Linguistic Research 42(Special Edition): 291-316
DOI: 10.17250/khisli.42..202509.011

High vs. low negation in English NPQs: How Korean EFL and ESL learners diverge

Keunhyung Park
(Kyungpook National University)

Park, Keunhyung. 2025. High vs. low negation in English NPQs: How Korean EFL and ESL learners diverge. *Linguistic Research* 42(Special Edition): 291-316. This study investigates how Korean learners of English interpret positive polar questions (PPQs) and the two structurally distinct negative polar questions (NPQs). Thirty classroom-based EFL learners in Korea and sixty immersion-based ESL learners in the United States completed a series of three online experiment. Each task paired an unambiguous visual prompt with either a PPQ, a high-negation NPQ (*Didn't you see…?*), or a low-negation NPQ (*Did you really not see…?*). Response times and answering patterns reveal a clear dissociation. Immersive exposure yields near-native performance on the high-negation NPQs, yet both learner groups—most strikingly the higher-proficiency ESL cohort—show severe difficulty with low-negation NPQs, producing up to 74 % unexpected answers and significantly slower responses. These results challenge the traditional "polarity- vs. truth-based" dichotomy, demonstrating that acquisition hinges on input frequency and explicit form–meaning mapping rather than overall proficiency. Pedagogically, targeted instruction on inner-VP negation is recommended. **(Kyungpook National University)**

Keywords L2 English, L1 Korean, L2 acquisition, EFL, ESL, negative polar question

1. Introduction

It is widely recognized in literature that polar questions (PQs) are present in all languages around the world and typically consist of both positive polar questions (PPQs) and negative polar questions (NPQs). While PPQs are relatively easy to interpret and their truth conditions can be determined easily, the meaning of NPQs can be more ambiguous, and the interpretation of simple yes-noanswers to NPQs can be unpredictable, according to several studies (Ladd 1981; Kramer and Rawlins

2009; Holmberg 2013; Sudo 2013; Claus et al. 2017; Kim 2017; Krifka 2017; Park and Dubinsky 2019; Wee 2019; Kim 2024; Kim et al. 2024).

Previous studies on the interpretation of NPQs have proposed that languages are typologically classified into "polarity-based" and "truth-based" languages based on how *yes-no* answers to NPQs are interpreted (Pope 1976; Jones 1999). Under this traditional view that languages have distinct interpretations of NPQs, literature has reported that most L2 learners have difficulties in the acquisition of L2 NPQs because of the L1-L2 differences. Given that the answering dichotomy is widely accepted in the literature, how L2 learners acquire native-like attainment of L2 NPQ if L1 NPQ is different from the target language has been a major question. However, upon closer examination of answering patterns to NPQs in various languages, it has been recognized that the traditional classification may have been a premature generalization. In other words, answering patterns are not strictly distinguished into two language groups. In the current study, we will examine how languages have two different types of NPQs with either polarity- or truth-based interpretations, which challenges the traditional classification. Casting doubt on the traditional typology, several recent studies have raised concerns about the ambiguous interpretations of NPQs in diverse languages, where both interpretations are possible depending on the structure of the NPQs.

The goal of this study is to analyze how Korean individuals learning English as a second/foreign language respond to PQs in L2 English. The study is specifically interested in identifying the developmental differences between two groups of learners: those learning English as a foreign language (EFL) in Korea and those learning English as a second language (ESL) in the United States. In Section 2, this paper first considers similarities and differences between English and Korean NPQs, that may lead to misinterpretations when questions are translated from one language to the other. In Section 3, we examine some limitations in the traditional dichotomy of answering patterns and report common misunderstandings in the previous L1/L2 studies regarding the acquisition of NPQs. In Section 4 and 5, the paper presents the results of a series of online experiments that measure L2 participants' response times (RTs) and unexpected responses (UERs) to English PQs under linguistically decontextualized conditions. Finally, Section 6 discusses why two different L2 groups show distinguished developmental sequences in responding L2 English NPQs and establish research agenda for future studies.

2. English and Korean PPQs and NPQs

It should be noted, first, that how the traditional dichotomy of answering patterns deal with English and Korean *yes-no* questions and felicitous answers to them. Previous accounts of NPQ interpretation in two languages note that their interpretations are the opposite of each other. That is, a *yes* answer to English NPQs tends to confirm the positive assertion and be interpreted as if negation were not there. Contrastively, a *yes* answer to Korean NPQs tends to confirm a negation of the underlying assertion, taking negation as a meaningful contribution to the question proposition. This contrast can be observed explicitly in examples (1)-(4), which illustrate the interpretation of short affirmative or negative responses to positive and negative questions.

(1) Q: Did you see him? (English PPQ)
A: Yes. = 'I did see him.' / No. = 'I did not see him.'

(2) Q: Did**n't** you see him? (English NPQ)
A: Yes. = 'I did see him.' / No. = 'I did not see him.'

(3) Q: ne-nun ku-lul po-ass-ni? (Korean PPQ)
you-TOP he-ACC see-PST-Q
'Did you see him?'
A: ung. = 'I did see him.' / ani. = 'I did not see him.'
yes no

(4) Q: ne-nun ku-lul **an** po-ass-ni? (Korean NPQ)
you-TOP he-ACC NEG see-PST-Q
'Didn't you see him?'
A: ung. = 'I did not see him.' / ani. = 'I did see him.'
yes no

Many researchers have, heretofore, claimed that these different patterns to be typological. That is, as in (2), English ignores the negation in NPQ interpretations (a "polarity-based" language), and as in (4), Korean counts it (a "truth-based" language). A numerous number of previous studies on the interpretation of English NPQs have simply assumed that the answering patterns in (1) and (2) are the only possible question-answer pairs in English. However, a closer examination reveals that both languages tend to ignore negation when it is "high" in the NPQ, outside VP,

and to count it when it is "low", inside/adjacent to VP. For instance, as shown in (5) and (6), we could find numerous examples of low-negation NPQ in English and high-negation NPQ in Korean, respectively, which show opposite interpretations to the typical English high-negation NPQ in (2) and Korean low-negation NPQ in (4).

(5) Q: Did you really **not** see him? (English Low-NPQ)
A: Yes. = 'I really didn't see him.'/ No. = 'I did see him.'

(6) Q: ne-nun ku-lul po-ci **an**h-ass-ni? (Korean High-NPQ)
he-NOM he-ACC see-NMLZ NEG.do-PST-Q
'Didn't you see him?'
A: ung. = 'I did see him.' / ani. = 'I did not see him.'
yes no

In (5) and (6), the meanings of short *yes* and *no* answers are opposite to those of answers in (2) and (4). One who adheres solely to the conventional typology of response systems may struggle to account for the atypical response patterns in (5) and (6). Despite the fact that each of these languages possesses both high and low negations, each shows a clear preference for one type of NPQ over the other, resulting in differing degrees of ambiguity—and these preferences are directly opposed to one another. This observation is that, rather than being attributable to any "typological" difference, the divergent interpretive inclinations of the two languages seem due to the fact that English high-negation NPQs as in (2) being used more frequently and being less ambiguous than those in (5). Conversely, in Korean, the situation is reversed. Korean low-negation NPQs as in (4) are more frequent in natural language and less ambiguously have a negative meaning than those in (6).

The inherent ambiguity of NPQs and cross-linguistic discrepancy give rise to several interesting inquiries about the contrast between how native and non-native speakers interpret them. Although English and Korean have both high- and low-negation in common, each language has its own unique structures of NPQs. In English, two distinct NPQs are presented as contracted and non-contracted NPQs, while in Korean, the two equivalents are expressed as LFN- and SFN-NPQs. The distinguished linguistic features and the skewed use of NPQs by native speakers consequently affect the proper interpretation of two NPQ forms in each language. These cross-linguistic differences are likely to cause severe misunderstanding if L2

learners do not have sufficient proficiency to distinguish syntactic differences of two NPQ structures and do not understand the subtle nuance of ambiguous NPQs in the target language. Moreover, if L2 learners have not received adequate instruction on the meaning of two NPQ forms, it would be incredibly challenging to acquire native-like proficiency (i.e., in rapidity and accuracy). In the following section, we highlight several issues related to the traditional dichotomy of answering pattern that have not been addressed in prior studies on L2 acquisition.

3. Literature review and critique

3.1 From answering-system typology to structure-sensitive accounts

Early work portrayed NPQ interpretation as a binary opposition between truth-based systems (e.g., Korean, Japanese, Chinese) and polarity-based systems (e.g., English, Spanish). In this view, *yes* either agrees with the truth value (truth-based) or simply affirms positive polarity while ignoring sentential negation (polarity-based). Acquisition studies in the 1980s–1990s therefore asked whether children and L2 learners must "re-map" an L1 answering system when learning an L2 with the opposite pattern, often citing contrasts such as Korean children answering *Are you not a baby?* with *Yes, I'm not* versus English children's *No, I'm not* (Pope 1976; Kim 1985; Choi 1991; Akiyama 1992; Jones 1999). The premise that English NPQs are semantically equivalent to PPQs—negation being inert—underpinned these claims.

Subsequent theoretical research has shown that this typology is too coarse. Following Ladd (1981), English distinguishes outer negation (e.g., *Didn't you see him too?*) and inner negation (e.g., *Didn't you see him either?*) yielding opposite short-answer patterns. Comparable structural splits have been argued cross-linguistically (Büring and Gunlogson 2000; Romero and Han 2004; Kramer and Rawlins 2009; Holmberg 2013, 2016; Sudo 2013; Krifka 2017). Converging experimental evidence likewise indicates structure-sensitive interpretation among native speakers: English-speaking children master tag questions before high-negation NPQs and struggle longest with low-negation forms (Woods and Roeper 2020, 2021), and eye-tracking shows that listeners entertain both p and $\neg p$ alternatives across NPQ types (Tian et al. 2021). Taken together, these findings suggest that the position of

negation, pragmatic licensing, and usage frequency—rather than an immutable typological "answering system"—govern default interpretations in English and beyond.

3.2 Methodological pitfalls and design implications

A re-examination of prior L1/L2 studies on NPQs reveals three recurring problems. (i) Stimulus ambiguity: items such as *Are you not…?/ Isn't it a cup?* were treated as the same low-negation NPQs, though they are structurally ambiguous between high- and low negation and interpreted differently depending on the conversational context; answers classified as "errors" often match a high-negation parse (e.g., *Yes, it is* to *Isn't it a cup?*). (ii) Partial cross-language comparisons: many studies simply contrasted English high-negation NPQs with Korean (or Chinese) low-negation NPQs, omitting the structurally matched counterparts and thereby guaranteeing a difference. (iii) Over-interpretation of developmental data: inconsistent responses by children or L2 learners were taken as non-acquisition, despite evidence that even monolingual children oscillate until complex semantic and pragmatic factors with regard to the ambiguous NPQs are mastered; adult baselines also show instability for low-negation NPQs (Park and Dubinsky 2019; Woods and Roeper 2021).

These pitfalls motivate the present design. We (a) balance the stimulus set by testing PPQs, high-negation NPQs, and low-negation NPQs; (b) eliminate discourse bias via linguistically decontextualized but ontologically certain visual prompts; (c) compare exposure profiles by testing Korean EFL and ESL learners; and (d) separate processing cost from categorical misunderstanding using response times and unexpected-answer rates. This structure-sensitive, frequency-aware approach directly addresses the limitations of earlier work and prepares the ground for the methodological details in Section 4.

4. Current study

A series of experiments aims to clarify how Korean L2 learners of English interpret English NPQs and compare their responses to those of native speakers. By comparing two distinct L2 learner groups, who differ in their level of proficiency and type of exposure to the language, this study will show how L2 learners gradually acquire

semantically and pragmatically ambiguous structures. To achieve these objectives, the study will:

- Present results from a set of online experiments designed to elicit L2 judgments on positive and negative polar questions under linguistically decontextualized conditions.
- Analyze how Korean learners of English—residing either in the United States or in Korea—interpret English PQs.
- Assess the influence of English negation on learners' comprehension of the two structurally distinct types of NPQs.

4.1 Expectation

As examined in previous sections, both L1 Korean and L2 English show two distinct positions of negation in NPQ constructions. It would be a reasonable assumption that both L1 and L2 response to be identically sensitive to the position of negation in NPQs. Since both languages allow negation to appear in different positions (inside vs. outside VP), these positions appear to correlate with them being interpreted either internal or external to the questioned proposition. If L2 learners are aware of the grammar and naturally follow the syntactic distinction, Korean learners of English should show very similar answering patterns to NPQs in L1 Korean and L2 English. However, L2 learners who have underdeveloped L2 proficiency may not be able to distinguish between two NPQ structures as they do in their L1. Therefore, it is expected that some L2 learners may answer all English NPQs as low-negation NPQs because of the presence of negation in the question (without regard to the semantics and pragmatics of the negation), or they might interpret all NPQs randomly, and show themselves to be unaware of the differences between high- and low-negation. In either case, there should be a significant gap between L1 and L2 interpretations.

From the baseline experiment with L1 speakers of English and Korean (Park and Dubinsky 2019), it is already ascertained that even L1 speakers are confused with certain NPQs and interpret them differently from the majority of L1 speakers, revealing the inherent ambiguity/complexity of NPQs. Thus, L2 learners might encounter great difficulties while acquiring a native-like understanding of NPQs. With regards to the

issues of teachability and learnability issues in the acquisition of NPQs, it is expected that Korean learners of English who have recently exposed to English as a foreign language in classroom-based language learning settings in their home country (e.g., EFL context) will face more challenges compared to L2 learners who have exposed to authentic native English in the USA (e.g., ESL context).

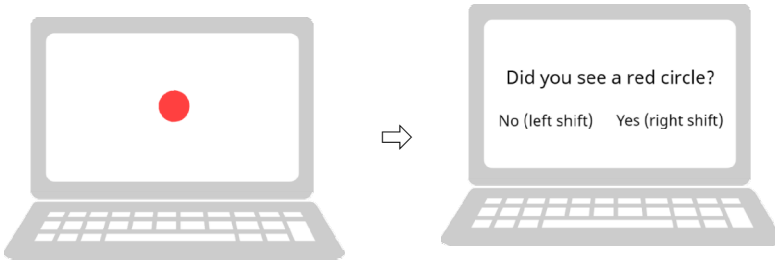
4.2 Hypotheses and method

We tested L2 interpretation of English PQs conducting a series of online experiments which was utilized in Park and Dubinsky (2019). The central hypothesis was that morpho-syntactic position of negation and pragmatic ambiguity increase processing cost, yielding longer RTs and more UERs. To examine learners' responses to PPQs, high-negation NPQs (outer/high), and low-negation NPQs (inner/low), we ran three experiments:

- Experiment 1 tests PPQs vs. high-negation NPQs
- Experiment 2 tests PPQs vs. low-negation NPQs
- Experiment 3 tests PPQs vs. high-negation NPQs vs. low-negation NPQs

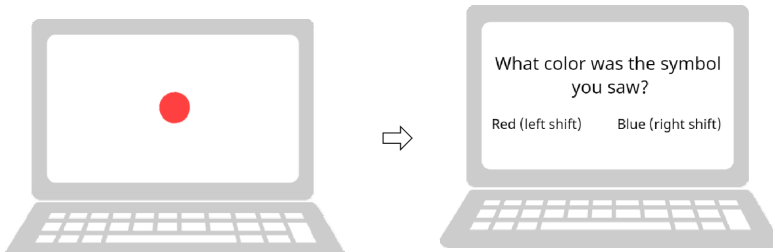
To control beliefs independently of discourse, each trial paired a linguistically decontextualized but ontologically certain visual prompt with a question. As shown in Figures 1–2, each item comprised two slides: (1) a symbol prompt (1000ms) showing one of five shapes (triangle, star, square, X, circle) in one of five colors (red, blue, black, yellow, green)[1], yielding 25 possible prompts; (2) a Q/A slide (5000ms) presenting the question and two on-screen response options.

1 The present test uses five different symbols in five different colors: red, yellow, green, blue, and black. According to Kay and Maffi (1999), these five colors are categorized in the six basic colors including white. These colors are most frequently used primary colors in psychology, psycholinguistics, and neurolinguistics.



1st symbol prompt slide (1000ms) 2nd Q/A slide (5000ms)

Figure 1. Illustration of slides in experimental items



1st symbol prompt slide (1000ms) 2nd Q/A slide (5000ms)

Figure 2. Illustration of slides in filler items

Participants responded by pressing the left or right shift key; the relevant labels (e.g., *No*/*Yes*, *Red*/*Blue*) were displayed on every trial to minimize memory load. This design ensures that, for example, a red-circle prompt unambiguously supports the proposition *I saw a red circle*, allowing us to attribute variability in answers to the question type rather than to discourse context.[2]

2 In conversational English, bare *yes*/*no* answers to NPQs can be dispreferred when discourse leaves both and *p* and ¬*p* salient; full answers (e.g., *Yes, he is*/*No, he isn't*) then help disambiguate (cf. Holmberg 2013, 2016). Our experiments do not target production preferences. Instead, they test interpretation under decontextualized, materially certain evidence: each trial first presents an unambiguous visual prompt that fixes the proposition at issue. Within this setting, bare particles are felicitous and sufficient, and they prevent confounds from auxiliary morphology in full sentential answers.

4.3 Experimental stimuli

Critical items. The examples in (7) present English PQs used in the experiments. There are examples of a PPQ and two distinct forms of NPQs. The answer choices for such questions, as presented to participants in the experiment, are limited to simple *Yes* and *No*.[34]

(7) a. Did you see a red circle? (PPQ)
b. Did**n't** you see a red circle? (High-NPQ)
c. Did you really **not** see a red circle? (Low-NPQ)

Filler items. To prevent strategy formation and keep attention balanced across color and shape, we interleaved two kinds of fillers with the critical PQs as shown in (8). *Wh*-fillers querying the color (8a) or shape (8b) of the symbol; response options were two content alternatives printed on screen (e.g., *Red* vs. *Blue*, *Circle* vs. *Square*). Participants pressed the key matching the correct alternative. Tag-fillers (8c) consisting of a positive proposition with a negative tag; response options were *Yes*/*No* as with PQs.

(8) a. What color was the symbol you saw?
b. What shape was the symbol you saw?
c. You saw a red circle, didn't you?

---

3 Following Ladd (1981), high-negation strings (e.g., *Didn't John...?*) are between outer/high (Q about *p*) and inner/low (Q about ¬*p*) negation. Diagnostics show both readings are available—e.g., NPI *either* (inner) vs. PPI *too* (outer): *Isn't John coming either/too?* (see also Kramer and Rawlins 2009; Krifka 2017). In our materials, questions were bare (no NPI/PPI or discourse cues) within a decontextualized *yes*/*no* task, which standardly favors the outer-negation reading; we analyze them accordingly. If some participants nevertheless accessed the inner reading, their answers were counted as "unexpected," rendering our High-NPQ effects conversation.

4 Following Romeo and Han (2004), while "really" can function as a Verum operator, it does not by itself introduce a *p*/¬*p* ambiguity in low-negation NPQs (see their in footnote #17). In our stimuli, "really" was used only to mark low scope of negation; it was not intended to license an outer-negation *p* reading. Nevertheless, we acknowledge that "really" can convey an epistemic/Verum bias (Romero and Han 2004; Goodhue 2018, 2019), which can make such questions sound follow-up-like when the relevant proposition has just been visually established. This may contribute to longer RTs and higher UERs for Low-NPQs in our design. Accordingly, we emphasize within-task comparisons (Low- vs. High-NPQ) under identical presentation and response demands. We greatly appreciate the anonymous reviewer for prompting this clarification.

Instructions and handling. Participants were explicitly told that *wh*-fillers required selecting the correct alternative shown on screen, and that tag-fillers should be answered *Yes*/*No* according to the symbol just seen. Filler trials served as attention checks and were excluded a priori from RT and UER analyses.[5]

Item counts. Experiments 1–2 presented 24 PPQs and 24 NPQs plus 24 fillers (total = 72 items). Experiment 3 presented 24 PPQs, 24 high-NPQs, 24 low-NPQs, and 48 fillers (total = 120 items). Within each 24-item set of critical questions, 12 items were constructed to elicit *Yes* responses and 12 to elicit *No* responses. Each 24-item filler set contained 12 *wh*-fillers (balanced for color/shape) and 12 tag-fillers.

4.4 Participants

We analyzed L2 data from two groups of native Korean speakers learning English, chosen to represent real-world input environments: classroom-based exposure in Korea (EFL) and immersive exposure in the United States (ESL). This grouping targets our primary question—whether exposure context relates to how learners interpret NPQs—while recognizing that proficiency distributions also differ across these contexts. We therefore treat group differences as the joint effect of exposure + proficiency, characterize proficiency descriptively. We return to this limitation explicitly in Section 7.

EFL group (n = 30). Participants were tested in South Korea. None had prior residence or study abroad in an English-speaking country. Their English experience was classroom-based, with limited opportunities for out-of-class use; exposure was predominantly receptive (reading, listening) rather than productive (speaking, writing).

ESL group (n = 60). Participants were tested in the USA and had been living/studying there for ≥ 12 months at the time of testing. All were born and educated (K-12) in South Korea and were enrolled as undergraduate or graduate students (e.g., University of South Carolina; Iowa State University), receiving immersive day-to-day input.

---

5 An anonymous reviewer asked whether interleaving *wh*-questions and tag questions with *yes*/*no* NPQs might influence processing times or answer tendencies. Two clarifications: (i) filler trials were randomly intermixed, served only as attention checks, and were excluded from all RT and UER analyses; (ii) *wh*-fillers displayed two on-screen alternatives (e.g., *red*/*blue*), and tag-fillers used *Yes*/*No* options, and participants responded to them rapidly with near-ceiling accuracy.

Proficiency characterization and screening. Groups were not assigned by proficiency, but the ESL cohort typically occupies a higher proficiency range. Where available, standardized scores were recorded: ESL participants commonly met thresholds of TOEIC ≥ 700 or TOEFL iBT ≥ 80. EFL participants generally ranged from basic to low-intermediate based on program placement. These measures are reported to describe the samples and, where possible, are used as continuous covariates to assess whether key NPQ effects hold over and above individual proficiency differences.

Interpretive scope. By comparing learners across input environments while controlling/acknowledging proficiency, we ask whether immersive exposure facilitates native-like mappings for high- vs. low-negation NPQs. Because exposure and proficiency are not fully separable in this design, causal claims about exposure alone are tempered; we outline matched/within-group follow-ups as future work in Section 7.

5. Results

5.1 Data preparation and model choice

Analyses were conducted on the complete trial-level dataset. Each experiment comprised 10 EFL and 20 ESL participants. In Experiments 1 and 2, participants completed 48 trials each, consisting of 24 PPQs paired with either 24 High-NPQs or 24 Low-NPQs, respectively. In Experiment 3, participants completed 72 trials, with 24 of each type. This design yielded a total of 5,040 trials (see Table 1 for descriptive statistics).

Given the repeated-measures structure of the data, mixed-effects modeling was employed. For the RT data, which were generally positively skewed (see Figure 3), residual diagnostics indicated that log transformation was more appropriate than using raw values. Accordingly, a linear mixed-effects model with a Gaussian distribution was fitted to the log-transformed RT. The model included group (EFL vs. ESL), question type (PPQ, High-NPQ, Low-NPQ), and experiment (1–3) as fixed effects, with random intercepts for participants and items. For the UER data, a binomial generalized linear mixed-effects model with a logit link was fitted to model the probability of an unexpected response, using the same fixed and random effects structure.

Table 1. Descriptive statistics for response time (ms) and unexpected responses by group and question type

Group	Q-type	No. of trials	No. of subj.	Mean RT	SD of RT	Median RT	IQR of RT	No. of UERs
EFL	PPQ	720	30	1552	525	1477	666	31
EFL	HNPQ	480	20	1882	828	1718	1046	97
EFL	LNPQ	480	20	2392	866	2363	1082	243
ESL	PPQ	1440	60	1475	556	1383	632	48
ESL	HNPQ	960	40	1704	806	1516	914	94
ESL	LNPQ	960	40	2390	1002	2218	1348	624

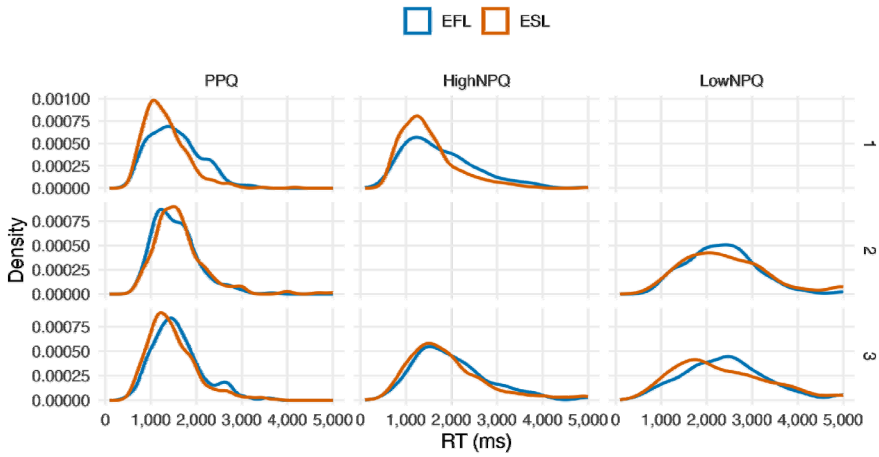


Figure 3. Response time (RT) densities by experiment, question type, and group

5.2 Response times (RT)

A linear mixed-effects model on log RT (Satterthwaite df) revealed a robust main effect of question type. Estimated marginal means (EMMs) contrasts indicated that responses to PPQs were faster than to High-NPQs (EMM contrast = −0.161 on the log-RT scale, SE = 0.0186, df = 84.6, t = −8.64, p < .0001, 95% CI [−0.205, −0.116]) and faster than to Low-NPQs (−0.388, SE = 0.0169, df = 90.5, t = −23.02, p < .0001, 95% CI [−0.429, −0.348]). High-NPQs were also faster than Low-NPQs (−0.228, SE = 0.0155, df = 725.3, t = −14.65, p < .0001, 95% CI [−0.264, −0.191]).

Back-transforming these log differences corresponds to approximate reductions of 15% (PPQ vs. High-NPQ), 32% (PPQ vs. Low-NPQ), and 20% (High-NPQ vs. Low-NPQ) in RT on the original scale.

There was no reliable main effect of group (EFL vs. ESL; β= 0.030, SE = 0.0265, df = 90.44, t = 1.13, p = .261), and the group × question type interactions were not significant (both p-values ≥ .81), indicating that the question type pattern was comparable across groups (as illustrated in Figure 4). Experiment 1 produced faster responses than Experiment 2 (EMM contrast on the log-RT scale = −0.1815, SE = 0.0632, df = 97.7, t = −2.87, p = .0137), which corresponds to an approximately 16.6 percent reduction in RT on the original scale. Experiment 1 was also marginally faster than Experiment 3 (contrast = −0.1458, SE = 0.0626, df = 94.5, t = −2.33, p = .0568), an approximately 13.6 percent reduction.Experiment 2 and Experiment 3 did not differ reliably (contrast = 0.0357, SE = 0.0629, df = 96.0, t = 0.57, p = .838).

Random-effects variances were 0.054 (SD = 0.233) for participants and 0.0025 (SD = 0.050) for items, with residual variance 0.093 (SD = 0.305), based on 5,040 observations (90 participants; 72 items). Intraclass correlation coefficients (ICC; from the no-interaction model) indicated that 36.3% of the variance in log RT was attributable to between-participant differences and 1.7% to between-item differences, with the remainder due to residual variation.

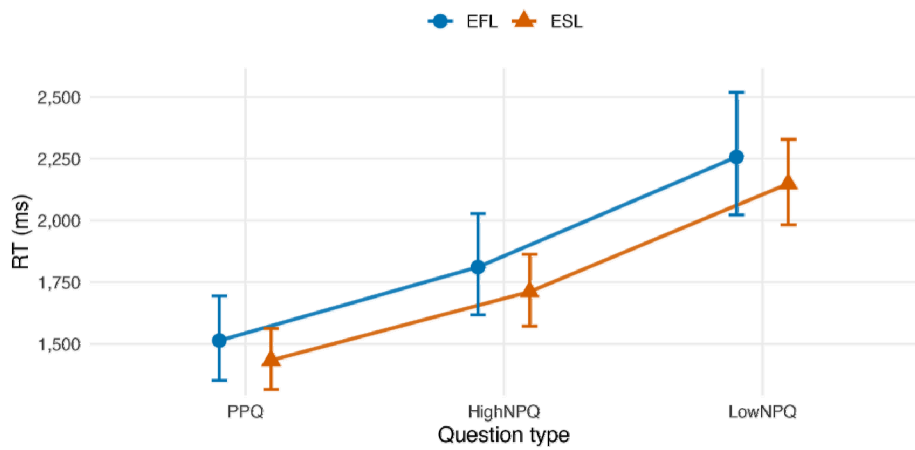


Figure 4. Estimated marginal mean RT (ms)

5.3 Unexpected responses (UERs)

The binomial mixed-effects model results showed that there was a strong effect of question type that depended on group. The group × question type interaction was significant for the Low-NPQ versus High-NPQ comparison (interaction estimate = 0.369, SE = 0.086, z = 4.30, p < .001), while the interaction involving PPQ versus High-NPQ was not significant (p = .871). Main effects of group and experiment were not statistically reliable (all p-values ≥ .58).

Estimated marginal means (on the probability scale) clarified the pattern. Predicted UERs were very low for PPQs (EFL: 1.64%, 95% CI [0.83, 3.22]; ESL: 1.36%, [0.78, 2.39]), moderate for High-NPQs (EFL: 15.3%, [8.81, 25.2]; ESL: 6.87%, [4.14, 11.2]), and high for Low-NPQs (EFL: 53.6%, [40.4, 66.3]; ESL: 68.1%, [58.3, 76.5]) (see Figure 5).

Planned within-group contrasts quantified these differences. In the EFL group, Low-NPQ trials were much more likely than High-NPQ trials to elicit an unexpected response (odds ratio [OR] = 6.41, 95% CI [3.77, 10.9], p < .001), whereas PPQs were much less likely than High-NPQs (OR = 0.093, 95% CI [0.045, 0.189], p < .001). In the ESL group, the Low-vs-High contrast was markedly larger (OR = 29.0, 95% CI [17.8, 47.0], p < .001), and the PPQ-vs-High reduction, though still substantial, was smaller than in EFL (OR = 0.188, 95% CI [0.096, 0.366], p < .001).

Taken together, question type robustly modulated the probability of an unexpected response, and this effect differed by group: ESL participants showed a much steeper increase from High-NPQ to Low-NPQ than EFL participants, while both groups exhibited very low UERs for PPQs.

Random-effects variances were 1.302 (SD = 1.141) for participants and 0.617 (SD = 0.785) for items, based on 5,040 observations (90 participants; 72 items). Intraclass correlation coefficients (ICC) indicated that 26.2% of the variance in the log-odds of an unexpected response was attributable to between-subject differences and 11.5% to between-item differences, with the remainder due to trial-level variability on the latent (logit) scale.

In summary, UERs increase with polarity complexity, following the pattern PPQ < High-NPQ ≪ Low-NPQ for both groups. Notably, immersion (ESL) enhances performance on High-NPQs but has no effect on reducing Low-NPQ errors.

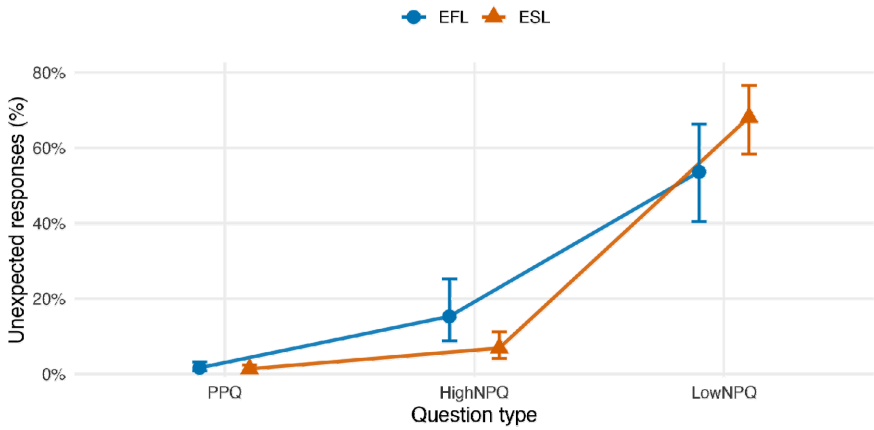


Figure 5. Estimated marginal UER

6. Discussion

RTs and UERs collected from a series of experiments show that L2 learners of English are fully aware of the semantic difference between PPQs and NPQs and respond differently to each question type. Responding to PPQs is quite straightforward and does not demand a high processing cost in choosing either *yes* or *no* answer. Most L2 learners from both groups were able to choose the correct answer quickly and accurately for PPQs. However, when negation was involved in the two distinct types of NPQs, high-negation NPQs and low-negation NPQs, two L2 groups showed distinguished developmental trajectories. The dichotomy of high-low NPQs, which is consistent with the structure of NPQs, resulted in significant differences in L2 learners' RTs and UERs.

It is still unclear why advanced ESL learners have difficulty exclusively with low-negation NPQs. It could be because these forms of English NPQs are not taught adequately in the L2 classroom, or the L2 learners may lack explicit knowledge of the distinction. Additionally, it is somewhat uncertain if this grammatical knowledge can be taught or if the current L2 participants are not yet capable of understanding the two NPQ structures. Further research is needed to address these questions

In Experiment 1, when PPQs and high-negation NPQs were presented

simultaneously, two L2 groups displayed different but anticipated RTs and UERs. The more proficient participants in the ESL group responded more accurately and quickly to high-negation NPQs than the less proficient participants in the EFL group. Analysis of the response patterns revealed by the UERs indicated that the ESL learners' responses to high-negation NPQs closely resembled those of native English speakers. In other words, ESL learners who have been exposed to native English for a long time in the USA are fully aware that negation in high-negation NPQs does not affect the truth of the NPQ, and their response pattern to high-negation NPQs is similar to that of PPQs. In contrast, EFL learners who have low proficiency and have only been exposed to English in a classroom setting make more UERs and take longer time in finding a correct *yes*/*no* answer. Two participants in EFL group answered most high-negation NPQs in the opposite pattern, and this was extremely rare among both L1 speakers and ESL learners. In sum, ESL group was much more native-like in answering high-negation NPQs.

Interestingly, the nativelikeness, which is consistent with L2 learners' proficiency, was completely reversed in Experiment 2 and 3. In these experiments, low-negation NPQs were presented along with other types of questions. Since ESL group was much better than EFL group in Experiment 1, the results from Experiment 2 and 3 should have shown that ESL learners are able to answer all question types better than EFL learners. However, it was consistently detected that ESL learners took longer time and committed more UERs in answering low-negation NPQs than less proficient EFL learners. This unexpected result cannot be explained by their native-like answering patterns to high-negation NPQs in Experiment 1. Due to their lower proficiency (as confirmed in Experiment 1 with slower RTs and higher UERs to high-negation NPQs), it was expected that EFL learners would have more difficulty answering low-negation NPQs in Experiments 2 and 3, which are the most complex and ambiguous type of questions. Two EFL participants almost oppositely answered low-negation NPQs, and the worst case was that many EFL learners randomly answered low-negation NPQs, resulting in inconsistent responses between two opposite answering patterns. Nevertheless, the responses of the ESL group to low-negation NPQs were even more problematic than those of the EFL group.

6.1 The influence of L1 transfer?

Why did ESL learners perform worse than EFL learners specifically in answering low-negation NPQs? Are they less proficient to properly answer English low-negation NPQs than EFL learners? If it is the case, how did they show better (almost native-like) answering choices to high-negation NPQs in Experiment 1? One might assume that L2 interpretation of NPQs can be ruined by L1-L2 differences. As discussed in Section 2, L1 Korean has two different forms of NPQs, SFN-NPQ and LFN-NPQ, which are usually answered in opposite ways. When compared to the English NPQs utilized in the current experiments, Korean SFN-NPQs are typically interpreted as English low-negation NPQs, and Korean LFN-NPQs are typically interpreted as English high-negation NPQs. Repeated below, answering patterns to English and Korean NPQs in (9) and (10) are parallel, and those to NPQs in (11) and (12) are parallel each other.

(9) Q: Did**n't** you see him? (English high-NPQ)
A: Yes. ='I did see him.' / No. = 'I did not see him.'

(10) Q: ne-nun ku-lul po-ci **an**h-ass-ni? (Korean LFN-NPQ)
you-NOM he-ACCsee-NMLZ NEG.do-PST-Q
'Didn't you see him?'
A: Ung. = 'I did see him.' / Ani. = 'I did not see him.'
yes no

(11) Q: Did you really **not** see him? (English low-NPQ)
A: Yes/ = 'I really didn't see him.' / No. = 'I did see him.'

(12) Q: ne-nun ku-lul **an** po-ass-ni? (Korean SFN-NPQ)
you-NOM he-ACC NEG see-PST-Q
'Didn't you see him?'
A: Ung. = 'I did not see him.' / Ani. = 'I did see him.'
yes no

Both Korean NPQs in (10) and (12) are frequently used in everyday conversation, so Korean participants are already very familiar with these forms of NPQs. Despite the observable fact that English and Korean have both structures of NPQs, it should

be noted that, being attributable to the frequency of use in natural language, the contrary interpretive tendencies of the two languages seem due to English high-negation NPQs (e.g., (9)) being used more frequently and being less ambiguous than low-negation NPQs (e.g., (11)). For Korean, in contrast, SFN-NPQs (e.g., (12)) are more frequent and less ambiguous than LFN-NPQs (e.g., (10)). If Korean participants' L1 affects L2 interpretation of English NPQs, they should obviously be conscious of the distinction between two English NPQs. If Korean speakers' native language affects their understanding of English NPQs (i.e., L1 transfer), Korean learners of English must be more biased toward English low-negation interpretation which is similar to the more frequent Korean SFN-NPQ interpretation. However, ESL learners in the present study struggled mainly with low-negation NPQs, while their interpretation of high-negation NPQs was almost native-like. This asymmetry reveals that ESL learners specifically face difficulties when answering low-negation NPQs. The significantly higher UERs for low-negation NPQs by ESL group were attributed to the unexpected answering pattern that ESL learners very systematically oppositely answered low-negation NPQs as high-negation NPQs.

6.2 Poverty of the stimulus

Regarding Korean L2 learners' uniform answering patterns to two forms of English NPQs, we could find an answer in post-test interviews. Some ESL and EFL participants mentioned that they had learned that English NPQs should be interpreted oppositely to Korean NPQs from non-native English teachers in middle and high school. During the experiment, these participants consciously tried to answer all English NPQs in the same way regardless of the forms of NPQs. That is, they recognized that there were two distinct NPQs in the experiment, but they anyway treated them as the same English NPQs following the grammar they were taught in school. The idea that English and Korean NPQs are answered oppositely is the key mechanism in the traditional answering dichotomy. In many cases, this simplification of the dichotomy would cause serious misunderstanding for L2 learners when interpreting NPQs in foreign or second languages.

Other participants in the experiment also expressed concerns about the ambiguity of English NPQs. Many ESL participants admitted that they had been consistently

confused when native English speakers used NPQs in natural conversation. Despite having lived in the USA for an extended period and self-evaluating their English proficiency as adequate, the use of English NPQs had always been a source of confusion for them. They also reported previous experiences that they often chose a wrong answer, which was opposite to their own intention, and then they recast the answer immediately by themselves. Even in worse cases, questioners who asked an NPQ sometimes tried to ask again to make sure the meaning of the *unexpected* answer. Then, L2 learners might realize that they had provided an incorrect response.

Our experimental results, together with post-test interviews, indicate that the distribution and quality of L2 input are crucial for ultimate attainment. In debates on the "poverty of the stimulus" (PoS)—the extent to which learners can acquire grammar without robust, direct evidence (Chomsky 1965, 2012)—L1 acquisition is often cited as demonstrating success despite limited explicit input. By contrast, our data suggest that adult L2 acquisition of interface-heavy phenomena such as NPQs is strongly constrained by how often and how transparently relevant forms appear in the input.

In particular, even advanced ESL learners failed to interpret low-negation NPQs in a native-like way. They consistently produced opposite answers, mirroring a school-taught "English NPQs are answered opposite to Korean" rule rather than mapping answers to the actual truth-conditions. This pattern was reflected in slower RTs and higher UERs in Experiments 2-3. Despite at least a year of immersion, these learners appear not to have encountered enough clear instances of low-negation NPQs —forms that are rare and pragmatically marked in everyday English—to override the classroom generalization. The finding supports a frequency-and-salience bottleneck account: syntactically complex, pragmatically nuanced, and infrequent constructions may not be acquired from exposure alone, even in immersive settings.

This conclusion does not imply that adult L2 learners are categorically unable to acquire low-frequency phenomena. Work on German scrambling and remnant movement shows that English learners can master structures that are both rare and absent in the L1 (Schreiber and Sprouse 1998; Hopp 2002, 2005). The contrast with NPQs is instructive: scrambling provides clear structural cues and relatively unambiguous acceptability contrasts, whereas low-negation NPQs require integrating syntax, semantics, and pragmatics, and are further confounded by misleading pedagogical rules. A plausible interpretation is that NPQs pose a tri-factor challenge—

low frequency, high interface complexity, and competing explicit instruction.

A practical implication follows: L2 learners may need targeted, explicit instruction that (i) contrasts high- vs. low-negation NPQs, (ii) practices short-answer mappings with minimal pairs, and (iii) raises the salience of inner-VP *not* (e.g., input enhancement). Future work should test whether such interventions reduce UERs for low-negation NPQs and whether gains generalize beyond the laboratory.

6.3 Teachability, developmental readiness, and NPQs

Post-test interviews revealed that many participants had been taught a simple rule: *English NPQs should be answered opposite to Korean NPQs*. This raises two questions: (i) can explicit classroom instruction guide sequential development in L2 NPQ interpretation, and (ii) can adult learners acquire the relevant mappings through exposure alone?

Pienemann's Teachability Hypothesis (1989) predicts that instruction is effective only if learners are developmentally ready for the next processing operation. In other words, the order of acquisition is constrained by processing prerequisites, even if the rate can be accelerated by teaching. Structures whose processing depends on operations at Stage X + 2 cannot be stably acquired by learners still at Stage X or X + 1.

Evidence from L1 English supports a staged progression in the NPQ domain. Children produce PPQs productively, then tag questions, later high-negation NPQs with clitic *n't*, and only much later—if at all—low-negation NPQs with inner-VP *not* (Woods and Roeper 2020, 2021). The last step is delayed because it requires integrating syntactic scope with pragmatic licensing. Viewed through this lens, our L2 results are consistent with a readiness-based sequence. Novice EFL learners often failed to differentiate NPQ types—consistent with being below the stage where negation height is reliably computed. Advanced ESL learners patterned like natives on high-negation NPQs (a later but still robust stage), yet remained non-native on low-negation NPQs. We attribute this asymmetry to a tri-factor hurdle: i) low-negation NPQs are infrequent, ii) interface-heavy (syntax–semantics–pragmatics), and iii) confounded by explicit but over-general classroom rules. Under the Teachability Hypothesis, instruction that simply repeats the "opposite answering" rule will not shift learners to the next stage; what is needed is instruction targeted at the required processing

operation (computing inner-VP negation and its truth-conditional mapping). A concise way to frame the developmental target is as in Table 3 below.

Table 3. The developmental stages in the acquisition of NPQs

Putative stage	Dominant operation	Representative form	Short-answer mapping
Stage1	Polarity mapping without negation	*Did you see…*?	Yes = p, No = $\neg p$
S2	Illocutionary bias handling	Tag questions	Bias-consistent confirmation
S3	Outer/high negation with clitic *n't*	*Didn't you see…*?	Yes = p, No = $\neg p$
S4	Inner/low negation (scope inside VP)	*Did you really not see…*?	Yes = $\neg p$, No = p

Instruction is most likely to help when learners are at Stage 3 and poised to acquire Stage 4, and materials contrast high- vs. low-negation forms with minimal pairs, require forced short-answers, and enhance inner-VP *not* (e.g., typographic highlighting, prosodic cues in audio). Brief metalinguistic prompts can be useful, but only when accompanied by practice that targets the Stage 4 operation rather than re-asserting a global "English vs. Korean" rule.

To test teachability directly, future work should (i) control prior instruction via a short survey and include it as a predictor in mixed-effects models; (ii) run a readiness-stratified pretest–teach–posttest, with Stage 3 vs. Stage 4 learners receiving the same targeted training on low-negation NPQs; and (iii) assess near- and delayed-post gains on both UERs and RTs. The Teachability Hypothesis predicts larger gains for learners already at Stage 3, with minimal improvement for those below Stage 3 even under explicit instruction.

7. Conclusion

In the present study, we conducted experiments to compare two L2 groups (EFL vs. ESL) and found that acquiring syntactically and semantically ambiguous structures is challenging for L2 learners. The data collected from online experiments revealed that L2 learners of English are aware of the semantic difference between PPQs and

NPQs and respond differently to each question type. Responding to PPQs is relatively straightforward, and most L2 learners quickly and accurately answered them. However, when negation is involved in the two distinct types of NPQs, Korean L2 learners of English showed different acquisitional trajectories depending on their proficiency level.

As expected, the ESL group outperformed the EFL group in terms of accuracy and speed in answering high-negation questions. The UERs show that ESL learners' responses to high-negation questions almost analogizes with those of native English speakers. This means that ESL learners who have been exposed to native English for a considerable period of time are fully aware that negation in high-negation questions is not usually considered in the truth condition of the NPQ. Interestingly, unlike the finding that ESL learners performed almost as well as native speakers in answering high-negation questions, their responses to low-negation questions were much worse than those of EFL learners. The high UERs generated by EFL learners can be attributed to their inadequate proficiency, leading to random responses to both types of English NPQs. On the other hand, ESL learners consistently answered low-negation questions as if they were high-negation questions. Many L2 participants reported that they were taught in middle and high schools that English NPQs should be answered opposite to Korean NPQs. As a result, they answered all English NPQs in the same way, regardless of their form.

It is still unclear why advanced ESL learners have difficulty exclusively with low-negation NPQs. It could be because these forms of English NPQs are not taught adequately in the L2 classroom, or the L2 learners may lack explicit knowledge of the distinction. Additionally, it is somewhat uncertain if this grammatical knowledge can be taught or if the current L2 participants are not yet capable of understanding the two NPQ structures. Further research is needed to address these questions.

Finally, we note that Low-NPQs were instantiated with *really*, which may be pragmatically marked even in our decontextualized task. Our High- vs. Low-NPQ conclusions rest on relative contrasts under matched conditions, but future work will test Low-NPQs without *really* (or with alternative adverbials) and in contextualized follow-up settings.

References

Akiyama, M. Michael. 1992. Cross-linguistic contrasts of verification and answering among children. *Journal of Psycholinguistic Research* 21: 67–85.

Büring, Daniel and Christine Gunlogson. 2000. *Aren't positive and negative polar questions the same?* Ms. University of California at Los Angeles and University of California at Santa Cruz.

Choi, Soonja. 1991. Children's answers to *yes-no* questions: A developmental study in English, French, and Korean. *Developmental Psychology* 27(3): 407–420.

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, Noam. 2012. Poverty of stimulus: Unfinished business. *Studies in Chinese Linguistics* 33(1): 3–16.

Claus, Berry, A. Meijer Marlijn, Sophie Repp, and Manfred Krifka. 2017. Puzzling response particles: An experimental study on the German answering system. *Semantics and Pragmatics* 10(19): 1–51.

Goodhue, Daniel. 2019. High negation questions and epistemic bias. In M. Teresa Espinal, Elena Castroviejo, Manuel Leonetti, Louise McNally and Cristina Real-Puigdollers (eds.), *Proceedings of Sinn und Bedeutung* 23(1), 469–486. Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès).

Goodhue, Daniel and Michael Wagner. 2018. Intonation, *yes* and no. Glossa 3(1): 1–45.

Holmberg, Anders. 2013. The syntax of answers to polar questions in English and Swedish. *Lingua* 128: 31–50.

Holmberg, Anders. 2016. *The syntax of yes and no*. Oxford: Oxford University Press.

Hopp, Hogler. 2002. Constraints on word order variation: Learnability and UG in advanced English–German and Japanese–German interlanguage. Unpublished MA thesis. University of Durham.

Hopp, Hogler. 2005. Constraining second language word order optionality: Scrambling in advanced English–German and Japanese–German interlanguage. *Second Language Research* 21: 34–71.

Jones, Bob Morris. 1999. *The Welsh answering system*. Berlin: Mouton de Gruyter.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1): 3–44.

Kay, Paul and Luisa Maffi. 1999. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist* 101(4): 743–760.

Kim, Jong-Bok. 2017. On the anaphoric nature of particle responses to the polar questions in English and Korean. *Korean Journal of Linguistics* 42(2): 153–177.

Kim, Jong-Bok. 2024. Parametric differences between English and Korean response particles: A discourse-based HPSG approach. *Studia Linguistica* 78(2): 280–314.

Kim, Jong-Bok, Jungsoo Kim, and Yunju Nam. 2024. Variations in answering negative polar

questions in Korean: An experimental study. *Lingua* 310: 103792.
Kramer, Ruth and Kyle Rawlins. 2009. Polarity particles: An ellipsis account. In Suzi Lima, Kevin Mullin, and Brian Smith (eds.), *Proceedings of the 39th Meeting of the Northeast Linguistics Society* (NELS) 39, 479–492. North Charleston, SC: CreateSpace Independent Publishing Platform.
Krifka, Manfred. 2017. Negated polarity questions as denegations of assertions. In Chungmin Lee, Ferenc Kiefer, and Manfred Krifka (eds.), *Contrastiveness in information structure, alternatives, and scalar implicatures*, 359–398. Dordrecht: Springer.
Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In Robert A. Hendrick, Carrie S. Masek, Mary Frances Miller (eds.), *Papers from the Seventeenth Regional Meeting of the Chicago Linguistic Society* (CLS) 17, 164–171. Chicago, IL: Chicago Linguistic Society.
Park, Keunhyung and Stanley Dubinsky. 2019. The syntax and semantics of negative questions and answers in Korean and English. In Patrick Farrell (ed.), *Proceedings of the Linguistic Society of America* (PLSA) 4(19), 1–9. Washington DC: Linguistic Society of America.
Pienemann, Manfred. 1989. Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics* 10: 52–79.
Pope, Emily. 1976. *Questions and answers in English*. The Hague: Mouton.
Romero, Maribel and Chung-Hye Han. 2004. On negative *yes*/*no* questions. *Linguistics and Philosophy* 27(5): 609–658.
Schreiber, Tim and Rex A. Sprouse. 1998. Knowledge of topicalization and scrambling in English –German interlanguage. *McGill Working Papers in Linguistics* 13: 162–72.
Sudo, Yasutada. 2013. Biased polar questions in English and Japanese. *Beyond expressives: Explorations in use-conditional meaning*: 275–295.
Wee, HaeKyung. 2019. Ambiguity of response particles to negative utterances in Korean and English. *Language Research* 55(3): 579–599.
Woods, Rebecca and Tom Roeper. 2021. The acquisition path of 'high' negation in English. In Danielle Dionne and Lee-Ann Vidal Covas (eds.), *Proceedings of 45th annual Boston University Conference on Language Development* (BUCLD 45), 761–775. Somerville, MA: Cascadilla Press.

Keunhyung Park
Lecturer
Department of English Education
Kyungpook National University
80 Daehak-ro, Buk-gu,
Daegu 41566, Korea
E-mail: khpark0115@gmail.com

Received: 2025. 08. 18.
Revised: 2025. 09. 20.
Accepted: 2025. 09. 25.