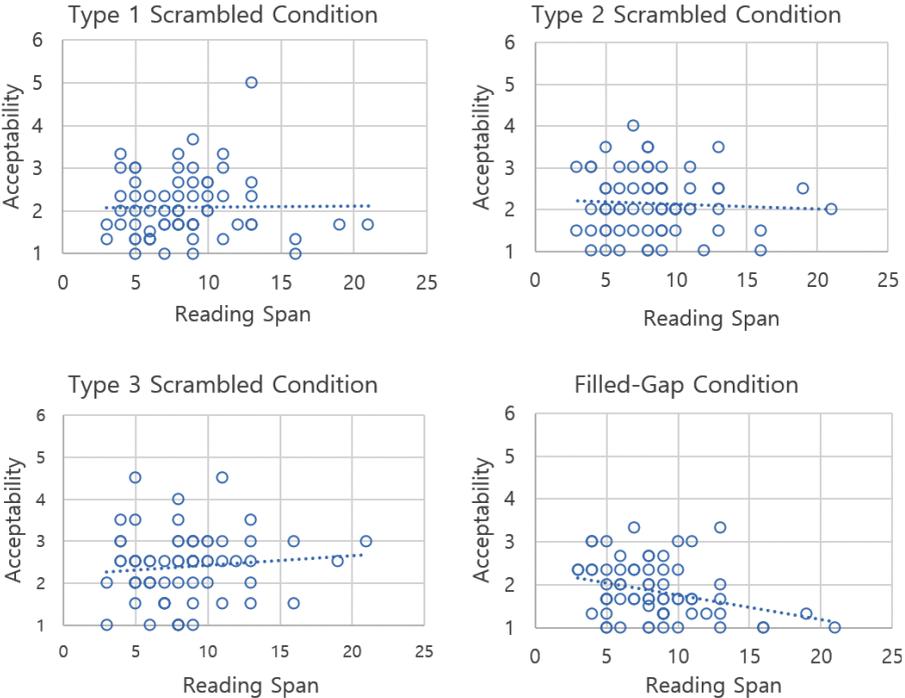
	1	2	3	4	5	6	7	8	9
1 Rounded Rspan									
2 Item-based Rspan	.936**								
3 Digit Span	.296*	.305*							
4 Type 1 Scrambled	.030	.012	-.041						
5 Type 2 Scrambled	-.107	-.054	-.154	.363**					
6 Type 3 Scrambled	.106	.103	.079	.468**	.330**				
7 Type 1 In-situ	.191	.178	.179	-.065	.121	.221			
8 Type 2 In-situ	-.021	-.045	-.016	.111	-.015	.225	.344**		
9 Type 3 In-situ	-.046	-.056	.167	.147	.155	.123	.445**	.259*	
10 Filled-gap	-.316**	-.316**	-.297*	.362**	.370**	.250*	.002	.085	.277*

Note. \* indicates  $p < .05$ ; \*\* indicates  $p < .01$

As shown, the participants' two Rspan measures (rounded Rspan and item-based Rspan) were in almost complete correlation ( $r = .936$ ,  $p < .001$ ). There was also a moderate degree of correlation between rounded Rspan and digit span ( $r = .296$ ,  $p < .05$ ), and between item-based Rspan and digit span ( $r = .305$ ,  $p < .05$ ). The acceptability ratings on the three types of scrambled constructions also strongly correlated with one another (all at the .01 significance level:  $r = .363$  between Types 1 & 2;  $r = .468$  between Types 1 & 3;  $r = .330$  between Types 2 & 3). However, none of the memory span measures correlated with the acceptability judgments in any of the scrambled conditions. If the Korean *wh*-island constructions were more acceptable to high-span readers, we would expect a positive correlation between individual participants'

acceptability ratings on these constructions and their memory span. Such a positive correlation was not found in the present study. Finally, the acceptability ratings on Types 1-3 in-situ constructions correlated only among themselves but not with the memory variables, which is not surprising in view of the previous findings that WM effects emerge only in the processing of complex sentences (King and Just 1991; B-t Lee et al. 1996; Staum Casasanto et al. 2010).

Interestingly, the ratings on all three types of scrambled constructions were found to correlate with the ratings of the filled-gap constructions. Also interesting to note is that the acceptability judgments of the filled-gap constructions showed a strong negative correlation with the Rspan measures. As demonstrated by the scatter plot matrices in Figure 2, the acceptability ratings of the three types of long-distance scrambling were almost evenly distributed along the acceptability scale between 1 and 4, neither increasing nor decreasing as the rater's Rspan increased. On the other hand, the acceptability ratings for the filled-gap constructions tended to decrease as Rspan increased, indicating that high-span readers performed better in rejecting these clearly ungrammatical sentences.



**Figure 2.** Reading span (item-based) and acceptability judgment: Scatter plots.

The negative correlation between the memory spans and the acceptability judgments on the filled-gap construction suggests the possibility that WM in fact contributes to the long-distance association between the filler and the gap. The finding also recalls Staum Casasanto et al.'s (2010) finding that their high-span participants judged ungrammatical sentences as worse than their low-span participants did. We will return to this issue in the discussion section.

A subsequent factor analysis with a Varimax rotation produced three factors with eigenvalues over 1. Table 4 presents the factor matrix, where loadings lower than .30 have been deleted. The remaining values of the variables all had loadings of .50 or more. As shown, Factor 1 received heavy loadings above .50 from the memory measures and the acceptability judgments of the filled-gap constructions; Factor 2 received loadings above .50 from Types 1–3 scrambled island constructions and the filled-gap constructions; and Factor 3 received heavy loadings from Types 1–3 in-situ constructions. The memory-related variables exclusively loaded on Factor 1. Likewise, all the scrambled constructions exclusively loaded on Factor 2. These results further confirm that individual variation in acceptability judgments on *wh*-island constructions is separate from individual variation in WM.

**Table 4.** Factor analysis

Variable	Factor		
	1	2	3
Rounded Rspan	.932		
Item-based Rspan	.931		
Digit Span	.534		
Filled-gap	-.520	.572	
Type 1 Scrambled		.823	
Type 3 Scrambled		.751	
Type 2 Scrambled		.719	
Type 3 In-situ			.763
Type 1 In-situ			.762
Type 2 In-situ			.736

Again, it is noteworthy that one variable—the acceptability of the filled-gap construction—relates about equally to the first and second components, but not to the third component. As mentioned earlier, the ratings on the ungrammatical

filled-gap sentences negatively correlated with the memory variables, suggesting that memory is involved in identifying filled-gap constructions as ungrammatical. The first factor is thus construed to best represent “memory span.”

Factor 2 is somewhat tricky to identify. The variables that loaded on Factor 2 concerned Types 1–3 scrambled constructions and the filled-gap construction, which all received low ratings in the acceptability judgment task. As mentioned, the ratings on these less acceptable patterns showed covariance, indicating that those who graded filled-gap constructions leniently also graded scrambled constructions leniently, and vice versa. Therefore, it is reasonable to conclude that this variance rather reflects individual participants’ level of permissiveness for or tolerance of “bad” sentences. The second component is thus best captured as “tolerance to degradedness.”

Finally, the third factor is loaded with variables concerning in-situ constructions, whose processing does not demand much memory resources, as their arguments and predicates are locally associated with each other in their canonical positions. The third factor thus seems to simply represent “responses to in-situ constructions.”

## 5. Discussion

The lack of correlation between reading span and acceptability judgment on the Korean *wh*-island constructions suggests that WM differences do not affect the acceptability of these configurations. The finding thus seems to lend support to Sprouse et al.’s (2012a, 2012b) endorsement of acceptability independent of WM limitation. On the other hand, a strong negative correlation between reading span and acceptability judgment on the clearly ungrammatical filled-gap sentences can be interpreted as evidence that individual WM is indeed related to acceptability judgment, in support of Hofmeister and colleagues (Hofmeister et al. 2012a, 2012b; Staum Casasanto et al. 2010).

Despite these seemingly contradictory roles of WM, however, the native speaker participants’ judgments on the filled-gap constructions in fact give us an important clue as to the interpretation of the results concerning the island constructions. The fact that the high-span participants better rejected the filled-gap sentences indicates that they were better at storing the fronted element in memory while parsing the intervening lexical elements, and that they accurately posited the gap site, and accordingly, better noticed that the gap site was illegally occupied by some other

argument. This, in turn, suggests that the high-span participants were also better at identifying the gap position in the scrambled conditions of the experiment. Therefore, if grammar allowed long-distance scrambling out of a *wh*-clause, high-span readers would have accepted the island constructions. Conversely, if long-distance scrambling from a *wh*-clause were forbidden by grammar, the high-span participants would have better rejected the scrambled constructions, judging them as even worse than the low-span readers would (Staum Casasanto et al. 2010). Neither of these judgment patterns was observed in this study. Both high-span and low-span participants showed individual variation in acceptability judgments on these constructions, independent of their memory span.

Our data therefore are not compatible with the claim that long-distance scrambling from a *wh*-clause is legitimate in Korean grammar, and is only unacceptable to those whose WM capacity falls short of the processing demand. Neither can we make a claim that long-distance extraction from a *wh*-island is disallowed for syntactic reasons, because in such a case, individual WM is expected to negatively correlate with the acceptance of the scrambled constructions. In short, our participants do not seem to have made rule-based judgments for *wh*-island sentences.

Where grammar does not draw a clear line between acceptable and unacceptable configurations, processibility may come into play. As mentioned, there are multiple reasons why *wh*-island constructions are hard to process in East Asian languages as well as in English. From this perspective, Types 1-3 scrambled conditions are generally undesired because they make such heavy processing demands. As mentioned earlier, Type 3 (WH/Ø) should be the easiest to process among the three types because it involves only one filler-gap dependency resolution and further, there is no intervener for the dependency. Between Type 1 (WH/WH) and Type 2 (NP/WH), both of which involve two filler-gap dependencies, Type 2 should be easier to process than Type 1 because the similarity between the filler NP and the intervening *wh*-phrase is minimal; the expected intervention effects are the strongest in Type 1 in that the target and the intervener are both bare *wh*-pronouns. Indeed, the acceptability hierarchy among the three types derived from the experiment seems to reflect their processibility hierarchy:

Type 3 (WH/Ø) > Type 2 (NP/WH) > Type 1 (WH/WH)

The processing-based explanation, however, cannot be extended far enough to account for the acceptability judgment contrast in terms of individual differences in

WM, as individuals' reading spans and their acceptance of the island sentences were not correlated. The findings so far do not lend support to the hypothesis that native speakers accept long-distance scrambling from a *wh*-clause because they have larger memory capacity, or reject it because they have smaller memory capacity. A more plausible interpretation for the varying acceptability perceived by different individuals, I argue, concerns the individual's permissiveness regarding these marginal constructions. Recall that the individual native speakers in this study were generally consistent in their judgments across all three types of scrambled constructions; those who accepted (to varying degrees) Type 1 scrambling also tended to accept Type 2 and Type 3 scrambling, and vice versa. In addition, the participants' ratings of these three constructions were strongly correlated with their ratings of the filled-gap construction, which suggests that an individual tended to be permissive of the three scrambled constructions in proportion to the same individual's level of permissiveness for the filled-gap construction.

Also interesting to note is that these island sentences were judged as more acceptable than clearly ungrammatical sentences such as the filled-gap sentences and further, showed wider individual variation, so that some participants, albeit few, actually judged the scrambled constructions (especially Type 3, WH/Ø constructions) as rather acceptable, agreeing with some syntacticians (Y-S Kang 1986; Miyagawa 2005; Saito 1989; H-M Sohn 1980).

The wide variation in the acceptability of the *wh*-island constructions might be related to the native speakers' inability to make a rule-based decision on these marginal constructions as opposed to other clearly grammatical constructions, as in the in-situ sentences, or clearly ungrammatical constructions, as in the filled-gap sentences. Some discourse-related theories such as D(iscourse)-linking (Pesetsky 1987, 2000; Rizzi 1990; J-M Yoon 2013) would suggest that those who better created a discourse context for the scrambled sentences could have judged them as more acceptable; the discourse issue, however, is beyond the scope of this paper.

Another question that deserves further consideration is whether the perceived degradedness of Type 3 (WH/Ø) sentences could be accounted for in terms of long-distance scrambling alone, without employing the notion of *wh*-island effects. As mentioned, processing long-distance filler-gap dependency across a clausal boundary is costly in itself (Frazier and Clifton 1989; Hofmeister and Sag 2010; Kluender 1998), and is likely to cause a certain degree of degradedness. In order to tease apart scrambling effects from *wh*-island effects, a fourth type of sentence pair with no *wh*-island structure as in (13) needs to be added as baseline:

(13)

Scrambled Condition (NP/Ø)

*Ku os-ul sahoeyca-nun [kasuw-ka pammuwtay-eyse ipulkela-ko] allyecuwes-ta.*  
That.dress-ACC show.host-TOP [singer-NOM night.show-LOC wear-DECL]  
informed-DECL

In-situ Condition

*Sahoyca-nun [kasuw-ka pammuwtay-eyse ku os-ul ipulkela-ko] allyecuwes-ta.*  
Show.host-TOP [singer-NOM night.show-LOC that.dress-ACC wear-DECL]  
informed-DECL

‘The show host informed that the singer would wear that dress at the late night show.’

If Type 3 scrambling (WH/Ø), which showed the weakest effects among the three types of scrambling, shows significantly larger effects than non-*wh* scrambling (NP/Ø) in terms of acceptability judgement, the difference can be interpreted as additional effects due to the *wh*-island structure. However, if Type 3 scrambling does not incur any additional effects, the degradedness in the Type 3 scrambled condition should be interpreted as mere scrambling effects.

From a sentence processing perspective, Type 3 scrambling (WH/Ø) is predicted to cause lower ratings than NP scrambling (NP/Ø) in (13). For Type 3 scrambled constructions, the parser must not only resolve the dependency between the *wh*-filler and its gap but also the dependency between the *wh*-phrase and its scope (marked by the embedded Q-particle *-ci*). In contrast, processing of the scrambled construction of (13) only requires resolving the dependency between the scrambled NP filler and its gap; it does not involve *wh*-scope fixing. Therefore, Type 3 *wh*-scrambling is predicted to incur higher processing costs. Whether such prediction is borne out is an important question that needs to be asked to better understand the nature of scrambling in general as well as *wh*-island effects in Korean. The current study did not explore this more fundamental question. The focus of the present study was limited to whether WM differences can explain the observed individual differences in the acceptability judgments of *wh*-island sentences in Korean.

## 6. Conclusion

The findings of the present study can be summarized as follows. First, the acceptability of Korean *wh*-island sentences was not related to individual WM differences. The previous findings that WM differences tend to correlate with differences in judgments of complex sentences as either grammatical or ungrammatical, and the findings of the current paper that individual WM correlated with the ratings for the ungrammatical filled-gap construction but not with the ratings for the *wh*-island constructions, together suggest that *wh*-island sentences in Korean belong to a grey zone, where grammar does not provide a clear, categorical cut-off line.

Second, the lack of correlation between memory spans and the judgments of Types 1–3 scrambled sentences on one hand, and the covariance of Types 1–3 island sentences with the filled-gap sentences on the other, suggest that the individual variation in the acceptability judgements for Korean *wh*-island sentences reflects individual differences in permissiveness for structural degradedness rather than WM differences.

Finally, despite the lack of direct involvement of individual WM in the acceptance of island constructions, this should not be interpreted as entirely abandoning the processing-based account of acceptability. Korean *wh*-island sentences were generally judged as unacceptable although there was individual variation. Further, their mean acceptability gradually decreased as their processing difficulty increased in terms of intervention and the number of filler-gap dependencies.

There are some limitations that should be considered in interpreting its findings. First, it did not include discourse variables in the study's design. Some of the individual variation in acceptability might be attributable to different individual levels of activation of pragmatic context. As suggested by D-linking accounts, providing rich contexts for island sentences might bring about increased acceptability. Second, this study did not include a baseline scrambled condition, where scrambling is from a non-*wh* embedded clause. Adding such baseline conditions in future experiments would be crucial for addressing the question whether there exist *wh*-island effects in Korean that is separate from mere long-scrambling effects. Finally, this study did not measure the processing costs involved in the different types of island constructions in terms of participants' reaction times or eye-movement patterns. The graded processing costs of the different types of constructions were simply presumed based on previous studies in the literature, and thus need to be further verified through time-sensitive measurements of processing difficulty.

## References

- Aoshima, Sachiko, Colin Phillips and Amy Weinberg. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 5, 23–54.
- Baddeley, Alan. (2003). Working memory and language: an overview. *Journal of Communication Disorders* 36(3), 189–208.
- Bader, Markus and Jana Häussler (2010). Toward a model of grammatical judgments. *Journal of Linguistics* 46, 273–330.
- Belletti, Adriana and Luigi Rizzi. (2013). Intervention in grammar and processing. In Ivano Caponigro and Carlo Cecchetto, eds., *From Grammar to Meaning: The Spontaneous Logicality of Language*, 294–311. Cambridge: Cambridge University Press.
- Cho, Jai-Hyoung. (1994). *Scrambling in Korean: Crossover, Reconstruction and Binding Theory*. Doctoral dissertation. University of Connecticut.
- Chomsky, Noam. (1962). The logical basis of linguistic theory. In Morris Halle, ed., *Preprints of the Papers for the Ninth International Congress of Linguists*, 509–574.
- Chomsky, Noam. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The MIT Press.
- Chung, Dae H. (1996). *On the Representation and Licensing of Q and Q-dependencies*. Doctoral dissertation. University of Southern California, Los Angeles.
- Chung, Dae H. (2005). Why is how in Korean insensitive to islands?: A revised nominal analysis. *Studies in Modern Grammar* 39, 115–131.
- Cowan, Nelson. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science* 19(1), 51–57.
- Daneman, Meredyth and Patricia Carpenter. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450–466.
- Diaconescu, Rodica C. and Helen Goodluck. (2004). The pronoun attraction effect for D(iscourse)-linked phrases: Evidence from speakers of a null subject language. *Journal of Psycholinguistic Research* 33(4), 303–319.
- Farmer, Thomas A., Alex B. Fine, Jennifer B. Misyak, and Morten H. Christiansen (2017). Reading span task performance, linguistic experience, and the processing of unexpected syntactic events. *The Quarterly Journal of Experimental Psychology* 70(3), 413–433.
- Frazier, Lyn and Charles Jr Clifton. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4, 93–126.
- Fukuda, Shin, Grant Goodall, Dan Michel, and Henry Beecher. (2012). Is magnitude estimation worth the trouble? In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, 328–336. Somerville, MA: Cascadilla Proceedings Project.
- Gibson, Edward. (1990). A computational theory of processing overload and garden-path effects. In *Proceedings of the 13th Conference on Computational Linguistics*, 114–119. Morristown, NJ: Association for Computational Linguistics.

- Ginzburg, Jonathan and Ivan A. Sag. (2000). *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford, CA: CSLI Publications.
- Gordon, Peter C., Randall Hendrick, and Marcus Johnson. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition* 27, 1411-1423.
- Gordon, Peter C., Randall Hendrick, and Marcus Johnson. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51, 97-114.
- Grodner, Daniel and Edward Gibson. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29(2), 261-290.
- Hahn, Hye-ryeong. (2015a). The acceptability of long-distance wh-scrambling in Korean from a real-time acceptability judgment perspective. *Studies in Linguistics* 34, 397-420.
- Hahn, Hye-ryeong. (2015b). Scrambling out of a wh-island: Types and acceptability. *The Journal of Studies in Language* 31(1), 157-177.
- Hahn, Hye-ryeong. (2015c). The resolution of wh-scope ambiguity in Korean and locality bias. *Language Research* 51, 91-120.
- Hahn, Hye-ryeong and Seungjin Hong. (2014). Processing scrambled wh-construction in head-final languages: Dependency resolution and feature checking. *Language and Information* 18(2), 59-79.
- Han, Jong-Im. (1992). Syntactic movement analysis of Korean relativization. *Language Research* 28, 335-357.
- Hofmeister, Philip and Ivan A. Sag. (2010). Cognitive constraints and island effects. *Language* 86, 366-415.
- Hofmeister, Philip, Laura Staum Casasanto, and Ivan A. Sag. (2012a). How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language* 88, 390-400.
- Hofmeister, Philip, Laura Staum Casasanto, and Ivan A. Sag. (2012b). Misapplying working-memory tests: A reductio ad absurdum. *Language* 88, 408-409.
- Hofmeister, Philip, Laura Staum Casasanto, and Ivan A. Sag. (2013). Islands in the grammar? Standards of evidence. In Jon Sprouse and Norbert Hornstein, eds., *Experimental Syntax and Island Effects*, 42-50. Cambridge: Cambridge University Press.
- Hong, Sun-Ho. (2004). On the lack of syntactic effects in Korean wh-questions. *The Linguistic Association of Korea Journal* 12(3), 43-57.
- Just, Marcel A. and Patricia A. Carpenter. (1987). Speed reading. In Marcel A. Just and Patricia A. Carpenter, eds., *The Psychology of Reading and Language Processing*, 425-452. Newton, MA: Allyn and Bacon.
- Kang, Young-Se. (1986). *Korean Syntax and Universal Grammar*. Doctoral dissertation. Harvard University, Cambridge.
- Kim, Bokyoung and Grant Goodall. (2014). An experimental investigation of island effects in Korean. *Japanese/Korean Linguistics* 23. Stanford, CA: CSLI Publications.
- Kim, Euhee and Myung-Kwan Park. (2016). The relation between working-memory capacity

- and island effects: A statistical modeling approach. *The Journal of Linguistic Science* 79, 43-70.
- Kim, Rhang-Hye-Yun. (2003). Scrambling and minimalist program. *Studies in Generative Grammar* 13, 3-19.
- Kimball, John. (1973). Seven principles of surface structure parsing in natural language. *Cognition* 2(1), 15-47.
- King, Jonathan and Marcel A. Just. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30, 580-602.
- Kluender, Robert. (1991). *Cognitive Constraints on Variables in Syntax*. Doctoral dissertation. University of California, San Diego.
- Kluender, Robert. (1992). Deriving island constraints from principles of predication. In Helen Goodluck and Michael Rochemont, eds., *Island Constraints: Theory, Acquisition and Processing*, 241-279. Dordrecht: Kluwer.
- Kluender, Robert. (1998). On the distinction between strong and weak islands: A processing perspective. In Peter Culicover and Louise McNally, eds., *Syntax and Semantics, Vol. 29: The Limits of Syntax*, 241-79. San Diego, CA: Academic Press.
- Ko, Heejeong. (2018). Scrambling in Korean syntax. In *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. Retrieved July 20, 2019 from [http://ling.snu.ac.kr/ko/publications/Papers/Ko\\_Scrambling%20in%20Korean%20Syntax%20FINAL.pdf](http://ling.snu.ac.kr/ko/publications/Papers/Ko_Scrambling%20in%20Korean%20Syntax%20FINAL.pdf)
- Lee, Byeong-taek. (2002). Reliability of the reading span test. *Psychological Science* 11.1, 15-33.
- Lee, Byeong-Taek, Kyoung-Joong Kim, and Myeong-han Zoh. (1996). Working memory and language: Comprehension individual differences in reading span and language processing. *Korean Journal of Experimental and Cognitive Psychology* 8(1), 59-85.
- Lee, Hyo-Sang. (1982). *Asymmetry in Island Constrains in Korean*. MS., University of California at Los Angeles.
- Lee, Yong-hun and Yeonkyung Park. (2015). Complex-NP islands in Korean: An experimental approach. In Hai Zhao, ed., *29th Pacific Asia Conference on Language, Information and Computation*, 240-249. Shanghai, China.
- Lee, Young-Suk. (1993). *Scrambling as Case-Driven Obligatory Movement*. Doctoral dissertation. University of Pennsylvania.
- Miyagawa, Shigeru. (2005). EPP and semantically vacuous scrambling. In Henk van Riemsdijk, Jan Koster, and Harry van der Hulst, eds., *The Free Word Order Phenomenon*, 181-220. Berlin, New York: Mouton de Gruyter.
- Miyamoto, Edson, T. and Shoichi Takahashi. (2003). *Typing Mismatch Effects in the Processing of Wh-Phrases in Japanese*. Manuscript submitted for publication.
- Nishigauchi, Taisuke. (1990). *Quantification in the Theory of Grammar*. Dordrecht: Kluwer Academic Publishers.
- O'Grady, William. (2008). Does emergentism have a chance? In H. Chan, H. Jacob and E. Kapia, eds., *Proceedings of 32nd Annual Boston University Conference on Language Development*, 16-35. Somerville, MA: Cascadilla Press.
- O'Grady, William. (2010). An emergentist approach to syntax. In Bernd Heine and Heiko

- Narrog, eds., *The Oxford Handbook of Linguistic Analysis*, 257–283. Oxford: Oxford University Press.
- Park, Hong-Keun. (2010). On some properties of in-situ and ex-situ wh-phrases in Korean. *Korean Journal of Linguistics* 31(1), 89–118.
- Roberts, Rose and Edward Gibson. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research* 31(6), 573–598.
- Pesetsky, David. (1987). Wh-in-situ: movement and unselective binding. In Eric Reuland and Alice G. B ter Meulen, eds., *The Representation of (In)definites*, 98–129. Cambridge, MA: MIT Press.
- Pesetsky, David. (2000). *Phrasal Movement and Its Kin*. Cambridge, MA: MIT Press.
- Rizzi, Luigi. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.
- Rizzi, Luigi. (2001). Reconstruction, weak island sensitivity, and agreement. In Carlo Cecchetto, Gennaro Chierchia, and Maria Teresa Guasti, eds., *Semantic Interfaces: Reference, Anaphora and Aspect*, 145–76. Stanford, CA: CSLI Publications.
- Saito, Mamoru. (1989). Scrambling as semantically vacuous A'-movement. In Mark Baltin and Anthony Kroch, eds., *Alternative Conceptions of Phrase Structure*, 182–200. Chicago: University of Chicago Press.
- Saito, Mamoru. (2004). Japanese scrambling in a comparative perspective. In David Adger, Cécile de Cat and George Tsulas, eds., *Peripheries: Syntactic Edges and Their Effects*, 143–163. Dordrecht: Kluwer.
- Sohn, Ho-Min. (1980). Theme-prominence in Korean. *Korean Linguistics* 2, 2–19.
- Sprouse, Jon, Matt Wagers, and Colin Phillips. (2012a). A test of the relation between working memory and syntactic island effects. *Language* 88, 82–123.
- Sprouse, Jon, Matt Wagers, and Colin Phillips. (2012b). Working-memory capacity and island effects: A reminder of the issues and the facts. *Language* 88, 401–407.
- Staum Casasanto, Laura, Philip Hofmeister, and Ivan A. Sag. (2010). Understanding acceptability judgments: Distinguishing the effects of grammar and processing on acceptability judgments. In Stellan Ohlsson and Richard Catrambone, eds., *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 224–229. Austin, TX: Cognitive Science Society.
- Suh, Chungmok. (1987). *WH-Constructions in Korean*. Seoul: Top Press.
- Ueyama, Ayumi. (2008). Two types of scrambling constructions in Japanese. In A. Barss and T. Langendoen, eds., *Anaphora: A Reference Guide*, 23–71. Cambridge: Blackwell.
- Yoshimura, Noriko. (1992). *Scrambling and Anaphora in Japanese*. Doctoral dissertation, University of Southern California, Los Angeles.
- Yoon, Jeong-Me. (2013). Undoing and wh-island effects of scrambling in Korean. *Studies in Generative Grammar* 23.1, 41–63.
- Watanabe, Akira. (1992). Subjacency and s-Structure movement of wh-in-Situ. *Journal of East Asian Linguistics* 1, 255–291.

Hye-ryeong Hahn  
Professor  
Department of English Education  
Seowon University  
Musimseoro 377-3, Seowon-gu, Cheongju 28674, Chungcheongbukdo, Korea  
E-mail: [hyerhahn@seowon.ac.kr](mailto:hyerhahn@seowon.ac.kr)

Received: July 1, 2019

Revised version received: July 31, 2019

Accepted: August 14, 2019