



Voice-mismatch asymmetry in Korean *why*-stripping*

Jeong-Seok Kim
(Korea University)

Kim, Jeong-Seok. 2023. Voice-mismatch asymmetry in Korean *why*-stripping. *Linguistic Research* 40(3): 409-429. The present study uses an acceptability judgment experiment designed to explore the nature of Korean *why*-stripping with reference to voice. First, it investigates whether the acceptability of Korean *why*-stripping is affected by the voice mismatch between antecedent and ellipsis clauses. Second, it examines whether there is a difference in acceptability between passive antecedents with active ellipsis clauses (Passive-Active) and active antecedents with passive ellipsis clauses (Active-Passive). One finding indicates that Korean *why*-stripping prefers voice matches over voice mismatches: voice matches make it easier to link the remnant to its correlate in the antecedent clause. The other finding indicates that Passive-Active *why*-stripping is more acceptable than Active-Passive *why*-stripping. The study also evaluates whether the second finding supports Arregui et al.'s (2006) Recycling Hypothesis—according to which passive clauses are more likely to be remembered as active than the other way around, creating an illusion of grammaticality—or Poppels and Kehler's (2019) hypothesis—according to which there exists a penalty against passive ellipsis clauses, which applies to both matched and mismatched cases of English VP ellipsis. The results confirm the hypothesis of Poppels and Kehler in that Active-Active *why*-stripping is more acceptable than Passive-Passive *why*-stripping and that Passive-Active *why*-stripping is more acceptable than Active-Passive *why*-stripping. (Korea University)

Keywords passive penalty, processing, recycling hypothesis, voice mismatches, *why*-stripping

1. Introduction

Why-stripping is a type of clausal ellipsis found in English (Weir 2014; Yoshida et al. 2015). It involves a *wh*-phrase *why* followed by a non-*wh*-remnant, as shown below:

* I would like to express my gratitude to the two anonymous reviewers and the participants at the 2022 Seminar at Korea University for their invaluable feedback and insights on this material. Any remaining errors in this study are my own responsibility. This study was supported by a faculty research grant from the College of Liberal Arts at Korea University in 2023.

- (1) Mary ate an apple, but why an apple? (= but why did she eat an apple?)

In accordance with the ellipsis tradition established by Merchant (2001) and others, the overt non-wh-phrase in the elided clause in (1) is called a *remnant*, and the corresponding phrase in the antecedent clause is called a *correlate*.

Korean also has a construction similar to English *why*-stripping (Cho and Lee 2017, 2018; Kim 2017; Bae and Park 2021; Kim et al. 2021), as illustrated in (2).

- (2) Mary-ka sakwa-lul mekess-nuntey, way sakwa-lul?
 Mary-Nom apple-Acc ate-but why apple-Acc
 ‘Mary ate an apple, but why an apple?’

It consists of a wh-phrase *way* ‘why’ followed by a non-wh-remnant. The remnant in *why*-stripping such as *an apple* in (1) or *sakwa-lul* in (2) corresponds to its correlate in the antecedent clause such as *an apple* or *sakwa-lul*, respectively. In *why*-stripping, the non-wh-remnant is usually the same as its correlate in the antecedent clause.

This study is interested in the acceptability of (3a) and (3b), which involves Korean *why*-stripping with voice mismatches between antecedent and ellipsis clauses.

- (3) a. [P(assive)-A(ctive)]
 Sinnmwun-i John-eyuyhay paytaltoyess-nuntey, way John-i?
 newspaper-Nom John-by be.delivered-but why John-Nom
 ‘The newspaper was delivered by John, but why John?’
- b. [A(ctive)-P(assive)]
 John-i sinnmwun-ul paytalhayss-nuntey, way John-eyuyhay?
 John-Nom newspaper-Acc delivered-but why John-by
 ‘John delivered the newspaper, but why by John?’

According to Kim et al.’s (2021) observation, voice mismatches between active and passive clauses in *why*-stripping are not acceptable. For example, a passive clause cannot precede an active remnant in the ellipsis site as in (3a), and vice versa as in (3b). However, Kim et al. cautiously note that some speakers find the *why*-stripping in (3a) to be less unacceptable than the *why*-stripping in (3b).

The main focus of this study is to investigate whether there are differences in

acceptability between distinct types of voice (mis)matches in ellipsis. Specifically, it has been noted that when a passive clause comes before an active VP ellipsis as in (4a), it is generally more acceptable than when an active clause comes before a passive VP ellipsis as in (4b).

- (4) a. The report was first read by John, and then Mary did too. [P-A]
 b. John read the report first, and then the confession was too. [A-P]

This observation was first made by Hardt (1993) and has been supported by subsequent studies conducted by Arregui et al. (2006), Kim et al. (2011), Kim and Runner (2018), Poppels and Kehler (2019), and others.

However, it is unclear if this pattern extends to other languages or ellipsis types. Given that Korean lacks a clear instance of English-style VP ellipsis, our study investigates if the acceptability of Korean *why*-stripping (a form of clausal ellipsis) is affected by differences in voice between antecedent and ellipsis clauses.

The remainder of this paper is organized as follows. Section 2 examines earlier research on the effects of voice mismatches and the penalty for passive clauses in VP ellipsis. We also explore some theoretical issues related to *why*-stripping in English and Korean. In Section 3, we present the results of a formal experiment on Korean *why*-stripping, with a focus on voice mismatches, and discuss the outcomes. Section 4 provides a general discussion of Korean *why*-stripping with voice mismatches. Finally, we conclude in Section 5.

2. Background

2.1 VP ellipsis and voice mismatches

English has VP ellipsis, as illustrated in (5).

- (5) John investigated the problem, and Bill did too.

In many cases of VP ellipsis, identifying the meaning relies on the relevant antecedent being present. According to the syntactic approach to ellipsis, when there is a missing

VP in a sentence, understanding the intended meaning depends on whether there is a similar VP available in the context to serve as a reference. This approach suggests that VP ellipsis is only acceptable if there is a syntactically-matching VP nearby (Sag 1976; Merchant 2008, 2013, etc.). Meanwhile, the referential approach treats VP ellipsis as a kind of placeholder that needs to be interpreted in the same way as other referential expressions like pronouns. Under this approach, understanding the meaning of VP ellipsis depends on the speaker's or listener's mental model of the discourse, and how it represents the relevant referents (Schachter 1977; Lobeck 1999, etc.).

To compare the predictions of different approaches, we can examine cases where the meaning of a VP is clear from the context but the structure does not match the expected form. If we adopt a syntactic approach, we expect such cases to be unacceptable, but a referential approach would predict that they are acceptable. Unfortunately, the data in such cases are often unclear. We can notice this by looking at different versions of (5):

- (6) a. #The problem was investigated by John, and Bob did too.
 b. The problem was to have been investigated, but obviously nobody did.
 (modified from Poppels and Kehler 2019: 2)

The example in (6a) demonstrates a voice mismatch where the second clause is active, but the first clause is passive, so the expected VP *investigated the problem* is missing from the context, making the sentence less acceptable. Referential analyses must explain why this sentence is unacceptable despite the meaning of the first clause providing the necessary information. In contrast, (6b) is considered acceptable, despite having the same voice mismatch as (6a), which is problematic for the syntactic approach since the required VP is still missing.

Arregui et al. (2006) and others (Kim et al. 2011; SanPietro et al. 2012; Kertz 2013; Kim and Runner 2018; Poppels and Kehler 2019; Clifton et al. 2019, etc.) have conducted many experiments to understand the acceptability of sentences with mismatches, using rating tasks to obtain more precise measurements. Consequently, two important findings have been revealed.

One finding is that sentences with a match, such as (7a), are consistently rated as more acceptable than paired versions where there is a voice mismatch, as in (7b).

- (7) a. John read the report first, and then Mary did too. [A-A]
 b. The report was first read by John, and then Mary did too. [P-A]

The other important finding is the variation in acceptability between distinct types of voice mismatches. As previously mentioned, passive antecedent-active VP ellipsis is generally more acceptable than active antecedent-passive VP ellipsis, which is shown again below in (8).¹

- (8) a. The report was first read by John, and then Mary did too. [P-A]
 b. John read the report first, and then the confession was too. [A-P]

Following Poppels and Kehler (2019), we will term this finding the *voice-mismatch asymmetry*. This raises a key question: Is this asymmetry due to some VP ellipsis constraint or a result of independent processing? Identifying the source is crucial to understanding ellipsis constraints and comprehension mechanisms.

Among others, Arregui et al. (2006) propose a processing-based explanation, particularly related to memory behavior, which attributes the voice-mismatch asymmetry to the Recycling Hypothesis (RH). The RH consists of two parts: (i) a grammatical restriction on the use of VP ellipsis, which requires syntactic similarity between a missing material and its antecedent; and (ii) a processing theory that accounts for any remaining variability in acceptability when grammar predicts ungrammaticality.

According to Arregui et al. (2006), syntactic identity dictates voice-mismatched VP ellipsis to be ungrammatical. As a result, both [P-A] and [A-P] mismatches are expected to be ungrammatical based on grammar alone. However, when the processor detects a grammatical violation, it attempts to reanalyze previous syntactic material and recycle it in a way that makes the input grammatical. In the case of ellipsis with non-identical antecedents, the processor may reanalyze the existing antecedent and create an alternative antecedent that satisfies the identity requirement. The degree of effort required by the processor to repair an ellipsis is assumed to determine the relative acceptability of the sentence, so that ungrammatical cases of ellipsis with non-identical antecedents may be perceived as relatively acceptable if an identical antecedent can be recycled from the existing one with minimal effort. The asymmetry between [P-A] and [A-P] mismatches

¹ This finding challenges Merchant's (2008, 2013) theory, which deemed VP ellipsis with voice mismatches as grammatical through intricate syntactic representation.

is explained as a by-product of the recycling process, with the help of an independently motivated auxiliary assumption based on syntactic misremembering by both speakers and listeners.

Grammar allows ellipsis only when the elided content matches with the antecedent in terms of syntax. To confirm this match, the second clause must retrieve the first clause from memory. In cases where there is no syntactically-matching antecedent available, the processor generates one using the available materials. For instance, if the second ellipsis clause is in the active voice and the first antecedent clause is in the passive voice, the antecedent clause might be recalled as active, resulting in a false [A-A] match, as seen in (8a). In simpler terms, speakers and listeners are more likely to mistakenly recall a passive clause as active (given its simplicity compared to the complex passive form it corresponds to) than to misremember an active clause as passive (which would be the alternative form of the simpler active structure) (Mehler 1963). According to this hypothesis, mismatches in (8b) are expected to be less acceptable. This means that the voice-mismatch asymmetry is explained as a processing phenomenon, rather than a property of ellipsis that requires further explanation.

Regarding the voice-mismatch asymmetry, Poppels and Kehler (2019) examine Arregui et al.'s (2006) RH via formal experiments, including the data in (8) and (9).

- (9) a. John read the report first, and then Mary did too. [A-A]
 b. The report was first read by John, and then the confession was too. [P-P]

The outcomes of (9) have an interesting implication for voice mismatches in ellipsis. If [P-P] in (9b) is as acceptable as [A-A] in (9a), it supports the RH. Conversely, if [P-P] in (9b) is less (or more) acceptable than [A-A] in (9a), it challenges the validity of the RH thesis. The RH predicts no difference in acceptability between matching conditions. In fact, Poppels and Kehler's experimental results showed that [P-P] is significantly less acceptable than [A-A]. Therefore, they conclude that there is a penalty for passive ellipsis clauses in terms of ellipsis interpretation. In other words, [P-P] is less acceptable than [A-A] because it involves a passive ellipsis clause, just as [A-P] is less acceptable than [P-A] due to the presence of a passive ellipsis clause.

2.2 *Why*-stripping and voice mismatches

As previously mentioned, it remains uncertain whether the voice-mismatch asymmetry observed in English VP ellipsis extends to other languages or distinct types of ellipsis, such as clausal ellipsis. While Korean does not exhibit a straightforward equivalent of English VP ellipsis, it remains unclear whether voice-mismatch asymmetry applies in this context. The counterpart of English VP ellipsis, as seen in (5), is ungrammatical in Korean (cf. Kim 1997):

- (10) *John-i ku mwuncey-lul cosahayss-ko Bill-i (ttohan) haysse.
 John-Nom the problem-Acc investigated-and Bill-Nom (also) did
 ‘John investigated the problem, and Bill did too.’

In light of this, we aim to investigate whether the acceptability of Korean *why*-stripping, a form of clausal ellipsis, is influenced by voice mismatches between antecedent and ellipsis clauses. To contextualize our discussion, we will briefly review informal observations in the literature regarding voice-mismatches in clause-type ellipsis and subsequently delve into the specific case of voice-mismatch in Korean *why*-stripping.

Generally, achieving parallelism between antecedent and ellipsis clauses appears to be a requirement for matching voice in ellipsis. Merchant (2008, 2013) notes that voice mismatches are not well-tolerated in structures like sluicing, gapping, and stripping, among others:

- (11) a. *Someone murdered Joe, but we don’t know by whom. (Merchant 2013: 93)
 b. *Some bring roses and lilies by others. (Merchant 2013: 83)
 c. *Max brought the roses, but not by Amy! (Merchant 2013: 83)

Shifting our focus to the concept of *why*-stripping, Yoshida et al. (2015) contend that the sentence in (12a) is formed following the procedure depicted in (12b).

- (12) a. Mary ate an apple, but why an apple?
 b. [_{CP} why C [_{FocP} an apple₁ Foc [_{TP} ~~Mary ate~~ t₁]]]

They assume that the focus-sensitive *why* in *why*-stripping is generated in Spec of CP, and claim that the non-wh-remnant *an apple* moves to Spec of Foc(us)P, which is located between CP and TP. The remainder of the clause (TP) is then deleted, resulting in the *why*-stripping construction.

Further, Yoshida et al. (2015) observe that *why*-stripping is not compatible with voice mismatches, as illustrated below:

- (13) a. Max brought the roses, but I don't understand why Max (but not John).
 b. The roses were brought by Max, but I don't understand why by Max (but not by John).
 c. *Max brought the roses, but I don't understand why by Max (but not John).
 (modified from Yoshida et al. 2015: 337)

The requirement for voice matching between the antecedent clause and the stripped clause in *why*-stripping suggests a common licensing condition on ellipsis, shared with sluicing and stripping. This condition demands syntactic parallelism between antecedent and ellipsis clauses, as proposed by Merchant (2008, 2013).

Now, let us turn to Korean *why*-stripping in (14).

- (14) Mary-ka sakwa-lul mekess-nuntey, way sakwa-lul?
 Mary-Nom apple-Acc ate-but why apple-Acc
 'Mary ate an apple, but why an apple?'

If *why*-stripping is considered an example of ellipsis, it should exhibit connectivity effects, such as case, binding, and voice (Ross 1969; Merchant 2001, etc.). As far as case and binding are concerned, this expectation seems to be met. Ross argues that in clausal ellipsis, particularly sluicing, the remnant and its antecedent must match in case. This is known as the case-matching generalization, which can be detected in languages like German where nominal case is overtly marked. In Korean, it is observed that *why*-stripping usually does not allow case mismatches, as shown below:

- (15) John-i samchon-eykey hyepcohayss-nuntey,
 John-Nom uncle-Dat assisted_{Dat}-but

- way samchon-eykey/*samchon-ul?
 why uncle-Dat /*uncle-Acc
 ‘John assisted_{Dat} his uncle, but why his uncle_{Dat}?’
- (16) John-i samchon-ul towass-nuntey,
 John-Nom uncle-Acc helped_{Acc}-but
 way samchon-ul/*samchon-eykey?
 why uncle-Acc/*uncle-Dat
 ‘John helped_{Acc} his uncle, but why his uncle_{Acc}?’

While the verbs *hyepcoha(ta)* ‘to assist’ and *top(ta)* ‘to help’ in (15) and (16) have similar meanings, they differ in their case licensing: *hyepcoha(ta)* only licenses dative case, while *top(ta)* only licenses accusative case. This case-matching effect is commonly observed in languages that mark case overtly.

Next, Kim (2017) observes binding connectivity in Korean *why*-stripping as follows:

- (17) A: Mimi₁-uy oppa₂-ka caki₂ kulim-ul phal-ass-e.
 Mimi-Gen brother-Nom self pictures-Acc sell-Past-Dec
 ‘Mimi’s brother sold pictures of himself.’
- B: Way caki₂/_{*1} kulim-ul?
 why self pictures-Acc?
 ‘Why pictures of himself?’ (Kim 2017: 741)

The pronoun *caki* in (17) is only referring to the subject *oppa* and not to the specifier *Mimi* of the subject in the antecedent clause.

Finally, Kim et al. (2021) have noted that in Korean *why*-stripping, there is no tolerance for active-passive mismatches: a passive correlate cannot come before an active remnant, and vice versa, as illustrated in (18a) and (18b), respectively.

- (18) a. *Sangphwum-i Mary-eyuyhay wunpantoyess-nuntey, way Mary-ka?
 goods-Nom Mary-by be.delivered-but why Mary-Nom?
 ‘The goods were delivered by Mary, but why Mary?’ [P-A]
- b. *John-i sinmwun-ul paytalhayss-nuntey, way John-eyuyhay?
 John-Nom newspaper-Acc delivered-but why John-by
 ‘John delivered the newspaper, but why by John?’ [A-P]

Kim et al. interestingly recognize that a portion of the speakers (namely, three out of 10 informants) perceive the voice-mismatched *why*-stripping in (18a) as more acceptable when contrasted with (18b).

Poppels and Kehler (2019) conducted a pilot experiment to examine the Passive-Ellipsis-Penalty Hypothesis (PEPH) in gapping and sluicing. A sample set of gapping is given below:

- (19) a. Mary scolded Wilma, and Susan, Nancy. [A-A] gapping
 b. Wilma was scolded by Mary, and Nancy, by Susan. [P-P] gapping
 c. Mary scolded Wilma, and Susan scolded Nancy. [A-A] control
 d. Wilma was scolded by Mary, and Nancy was scolded by Susan.
 [P-P] control
 (Poppels and Kehler 2019: 17)

It is important to note that their pilot study did not include cases with syntactic mismatches due to challenges in constructing relevant examples. The results of the pilot study indicated that the PEPH does not apply to gapping and sluicing. To date, there has not been a formal experimental investigation of the voice-mismatch asymmetry in clausal ellipsis. Given the limitations of their pilot study, we aim to explore whether the passive penalty is specific to ellipsis by examining Korean *why*-stripping.

3. Experiment

We aim to evaluate the following hypotheses: First, the acceptability of Korean *why*-stripping is influenced by the voice match between antecedent and ellipsis clauses. Second, the voice of ellipsis clauses affects acceptability. We predict that voice-matched *why*-stripping will be more acceptable than voice-mismatched *why*-stripping due to ellipsis parallelism. Additionally, we predict that there will be an acceptability difference between P-A and A-P *why*-stripping, consistent with Arregui et al.'s (2006) findings for English VP ellipsis. We expect to observe a similar effect in Korean *why*-stripping.

3.1 Participants, materials, and design

We recruited 51 native Korean speakers who were undergraduate students at a university in South Korea. Participants were between the ages of 18 and 27, with an average age of 21.81 years. Participants agreed to participate in the experiment and provided informed consent in exchange for course credit. The experiment was conducted online, and participants typically completed it in 10 minutes. However, the responses from three participants who were not paying attention during the experiment were excluded from the analysis, leaving only the responses from 48 participants (12 for each of the four lists) in the analysis.

The experiment used a 2×2 design which involved two variables: MATCH (Match (of voices) vs. Mismatch (of voices)) and VOICE.E (Active.E (active voice of ellipsis clauses) vs. Passive.E (passive voice of ellipsis clauses)), as shown in (20).

- (20) a. [A-A] (= [Match | Active.E])
 John-i sinmwun-ul paytalhayss-nuntey, way John-i?
 John-Nom newspaper-Acc delivered-but why John-Nom
 ‘John delivered the newspaper, but why John?’
- b. [P-P] (= [Match | Passive.E])
 Sinmwun-i John-eyuyhay paytaltoyess-nuntey, way John-eyuyhay?
 newspaper-Nom John-by be.delivered-but why John-by
 ‘The newspaper was delivered by John, but why John?’
- c. [P-A] (= [Mismatch | Active.E])
 Sinmwun-i John-eyuyhay paytaltoyess-nuntey, way John-i?
 newspaper-Nom John-by be.delivered-but why John-Nom
 ‘The newspaper was delivered by John, but why John?’
- d. [A-P] (= [Mismatch | Passive.E])
 John-i sinmwun-ul paytalhayss-nuntey, way John-eyuyhay?
 John-Nom newspaper-Acc delivered-but why John-by
 ‘John delivered the newspaper, but why by John?’

The experiment had four conditions, two of which had voice-matched clauses. In the [A-A] condition, both antecedent and ellipsis clauses had an active voice, while in the

[P-P] condition, both had a passive voice. The other two conditions had voice-mismatched clauses: the antecedent clause was passive and the ellipsis clause was active in the [P-A] condition, and vice versa in the [A-P] condition. The complete list of items used in the experiment is available online.²

The experiment used 16 sets of sentences that were matched in terms of their vocabulary and grammar, but varied in their voice-matching conditions. These sets of sentences were assigned to four lists using a Latin square design, ensuring that each list had one sentence from each set. Each list contained 16 experimental items, along with 48 filler items that were of similar length to the experimental items but varied in their acceptability. Therefore, each list had a total of 64 sentences.

3.2 Procedure and data analysis

We used an online experiment platform called PClbex (Zehr and Schwarz 2018) to conduct our study. Participants were shown sentences one at a time on a computer screen and asked to rate their acceptability using a 1-7 scale. Along with the main test items, we also included 16 filler items that were previously identified as either highly acceptable or highly unacceptable by a group of 200 participants. We assigned an expected value of 7 for the highly acceptable items and 1 for the highly unacceptable ones. For each filler item, we calculated the difference between each participant's rating and its expected value, squared the difference, and then summed these squared differences. This allowed us to identify participants whose ratings were significantly different from what was expected, and we excluded these participants from our analysis (cf. Sprouse et al. 2022).

Before analyzing the data, we converted the raw ratings for both experimental and filler items into z-scores. This helped to remove any potential biases in how participants used the rating scale, as it standardized all the ratings to the same scale (Schütze and Sprouse 2013). We used linear mixed-effects models to analyze the data, which allowed us to include both random participant and random item variables in our analysis (Baayen et al. 2008). We used the *lme4* package (Bates et al. 2015) in the R software environment (R Core Team 2020) to estimate these models. To estimate *p*-values for the fixed and random effects, we used Satterthwaite's approximation (Kuznetsova et al. 2017).

2 https://www.researchgate.net/publication/374948319_Appendix

3.3 Results

We analyzed the responses of 48 participants, with 12 participants assigned to each of the four lists. This gave us a total of 768 tokens for analysis, with 192 tokens for each of the four experimental conditions. In Figure 1, we present the mean z-scores for the acceptability judgments in each of the four experimental conditions. A score of zero indicates the overall mean acceptability rating, while positive z-scores indicate that the conditions were rated as more acceptable and negative z-scores indicate that the conditions were rated as less acceptable.

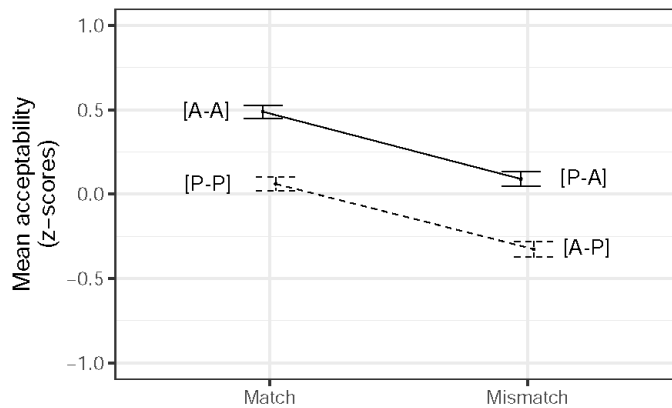


Figure 1. Mean acceptability of experimental conditions (Error bars indicate SE)

Table 1 provides a summary of the linear mixed-effects model we used. The model included MATCH and VOICE.E as fixed effects, as well as their interaction. It also included random intercepts for participants and items, and the maximum number of random slopes justified by the data (Barr et al. 2013).³

³ The complete formula was: $z.score \sim MATCH*VOICE.E + (1 + MATCH*VOICE.E | participant) + (1 | item)$.

Table 1. Fixed effect summary for the experiment

	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
(Intercept)	0.489	0.065	7.506	***
MATCH	-0.399	0.067	-5.924	***
VOICE.E	-0.428	0.072	-5.935	***
MATCH:VOICE.E	0.012	0.085	0.140	0.89

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The statistical analysis revealed a significant effect of MATCH. The voice-matched condition was rated as significantly more acceptable than the voice-mismatched condition. We also found a significant effect of VOICE.E, with the condition containing an active ellipsis clause rated as significantly more acceptable than that containing a passive ellipsis clause. Notably, these two effects were independent of each other.

To summarize, our study yielded two main findings. First, we found that syntactically matched cases of TP ellipsis were rated as more acceptable than syntactically mismatched cases. Specifically, the [A-A] condition was more acceptable than the [P-A] condition (*mean*: 0.489 vs. 0.090), and the [P-P] condition was more acceptable than the [A-P] condition (*mean*: 0.061 vs. -0.326). Second, we found that the mismatched [P-A] condition was rated as more acceptable than the mismatched [A-P] condition (*mean*: 0.090 vs. -0.326).⁴

4. Discussion

The purpose of the experiment was to evaluate two predictions. The first prediction was that syntactically matched cases of TP ellipsis would be rated as more acceptable than syntactically mismatched cases. This prediction was supported by the results. The second prediction was also confirmed, with [P-A] mismatches being rated as more acceptable

4 Two reviewers independently raised a question about the greater acceptability of the [P-A] mismatch condition (*mean*: 0.090) in comparison to the [P-P] match condition (*mean*: 0.061). It is essential to emphasize that the current experiment was not designed to compare [P-A] and [P-P]. To be precise, our statistical comparisons are limited to the following pairs: [A-A]-[P-P], [A-A]-[P-A], [P-P]-[A-P], and [P-A]-[A-P]. Therefore, we cannot draw any conclusion about the statistical differences between [A-A] and [A-P] or between [P-A] and [P-P]. Despite these experimental limitations, we conducted a one-way mixed-effects model to compare the acceptability ratings of [P-A] and [P-P]. The results indicated no significant difference between them ($\beta = 0.029$, $SE = 0.077$, $t = 0.383$, $p = 1.000$). This suggests that the effect sizes of the [P-A] penalty and the [P-P] penalty are fairly similar. However, it is worth noting that confirming this claim would require a different experimental design.

than [A-P] mismatches. This finding replicated the voice-mismatch asymmetry effect previously reported by Arregui et al. (2006).⁵

However, the [A-A] condition, where antecedent and ellipsis clauses had the same voice, was found to be more acceptable than the [P-P] condition where both clauses were passive (*mean*: 0.489 vs. 0.061). This finding suggests that the effect of voice (mis)matching is not solely due to the memory-based processes posited by the RH, but may also be influenced by a more general preference or penalty for certain types of syntactic structures. The data do not support the idea that there is a penalty against passive clauses in general, as this would predict no difference between [P-A] and [A-P] mismatches: both have one passive clause. The results thus suggest that there may be a separate penalty for passive ellipsis clauses, which affects their acceptability even when passive ellipsis clauses are part of a matched syntactic structure. Among the four conditions in (20), only the [P-P] in (20b) and the [A-P] in (20d) contained passive ellipsis clauses. This resulted in reduced acceptability for both [P-P] matches compared to [A-A] matches and [A-P] mismatches compared to [P-A] mismatches. The data align with Poppels and Kehler's (2019) Passive-Ellipsis-Penalty Hypothesis (PEPH), which predicts a distinction between [A-A] and [P-P] matches due to the presence of a passive ellipsis clause exclusively in the latter.

In the upcoming discussion, we will explain how Korean *why*-stripping is derived by building upon Kim et al.'s (2021) analysis. Our proposal consists of three parts:

- (21) a. Korean *why*-stripping has a structure similar to that of a *why*-question.
- b. *Way* 'why' in Korean *why*-stripping originates in Spec of CP.
- c. The non-wh-phrase in Korean *why*-stripping is a focus phrase associated with *why*.

To be more specific, we suggest that the experimental stimuli in (20) are derived in the following manner:

5 The findings of previous studies on matched VP ellipsis have been mixed. Kim et al. (2011) and Poppels and Kehler (2019) found that [A-A] matched cases were more acceptable than [P-P] matched cases. However, Kim and Runner (2018) reported an interaction between antecedent voice and mismatch, suggesting no difference in acceptability between voice-matched [A-A] and [P-P] stimuli.

- (22) a. John-Nom newspaper-Acc delivered-but, [A-A]
 [_{CP} why [_{FocP} John₁-Nom [_{TP} ~~t_i-newspaper-Acc delivered~~]]-Q]
- b. newspaper-Nom John-by be.delivered-but, [P-P]
 [_{CP} why [_{FocP} John₁-by [_{TP} newspaper-Nom t_i-be.delivered]]-Q]
- c. newspaper-Nom John-by be.delivered-but, [P-A]
 [_{CP} why [_{FocP} John₁-Nom [_{TP} ~~t_i-newspaper-Acc delivered~~]]-Q]
- d. John-Nom newspaper-Acc delivered-but, [A-P]
 [_{CP} why [_{FocP} John₁-by [_{TP} newspaper-Nom t_i-be.delivered]]-Q]

To explain why [A-A] in (22a) is highly acceptable, we suggest that *way* ‘why’ is initially positioned in Spec of CP, while the non-wh-remnant *John-i* ‘John-Nom’ moves to Spec of FocP before TP deletion.⁶ After TP deletion, the Q-particle is left behind but can be further deleted through extra deletion, following An’s (2016, 2019) proposal.⁷ On the other hand, for the moderately acceptable [P-P] in (22b), we suggest that *way* ‘why’ is base-generated in Spec of CP, and the non-wh-remnant *John-eyuyhay* ‘John-by’ moves to Spec of FocP, prior to TP deletion. Both [A-A] and [P-P] are syntactically well-formed.

The lower acceptability of [P-A] in (22c) and [A-P] in (22d) compared to [A-A] in (22a) and [P-P] in (22b), respectively, can be explained by the negative impact of voice mismatches on acceptability ratings for syntactic reasons. Apparently, the acceptability

6 A reviewer raised a question about how our current analysis could explain the reverse word order in *why*-stripped clauses, as illustrated in (i):

- (i) John-i sinmwun-ul paytalhayss-nuntey, John-i way?
 John-Nom newspaper-Acc delivered-but John-Nom why
 cf. ‘John delivered the newspaper, but why John (but not Bill)?’

It appears that the non-wh-remnant in (i) might function as a topic rather than a focus, unlike the non-wh-remnant in our baseline example in (20a). If the example in (i) were a variation of Korean *why*-stripping, it could result from subject scrambling over *way* ‘why’ in Spec of CP (or Spec of ForceP), or it might involve topic movement to Spec of TopP, which could be positioned higher than Spec of ForceP in Korean. In our present study, we do not attempt to address the question of potential similarities and differences between (20a) and (i). We leave this matter for future theoretical investigation.

7 Regarding non-constituent ellipsis, we follow An’s (2016, 2019) extra deletion framework. An observes that P/case markers on ellipsis remnants can only be absent in the final position of a string. He argues that PF deletion extends to parts of ellipsis remnants, like a P/case marker or a head noun, as long as they remain recoverable and adjacent. This extra deletion process is entirely reliant on the PF deletion operation. While syntax determines what should be deleted, PF deletion also maintains a condition that what is elided must be contiguous. According to An’s proposal, PF deletion can disregard syntactic constituency.

difference between voice-mismatch conditions (i.e., [P-A] in (22c) and [A-P] in (22d)) seems to support the RH account: passive clauses are more likely to be misremembered as active, and mismatches that involve a [P-A] clause are more likely to create an illusion of grammaticality than cases that involve an [A-P] clause.

Nevertheless, the finding that [P-P] in (22b) is less acceptable than [A-A] in (22a) raises concerns about the validity of the RH thesis. This is because both types of ellipses have a syntactically-matching TP available in the antecedent clause, so there should be no constraint violation and no need for a recovery mechanism like the RH. The RH predicts no difference in acceptability between matched conditions. The results therefore support Poppels and Kehler's (2019) Passive-Ellipsis-Penalty Hypothesis (PEPH).

It seems that the penalty should be addressed from the perspective of processing, given that both [A-A] and [P-P] were syntactically well-formed. We suggest that when the processor deals with the ellipsis clauses in (22a) and (22b), the non-*wh*-remnant, which is an agent argument *John-i* 'John-Nom' or *John-eyuyhay* 'John-by', refers to its correlate *John-i* or *John-eyuyhay* in the antecedent clause for a successful interpretation. As shown in the antecedent clause in (22b), a theme argument typically precedes an agent argument in passive clauses. While the word order in [A-A] parallels the order of thematic hierarchy (i.e., "The agent precedes the theme", Van Valin and Lapolla 1997), the word order in [P-P] does not. To be specific, we suggest that the theme-agent order in [P-P] places an extra processing challenge relative to the agent-theme order in [A-A], which would explain the difference in acceptability.⁸

5. Conclusion

In this study, we conducted an experiment to probe into the voice-mismatch asymmetry

8 As highlighted by a reviewer, Poppels and Kehler (2019) propose that the PEPH may arise from a conflict between the information structure of passive voice and VP ellipsis. In the passive voice, the subject is typically marked as topical, while VP ellipsis presupposes a topical meaning. In constructions like (i), this conflict emerges:

(i) The report was first read by John, and then the confession was too.

VP ellipsis requires the VP meaning (*read by John*) to be topical, but the passive construction implies that the subject NP (*the confession*) is topical. This creates a clash because focus must be present in the clause. The debate between the thematic hierarchy and the informational-structural approach to explain the passive-ellipsis-penalty is a topic for future research.

in clausal ellipsis (specifically, *why*-stripping) in Korean, and whether it supports the Recycling Hypothesis (RH) (Arregui et al. 2006) or the Passive-Ellipsis-Penalty Hypothesis (PEPH) (Poppels and Kehler 2019). We found that the voice-mismatch asymmetry in English VP ellipsis was replicated in Korean *why*-stripping: passive antecedents with active ellipsis clauses were more acceptable than active antecedents with passive ellipsis clauses. However, we also discovered that there was a penalty for syntactically matched cases in passive voice: the passive-passive *why*-stripping is less acceptable than the active-active *why*-stripping. This suggests that the source of the penalty is broader than just the domain where the RH applies. We thus concluded that the experimental results support the PEPH rather than the RH.

To explore whether the passive penalty is an independent condition unrelated to ellipsis, we can examine sentences without ellipsis. For instance: (i) *John-i sinmwun-ul paytalhayss-nuntey, way John-i sinmwun-ul paytalhayss-ni?* (= John delivered the newspaper, but why did John deliver it?) and (ii) *Sinmwun-i John-eyuyhay paytaltoyess-nuntey, way sinmwun-i John-eyuyhay paytaltoyess-ni?* (= The newspaper was delivered by John, but why was it delivered by John?). If we observe no similar effect in these non-ellipsis sentences, it could support the argument that the passive penalty only applies to sentences with ellipsis clauses. However, if we happen to detect a similar effect in full sentences, it could challenge the PEPH and suggest that the passive penalty is not specific to ellipsis.

References

- An, Duk-Ho. 2016. Extra deletion in fragment answers and its implications. *Journal of East Asian Linguistics* 25(4): 313-350.
- An, Duk-Ho. 2019. On certain adjacency effects in ellipsis contexts. *Linguistic Inquiry* 50(2): 337-355.
- Arregui, Ana, Charles Clifton, Lyn Frazier, and Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55(2): 232-246.
- Baayen, Rolf H., Douglas J. Davidson, and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4): 390-412.
- Bae, Suyoung and Bum-Sik Park. 2021. *Why*-stripping in Korean: A division of labor between syntax and processing. *Korean Journal of Linguistics* 46(1): 29-50.

- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255-278.
- Bates, Douglas, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.
- Bromberger, Sylvain. 1992. *On what we know we don't know: Explanation, theory, linguistics, and how questions shape them*. Chicago, IL: The University of Chicago Press.
- Cho, Sae-Youn and Han-Gyu Lee. 2017. Properties of *why* construction in Korean: A direct interpretation approach [in Korean]. *Studies in Humanities* 55: 97-120.
- Cho, Sae-Youn and Han-Gyu Lee. 2018. The *why-not* construction in Korean: A direct interpretation approach. *The Linguistic Association of Korea Journal* 26(2): 1-19.
- Clifton, Charles Jr., Ming Xiang, and Lyn Frazier. 2019. A note on the voice mismatch asymmetry in ellipsis. *Journal of Psycholinguistic Research* 48(4): 877-887.
- Hardt, Daniel. 1993. *Verb phrase ellipsis: Form, meaning and processing*. PhD Dissertation. University of Pennsylvania.
- Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89(3): 390-428.
- Kim, Christina S., Gregory M. Kobele, Jeffrey T. Runner, and John T. Hale. 2011. The acceptability cline in VP ellipsis. *Syntax* 14(4): 318-354.
- Kim, Christina S. and Jeffrey T. Runner. 2018. The division of labor in explanations of verb phrase ellipsis. *Linguistics and Philosophy* 41(1): 41-85.
- Kim, Jeong-Seok. 1997. *Syntactic focus movement and ellipsis: A minimalist approach*. PhD Dissertation. University of Connecticut.
- Kim, Jeong-Seok, Hye Jin Lee, Su-Hyuk Yoon, Jin Hyung Lee, Jee Young Lee, and Soohyun Kwon. 2021. An experimental investigation of *why*-stripping in Korean. *Linguistic Research* 38(2): 271-299.
- Kim, Jong-Bok. 2017. *Why*-stripping in English and Korean: A direct interpretation approach. *Studies in Generative Grammar* 27(4): 733-751.
- Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13): 1-26.
- Lobeck, Anne. 1999. VP ellipsis and the minimalist program: Some speculations and proposals. In Shalom Lappin and Elabbas Benmamoun (eds.), *Fragments: Studies in ellipsis and gapping*, 98-123. Oxford: Oxford University Press.
- Mehler, Jacques. 1963. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior* 2(4): 346-351.
- Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.
- Merchant, Jason. 2008. An asymmetry in voice mismatch in VP-ellipsis and pseudogapping. *Linguistic Inquiry* 39(1): 169-179.

- Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44(1): 77-108.
- Poppels, Till and Andrew Kehler. 2019. Reconsidering asymmetries in voice-mismatched VP ellipsis. *Glossa: A Journal of General Linguistics* 4(1): 60. 1-22.
- R Core Team. 2020. *R: A language and environment for statistical computing* (Version 3.6.3) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.
- Rohde, Hannah and Andrew Kehler. 2014. Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience* 29(8): 912-927.
- Ross, John Robert. 1969. Guess who? *Proceedings of the Fifth Annual Meeting of the Chicago Linguistic Society (CLS 5)*, 252-286. Chicago, IL: Chicago Linguistic Society.
- Sag, Ivan. 1976. *Deletion and logical form*. PhD Dissertation. Massachusetts Institute of Technology.
- SanPietro, Steven A., Ming Xiang, and Jason Merchant. 2012. Accounting for voice mismatch in ellipsis. In Nathan Arnett and Ryan Bennett (eds.), *Proceedings of the 30th West Coast Conference on Formal Linguistics (WCCFL 30)*, 32-42. Somerville, MA: Cascadilla Proceedings Project.
- Schachter, Paul. 1977. Does she or doesn't she? *Linguistic Inquiry* 8(4): 762-767.
- Schütze, Carson T. and Jon Sprouse. 2013. Judgment data. In Robert J. Podesva and Devyani Sharma (eds.), *Research methods in linguistics*, 27-50. Cambridge, MA: Cambridge University Press.
- Sprouse, Jon, Troy Messick, and Jonathan David Bobaljik. 2022. Gender asymmetries in ellipsis: An experimental comparison of markedness and frequency accounts in English. *Journal of Linguistics* 58(2): 345-379.
- Van Valin, Robert D. and Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge, MA: Cambridge University Press.
- Weir, Andrew. 2014. *Why*-stripping targets voice phrase. In Hsin-Lun Huang, Amanda Rysling, and Ethan Poole (eds.), *Proceedings of the 43rd North East Linguistics Society (NELS 43)*, 235-248. Amherst, MA: GLSA Publications.
- Yoshida, Masaya, Chizuru Nakao, and Iván Ortega-Santos. 2015. The syntax of *why*-stripping. *Natural Language and Linguistic Theory* 33(1): 323-370.
- Zehr, Jérémy and Florian Schwarz. 2018. *PennController for internet based experiments (IBEX)*. Retrieved July 8, 2023, from <https://doi.org/10.17605/OSF.IO/MD832>

Jeong-Seok Kim

Professor

Department of English Language and Literature

Korea University

145 Anam-Ro, Seongbuk-Gu,

Seoul, 02841 Korea

E-mail: uconnkim@korea.ac.kr

Received: 2023. 09. 06.

Revised: 2023. 11. 06.

Accepted: 2023. 11. 07.