



## Exploring the neurophysiological underpinnings of conversational implicature processing in Korean<sup>\*</sup>

Yunju Nam<sup>a\*\*</sup> · Hyenyeong Chung<sup>b\*\*\*</sup> · Shin-ae Yoon<sup>b\*\*\*\*</sup>  
(Hanyang University<sup>a</sup> · Konkuk University<sup>b</sup>)

Nam, Yunju, Hyenyeong Chung, and Shin-ae Yoon. 2026. Exploring the neurophysiological underpinnings of conversational implicature processing in Korean. *Linguistic Research* 43(2): 593-620. Conversational implicature involves recovering speaker-intended meaning beyond an utterance's literal content. This study examined the neural time course of pragmatic inference in Korean using event-related potentials (ERPs). Across four conditions, such as Direct Answer (DA), Moderately Implicit (MI), Highly Implicit (HI), and Unrelated Answer (UA), an offline norming study confirmed a graded appropriateness pattern (DA > MI > HI > UA). In the ERP experiment (N = 28), participants read dialogue pairs via rapid serial visual presentation. Reaction times were longer for MI and HI, reflecting greater processing demands. Three ERP effects emerged: (1) an N400 (300–450 ms) at the first word in HI, reflecting an unbridged referent cost; (2) a frontal P600 (450–600 ms) at the sentence-final predicate in HI, indexing implicature derivation; and (3) in UA, an N400 at the second word and a late left-anterior negativity (600–800 ms) at sentence-final predicate, reflecting semantic/pragmatic violation at the early stage and failed pragmatic repair. MI and DA showed no reliable ERP differences. These results support a model in which pragmatic inference engages both semantic integration (N400) and higher-order meaning construction (frontal P600). (Hanyang University · Konkuk University)

**Keywords** conversational implicature, pragmatic inference, N400, P600, ERPs, Korean, indirect reply, Gricean maxims

---

\* This work was supported by the research fund of Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5A8055449).

\*\* First author

\*\*\* Co-author

\*\*\*\* Corresponding author

## 1. Introduction

The difficulty of processing implicit meaning in conversation is not merely a theoretical concern: it has practical consequences in any system that depends on natural language understanding. A striking illustration comes from human–AI speaker interactions, where AI systems routinely fail to interpret utterances that deviate from literal information requests (Nam et al. 2023). Analysis of naturally occurring human–AI dialogues reveals that violations of the Gricean Maxim of Relation account for as many as 67.4% of pragmatic communication failures, underscoring the challenge of handling relevance-based implicature computationally. Understanding the neural and cognitive mechanisms by which humans process such implicature is thus a prerequisite for developing more naturalistic conversational AI.

Human verbal communication is not limited to the delivery of literal semantic content. Speakers routinely convey meanings that diverge substantially from the propositional content of their words, and listeners regularly recover these intended meanings with apparent ease. This gap between what is said and what is meant, referred to by Grice (1975) as conversational implicature, is among the most prevalent and communicatively consequential phenomena in natural language use (Levinson 1983; Sperber and Wilson 1986). Understanding how the brain processes such non-literal meaning in real time constitutes a fundamental question in the cognitive neuroscience of language.

The successful interpretation of conversational implicature requires listeners to recognize that an utterance cannot be taken with literal meaning, to identify relevant contextual and world-knowledge constraints, and to construct an enriched interpretation consistent with the assumption of speaker cooperativity (Grice 1975; Levinson 2000). Under Grice's Cooperative Principle, speakers are expected to follow the maxims of Quantity, Quality, Relation, and Manner. When a speaker violates the maxim of Relation by providing an apparently irrelevant reply, listeners do not simply abandon the cooperative assumption; instead, they infer that the speaker intends a non-literal, contextually appropriate meaning (Levinson 1983; Geurts 2010). This process of pragmatic enrichment—sometimes called indirect reply comprehension (Holtgraves 1999)—requires resources beyond the linguistic processing system proper, including semantic memory, common ground, and social cognition (Hagoort et al. 2009; Bambini 2010).

Neurolinguistic research has identified at least two partially dissociable networks supporting pragmatic language processing. The fronto-temporal language network—encompassing inferior frontal gyrus (IFG), middle temporal gyrus, anterior temporal lobe, and angular gyrus—supports context-dependent semantic integration (Ferstl and von Cramon 2001; Kuperberg et al. 2008; Hagoort 2013). A second network associated with Theory of Mind (ToM)—comprising medial prefrontal cortex (mPFC), posterior cingulate cortex, and temporo-parietal junction (TPJ)—is engaged when listeners must attribute communicative intentions to speakers (Frith and Frith 2006; Schurz et al. 2021). Neuroimaging studies of figurative language and indirect speech acts consistently implicate both networks, with the ToM network especially prominent under high inferential demands (Bambini and Bara 2012; Van Ackeren et al. 2012; Bašnáková et al. 2014).

The electrophysiological time course of pragmatic processing has been elucidated primarily through studies of figurative language and scalar implicature. The N400—a centro-parietal negativity peaking around 400 ms—is sensitive to the ease of semantic integration in context and has been widely observed during the processing of pragmatic anomalies (Kutas and Federmeier 2011; Bambini et al. 2016). Studies of scalar implicature have reported reliable N400 effects modulated by individual pragmatic ability, suggesting that contextual enrichment proceeds incrementally from the first available word of an utterance (Nieuwland et al. 2010; Breheny et al. 2013; Politzer-Ahles and Fiorentino 2013). Specifically, Nieuwland et al. (2010) found that listeners with higher pragmatic competence showed larger N400 effects for underinformative sentences (e.g., 'Some elephants are mammals'), demonstrating that scalar inference is a skill-dependent pragmatic operation rather than an automatic semantic computation (see also Katsos and Bishop 2011). Degen and Tanenhaus (2015) further demonstrated that scalar implicature computation is sensitive to fine-grained contextual constraints, proceeding flexibly rather than following a rigid Gricean algorithm.

A positivity component in the P600 range—particularly its right-lateralized, anterior-distributed pragmatic variant—has been reported during the processing of irony (Regel et al. 2011), indirect requests (Coulson and Lovett 2010), and metaphor (De Grauwe et al. 2010), and is thought to index higher-order meaning construction required for pragmatic enrichment. Regel et al. (2011) presented participants with ironic remarks embedded in short discourse contexts and found a fronto-central

positivity between 400 and 800 ms, distinct in topography from the posterior syntactic P600. They interpreted this effect as reflecting the additional cognitive effort needed to integrate a speaker's non-literal communicative intention. Coulson and Lovett (2010) examined indirect requests (e.g., 'Can you reach the salt?') and reported a late positive component between 700 and 900 ms over anterior sites, indexing the retrieval of conventionalized meaning and integration with prior discourse context. Together, these studies establish that frontal P600-like components are reliable signatures of higher-order pragmatic inference, dissociable from purely syntactic reanalysis processes.

Indirect replies—responses that are relevant to a question without directly addressing it—provide a particularly clear test case for relevance-based implicature. Holtgraves (1999) distinguished replies that share surface-relevant content with the question (comparable to the MI condition in the present study) from those that are more globally indirect (comparable to HI). Using a priming paradigm, Holtgraves demonstrated that listeners activate the intended meaning of indirect replies as rapidly as they process direct answers, regardless of the degree of indirectness. Critically, he argued that this rapid inference involves Gricean pragmatic computation rather than lexical priming alone. The fMRI study by Bašnáková et al. (2014) extended this work to the neural level, showing that indirect replies—compared with direct replies—activated ToM areas (mPFC, TPJ) as well as language-network regions (IFG, MTG), both when participants were addressees and when they were overhearers of a conversation.

However, relatively few studies have specifically examined relevance-based conversational implicature in the ERP literature. The primary study informing the present work is an fMRI study by Jang et al. (2013), which manipulated the degree of implicitness in answers to questions (e.g., “Is Dr. Smith in his office now?”) in Korean dialogues. The authors constructed three conditions: explicit answers (e.g., “Dr. Smith is in his office now”), moderately implicit answers (which shared a lexical bridge noun with the question, e.g., “Dr. Smith's car is parked outside the building”), and highly implicit answers (which lacked such a bridge noun, e.g., “The black car is parked outside the building”).

The results revealed greater activation in the left anterior temporal lobe, angular gyrus, and posterior middle temporal gyrus for implicit relative to direct answers, consistent with increased demands on both semantic integration and pragmatic

inference. Crucially, activation in the left anterior temporal lobe was greater for highly implicit than for moderately implicit answers, suggesting a gradient of inferential load corresponding to the availability of lexical scaffolding.

However, because fMRI lacks the millisecond temporal resolution required to track the incremental unfolding of pragmatic processing on a phrase-by-phrase basis, the precise time course of these processes could not be determined. In particular, it remains unclear which phrase triggers the initial integration difficulty and when implicature derivation is completed. The present ERP study was designed to address this limitation.

The present study addresses this gap by replicating and extending the Jang et al. (2013) paradigm using ERPs. The replication component retained the core three-way manipulation of implicitness (DA, MI, HI) with Korean dialogue pairs, allowing direct comparison of neural signatures with Jang et al.'s neuroimaging findings. The extension component introduced an Unrelated Answer (UA) control condition in which the question contains the same bridge noun as the MI condition but the answer is pragmatically irrelevant. This UA condition is critical for dissociating lexical overlap effects from genuine implicature processing: if MI effects relative to DA merely reflected lexical priming (shared noun), then UA—which matches MI in lexical content but not pragmatic relevance—should show a similar pattern. The absence of a P600 in UA but its presence in HI would thus confirm that the P600 indexes true implicature derivation rather than surface-level lexical overlap.

Crucially, the present study also improved upon the Jang et al. (2013) design in a way that affords stronger control over lexico-neural confounds. Whereas Jang et al. held the question constant across conditions and varied the implicitness of the answer, the present study held the answer constant and varied the question, so that identical answer sentences were embedded in contexts of differing pragmatic relevance. This manipulation ensures that any observed neural differences between conditions cannot be attributed to differences in the lexical content or surface form of the utterances themselves, but must instead reflect the pragmatic computation triggered by the question–answer relationship.

To validate the experimental materials, an offline norming procedure with a separate group of participants was first conducted; the same materials were then used in the ERP experiment, in which participants produced both online forced-choice judgments and EEG data within a single session.

## 2. Research questions

The present study addressed the following research questions:

**RQ1. Does the manipulation of relevance-based implicitness produce a graded appropriateness gradient in an offline norming task?**

If implicitness varies along a continuum of inferential demand, appropriateness ratings will decrease as a function of deviation from direct answering (DA > MI > HI), with UA ratings receiving the lowest score due to their pragmatic irrelevance.

**RQ2. Do online acceptability judgment latencies reflect the processing costs of pragmatic inference?**

If implicature derivation incurs measurable online cost, RTs should be prolonged for conditions requiring richer inferential work (MI, HI) relative to the baseline (DA). The UA condition provides a critical contrast: unrelated answers can be flagged as inappropriate more rapidly than pragmatically challenging ones because they don't need an additional reasoning process.

**RQ3. Does the HI condition elicit an early N400 effect at the first content word of the answer, reflecting referential integration cost?**

Drawing on evidence that N400 amplitude is sensitive to the availability of a pre-activated referent (Kutas and Federmeier 2011), greater N400 amplitude is expected for HI than for other conditions at Phrase 1, where only HI excludes a lexical bridge word that could pre-activate the target concept.

**RQ4. Does the HI condition produce a sentence-final P600 effect indexing implicature completion, and if so, does it exhibit a frontal rather than posterior topographic distribution?**

Based on prior ERP studies of indirect speech acts and irony (Coulson and Lovett 2010; Regel et al. 2011), a late positivity is anticipated for HI relative to DA at the sentence-final predicate (Phrase 4). Whether this positivity carries a distinctively frontal topography — distinguishing pragmatic reanalysis from syntactic repair — cannot be settled by prior literature alone, as frontal and posterior positivities are not always dissociable at the level of mean amplitude analyses, and the precise scalp

distribution must be established empirically.

**RQ5. Does the UA condition produce a distinct ERP profile reflecting pragmatic failure?**

An N400 at Phrase 2 is expected for UA, where the semantic mismatch between question and answer first becomes apparent. However, the neural signature of outright pragmatic failure — as distinct from pragmatic challenge — remains underexplored in the literature. One of the main purposes of this research is to explore the neuro-signal reflecting the pragmatic failure and which is distinguished from the signal for pragmatic inference.

### **3. Methods**

#### **3.1. Participants**

Thirty-five Korean native speakers were initially recruited for the ERP experiment. Data from seven were excluded due to low accuracy ( $n = 1$ ), incomplete Autism Questionnaire ( $n = 1$ ), or missing behavioural responses or high EEG artifact rates ( $n = 5$ ), yielding a final sample of 28 participants (15 males; Mean age = 22.3 years,  $SD = 2.1$ ). All were right-handed, reported no history of neurological or psychiatric illness, and had not resided abroad for more than six months. Prior to EEG recording, participants completed the Korean Autism-Spectrum Quotient (AQ; Baron-Cohen et al. 2001; Korean translation: Ko et al. 2018) to screen for atypical pragmatic processing (exclusion criterion:  $AQ > 32$ ; Nieuwland et al. 2010). No participant included in the final ERP analysis exceeded this threshold ( $M = 20.1$ ,  $SD = 4.7$ ). All participants provided written informed consent in accordance with the Declaration of Helsinki. For the norming procedure, a separate group of 20 Korean native speakers (11 females; Mean age = 28.65 years,  $SD = 4.83$ ) participated. All had been raised in Korea, identified Korean as their primary language, and reported no history of neurological or psychiatric illness. None participated in the ERP experiment.

### 3.2. Materials

160 Korean dialogue pairs were constructed for the present study and used in both the norming procedure and the ERP experiment. Each dialogue consisted of an interrogative utterance by Speaker A followed by a four-phrase declarative answer by Speaker B. The central design feature was that the same target utterance by Speaker B was held constant across four conditions while only the prior question varied, thereby isolating the pragmatic effect of context on the processing of an identical answer.

Four conditions of 40 items each were constructed. The target answer utterance (Speaker B) in all conditions was a four-phrase declarative sentence. For example, the target utterance 교수님 자동차 밖에 있거든요 ('The professor's car is outside') was segmented into four phrases: Phrase 1: 교수님 (the professor), Phrase 2: 자동차 (car), Phrase 3: 밖에 (outside/locative), and Phrase 4: 있거든요 (is/sentence-final predicate). It should be noted that the question utterance (Speaker A) varies across conditions in both content and phrase length and is presented all at once prior to the RSVP phase; only Speaker B's response is segmented for ERP analysis.

In the Direct Answer (DA) condition, Speaker A's question referred directly to the content of the answer—e.g., 교수님 자동차 어디에 있어요? [Where is the professor's car?]  
—so that Speaker B's reply was immediately interpretable. In the Moderately Implicit (MI) condition, the question shared a lexical bridge noun with the answer (e.g., 교수님 [professor]) but was answered only indirectly (e.g., 오늘 교수님 나오셨어요? [Did the professor come today?]), requiring a bridging inference. In the Highly Implicit (HI) condition, no bridge noun was shared between question and answer (e.g., 오늘 강연자 도착했나요? [Did the lecturer arrive today?]), maximizing inferential load: listeners must additionally identify 강연자 (lecturer) with 교수님 (professor) before deriving the car-presence implicature. In the Unrelated Answer (UA) condition, the question contained the bridge noun (e.g., 지도교수님 전공이 뭐예요? [What is your supervisor's major?]) but was pragmatically unrelated to the answer, creating a condition that matches MI in lexical overlap while eliminating pragmatic relevance, thereby serving as a critical control for lexical priming effects.

The 160 target dialogue pairs were supplemented with 144 filler dialogues to reduce task transparency. Of the 304 total dialogues, the acceptability judgment required a

'Yes' (appropriate) response for 228 (75%) and a 'No' (inappropriate) response for 76 (25%), reflecting the structure of the target conditions. However, the 144 filler items were constructed such that approximately half required an 'inappropriate' judgment, reducing the overall correct-'appropriate' response rate across the full experiment to approximately 55% and preventing systematic response bias. Specifically, fillers included 72 dialogues with clearly inappropriate answers, ensuring that participants could not adopt a default 'appropriate' response strategy.

All items were distributed across four counterbalanced lists via Latin square design, such that each list contained 40 target items per condition (one from each condition group of 40 items), and no target item appeared in more than one condition within a single list. To maximize the number of artifact-free ERP trials per condition—a critical requirement for reliable averaging (De Zubicaray and Schiller 2019)—all participants completed all four lists in sequence, with list order counterbalanced across participants. This quasi-experimental design (Cook and Campbell 1979) ensures that each participant encounters each target item exactly four times (once per condition), yielding 160 target trials per condition per participant, substantially exceeding the recommended minimum of 30 artifact-free trials per cell. Crucially, within each list, each item appears in only one condition, so that the condition assignment is never repeated for the same item within a single presentation. Across lists, repeated exposure to items in different conditions enables fine-grained within-item, within-subject condition comparisons while controlling for item-level lexical and discourse features. Representative examples from each condition are shown in Table 1.

Table 1. Examples of the four experimental conditions<sup>1</sup>

Condition	Question (Speaker A)	Target Answer (Speaker B)
Direct Answer (DA)	교수-님 자동차 어디-에 있-어요? professor-HON car where-LOC be-POL.PRES (Where is the professor's car?)	교수-님 / 자동차 / 밖-에 / 있-던데요. professor-HON / car / outside-LOC / be-RETRO.EVID.POL (The professor's car is outside.)
Moderately Implicit (MI)	오늘 교수-님 나오-셨-어요? today professor-HON come.out-HON-PST-POL.Q (Did the professor come today?)	
Highly Implicit (HI)	오늘 강연-자 도착-했-나요? today lecture-N.SUFF arrive-do-PST-Q.POL (Did the lecturer arrive today?)	

<sup>1</sup> Romanization follows the Yale system.

Unrelated Answer (UA)	지도-교수-님 전공-이 뭐-에-요? supervise-professor-HON major-NOM what-COP-POL.PRS (What is your supervisor's major?)	
-----------------------------	--	--

Note. All four conditions use an identical target utterance by Speaker B. Korean particles are attached to the preceding noun with a hyphen per Leipzig convention. Zero-marked arguments (e.g., 자동차 without an overt particle) are left without a gloss line suffix.

### 3.3. Material norming

Prior to the ERP experiment, an offline norming procedure was conducted with the separate group of 20 participants described in Section 3.1 to verify that the 160 items instantiated the intended pragmatic gradient. This validation step was treated as an integral part of material preparation rather than a standalone study. Participants completed a paper-and-pencil appropriateness judgment task in which each of the 160 dialogue pairs was presented on a separate sheet. For each item, participants rated the appropriateness of Speaker B's answer on a seven-point Likert scale (1 = 매우 부적절 [highly inappropriate]; 7 = 매우 적절 [highly appropriate]). Using the same four-list Latin square design described in Section 3.2, each participant was assigned to one list (five participants per list), rating 40 items total (10 per condition), with each item evaluated by exactly five participants. No time limit was imposed; the task took approximately 20–30 minutes.

Mean appropriateness ratings and standard deviations were computed per condition. A linear mixed-effects regression model (LMER) was fitted in R (R Core Team 2023) using lme4 (Bates et al. 2015), with Condition (DA as reference; MI, HI, UA as predictors) as a fixed effect and participant and item as crossed random effects. Ratings decreased monotonically with increasing implicitness: DA ( $M = 6.43$ ,  $SD = 1.02$ ) > MI ( $M = 4.69$ ,  $SD = 1.75$ ) > HI ( $M = 3.92$ ,  $SD = 1.68$ ) > UA ( $M = 1.68$ ,  $SD = 1.08$ ). The LMER revealed significant differences from DA for all three conditions (all  $ps < .001$ ; Table 2). Both MI and HI ratings fell above the scale midpoint (3.5), confirming that both indirect answer types are judged as pragmatically acceptable under Korean high-context norms, if less so than direct answers. The significant MI–HI difference ( $\beta = -0.775$ ,  $t = -5.87$ ,  $p < .001$  when MI was the reference) further confirms that the two implicit conditions differ in perceived pragmatic clarity, consistent with

the presence versus absence of a lexical bridge noun. These results confirm that the 160 items instantiate a reliable, graded pragmatic continuum suitable for use in the ERP experiment.

Table 2. Descriptive statistics and LMER results for appropriateness ratings

Condition	M	SD	SE	LMER coefficient vs. DA
DA	6.43	1.02	0.228	— (Reference)
MI	4.69	1.75	0.391	$\beta = -1.74, t = -13.14, p < .001$
HI	3.92	1.68	0.376	$\beta = -2.51, t = -19.00, p < .001$
UA	1.68	1.08	0.241	$\beta = -4.75, t = -35.97, p < .001$

Note.  $\beta$  coefficients are from LMER with DA as reference. M = Mean, SD = Standard Deviation. SE = Standard Error. All three contrasts are significant at  $p < .001$ .

### 3.4. EEG experiment

A within-subjects design was employed with Condition (DA, MI, HI, UA) as the main factor. The quasi-experimental design described in Section 3.2 was implemented across five sessions including one practice session; the order of lists was counterbalanced across participants. Within each list, items were pseudo-randomized with the constraint that no more than three consecutive trials came from the same condition. This yielded 160 target trials per condition per participant.

EEG was recorded continuously using a BrainAmp DC amplifier (Brain Products GmbH, Germany) with a 32-channel active antiCAP electrode cap positioned according to the International 10-10 system. The online reference was the tip of the nose; data were re-referenced offline to the average of the left and right mastoid electrodes (TP9, TP10). Vertical and horizontal eye movements were monitored via electrodes placed below both eyes and lateral to both eyes, respectively. Signals were digitised at 250 Hz with electrode impedances maintained below 5 k $\Omega$ . A band-pass filter of 0.1–30 Hz and a 60 Hz notch filter were applied online. Note that the draft version of this manuscript described a band-pass filter of 0.5–30 Hz; the correct value is 0.1 Hz for the high-pass cutoff, consistent with standard ERP methodology for the latency ranges examined here.

Stimuli were presented via E-Prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA). Each trial began with a 500 ms fixation cross and a 500 ms blank

screen. Speaker A's question was then displayed in black text at screen center; participants pressed the space bar when they had read and understood it (maximum display: 5000 ms). A further fixation cross and blank interval (500 ms each) preceded the target utterance. Speaker B's answer was presented phrase-by-phrase in blue text via Rapid Serial Visual Presentation (RSVP), visually distinguishing it from the prior utterance and enabling phrase-locked ERP recording. Each target utterance was divided into four phrases—e.g., 교수님 (Phrase 1) / 자동차 (Phrase 2) / 밖에 (Phrase 3) / 있던데요 (Phrase 4, sentence-final predicate)—each displayed for 500 ms followed by a 500 ms blank screen (SOA = 1000 ms). This SOA provided a sufficient recording window to capture both N400 and later pragmatic components (Coulson and Lovett 2010; Gisladdottir et al. 2015). Immediately after the final phrase, the judgment prompt appeared on screen. Participants responded by pressing F (appropriate) or J (inappropriate) within 5000 ms.

### 3.5. Data analysis

Accuracy and RT from the acceptability judgment task were analyzed in R (R Core Team 2023). Accuracy was near ceiling (> 99% across all conditions) and was not analyzed further. For RT, only correct trials were included; outliers (> 2.5 SDs from the participant's per-condition mean) were replaced by the condition mean (Van Selst and Jolicoeur 1994). RTs were submitted to LMER (lme4; Bates et al. 2015) with Condition (DA reference; MI, HI, UA as predictors) as a fixed effect and participant and item as crossed random effects. To obtain *p*-values for the fixed effects, we used the lmerTest package (Kuznetsova et al. 2017), which applies the Satterthwaite's approximation for degrees of freedom.

EEG data from correct trials were epoched from −100 ms to 900 ms relative to the onset of each phrase, with baseline correction over the 100 ms pre-stimulus interval. Artefact rejection excluded epochs with eye movements (vertical > ±80 μV; horizontal > ±50 μV) or other deflections exceeding ±100 μV. Grand-averaged ERPs were computed per condition and phrase position. Two time windows were defined a priori based on the existing literature and the visual inspection on the grand-averaged brain waveforms: 300–450 ms (N400; Van Berkum 2009; Regel et al. 2011) and 450–600 ms (P600; Regel et al. 2011; Hunt et al. 2013; Bambini et al. 2016). The 450–600

ms window was selected as the sole P600 analysis window, motivated a priori by the established temporal range of the pragmatic P600 in prior studies. A late window of 600–800 ms was examined following visual inspection of the UA waveforms and is treated as exploratory.

To address whether the observed frontal positivity for HI constitutes a true pragmatic P600 rather than an extension of the posterior syntactic P600, we conducted an additional topographic analysis. Mean amplitude in anterior ROIs (LA, MA, RA) was compared with that in posterior ROIs (LP, MP, RP) for the DA vs. HI contrast in the P600 window. Planned paired comparisons at individual ROIs were conducted to examine the topographic distribution of the HI effect.

Twenty-two electrodes were aggregated into seven regions of interest (ROIs) defined by the anteroposterior axis and laterality: Left Anterior (LA: F3, FC5), Right Anterior (RA: F4, FC6), Left Posterior (LP: CP5, P3), Right Posterior (RP: CP6, P4), Midline Anterior (MA: Fz, FC1, FC2), Midline Central (MC: Cz, C3, C4), Midline Posterior (MP: CP1, CP2, Pz). Mean amplitude in each ROI and time window was submitted to repeated-measures ANOVA with Condition and ROI as within-subjects factors. Analyses were conducted as planned pairwise comparisons (DA vs. HI; DA vs. MI; DA vs. UA) rather than omnibus ANOVAs, reflecting focused a priori predictions. For midline analysis, a Sentence Type (2)  $\times$  ROI (3) ANOVA was used; for lateral analysis, a Sentence Type (2)  $\times$  Laterality (2)  $\times$  Posteriority (2) ANOVA was used. Significant interaction effects were followed up with pairwise *t*-tests. Greenhouse–Geisser corrections were applied where sphericity was violated.

## 4. Results

### 4.1. Acceptability judgment results

Accuracy exceeded 99% in all conditions (DA, MI, HI were judged as appropriate and UA as inappropriate). Mean RTs were: DA = 380.67 ms, MI = 617.44 ms, HI = 701.25 ms, UA = 399.47 ms (Table 3). The LMER revealed significantly longer RTs for MI ( $\beta = 236.77$ , SE = 24.00,  $t = 6.665$ ,  $p < .001$ ) and HI ( $\beta = 320.58$ , SE = 23.99,  $t = 10.446$ ,  $p < .001$ ) relative to DA. UA did not differ significantly from DA ( $\beta = 18.80$ , SE = 24.00,  $t = 0.712$ ,  $p = .438$ ). Participants thus required substantially

more time to evaluate implicit than direct or unrelated answers, reflecting the additional processing demands of pragmatic inference in MI and HI.

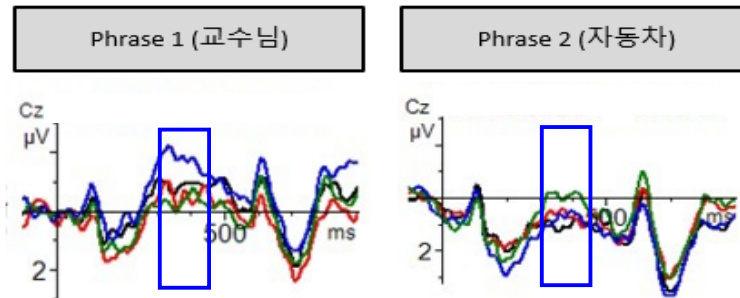
Table 3. LMER Results for acceptability judgment response times

Condition	Mean RT (ms)	Estimate	SE (ms)	<i>t</i>	<i>p</i>
DA (Reference)	380.67	380.67	29.93	7.858	—
MI	617.44	236.77	24.00	6.665	< .001 ***
HI	701.25	320.58	23.99	10.446	< .001 ***
UA	399.47	18.80	24.00	0.712	.438

Note.  $\beta$  coefficients are from LMER with DA as reference. SE = Standard Error.  
\*\*\*  $p < .001$ .

#### 4.2. ERP results<sup>2</sup>

The ERP results at critical electrodes are depicted in Figure 1.



<sup>2</sup> All F-statistics below are from planned pairwise repeated-measures ANOVAs (Condition [2 levels: DA vs. the condition of interest]  $\times$  ROI). Degrees of freedom for the Sentence Type factor are therefore (1, 27) throughout, as each comparison involves two conditions. This approach was adopted in line with a priori predictions detailed in Section 2, ensuring appropriate control of Type I error across planned contrasts.

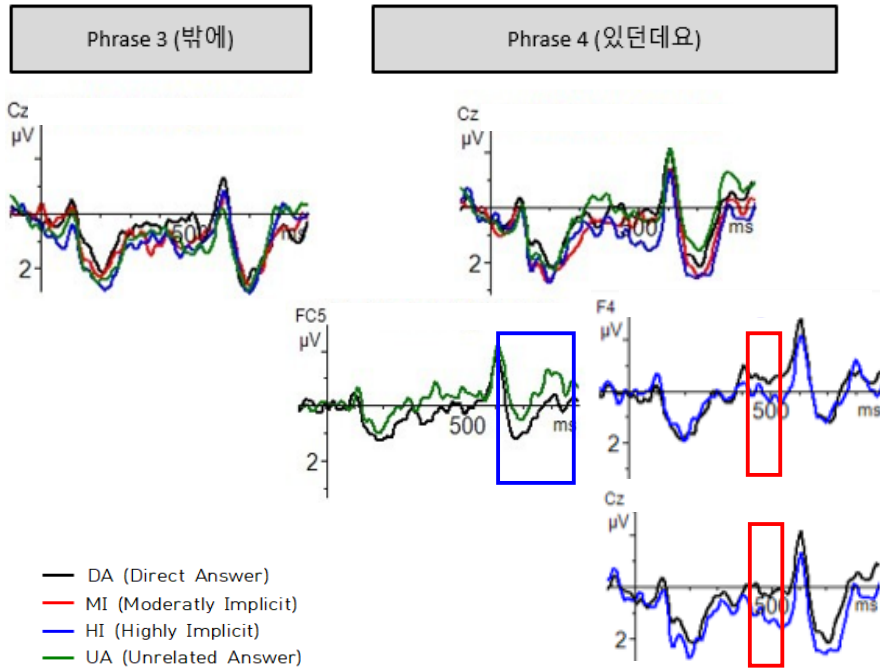


Figure 1. ERP results in each phrases

#### 4.2.1. Phrase 1 (first content word)

In the N400 window (300–450 ms) for the DA vs. HI comparison, a significant main effect of Sentence Type was observed over midline ROIs ( $F(1, 27) = 3.995, p < .05$ ), with a significant main effect of ROI ( $F(2, 54) = 4.495, p < .05$ ) but no significant Sentence Type  $\times$  ROI interaction ( $F(2, 54) = 1.251, p = .284$ ). HI elicited greater negativity than DA in the midline region. No significant condition differences were found in the lateral analysis, and no effects were observed in the P600 window at Phrase 1. No N400 differences were observed between DA and MI or between DA and UA at Phrase 1, confirming that the effect is specific to the absence of a lexical bridge noun in the HI condition.

#### 4.2.2. Phrase 2 (second content word)

In the N400 window (300–450 ms), UA condition elicited a significant main effect

compared to DA in the midline analysis ( $F(1,27) = 4.870, p < .05$ ), with a significant main effect of ROI ( $F(2,54) = 7.158, p < .01$ ) but no Sentence Type  $\times$  ROI interaction ( $F(2,54) = 0.732, p = .484$ ). In the lateral analysis, no significant Sentence Type main effect was found ( $F(1,27) = 0.648$ ), though a significant main effect of Anteriority emerged ( $F(1,27) = 4.887, p < .05$ ). Together, these results indicate an N400 effect for UA relative to DA over midline sites in the 300–450 ms window. No significant differences were found among DA, MI, and HI at this position.

#### 4.2.3. Phrase 3 (locative postposition or adverbs)

No significant condition effects were found in any time window at Phrase 3. This null result is expected. Phrase 3 (밖<sub>에</sub>), the locative postposition 'outside' is a closed-class functional element providing primarily spatial information that does not differentially influence pragmatic inference across conditions.

#### 4.2.4. Phrase 4 (sentence-final predicate)

##### 1) Frontal P600 in HI (DA vs. HI comparison)

In the P600 window (450–600 ms) for the DA vs. HI comparison, the midline analysis yielded a significant main effect of Sentence Type ( $F(1, 27) = 6.211, p < .05$ ) and a significant main effect of ROI ( $F(2, 54) = 5.906, p < .05$ ). The interaction between Sentence Type and ROI was not significant ( $F(2, 54) = 1.040, n.s.$ ). In the lateral analysis, no significant effects were obtained. No significant P600 difference was found in MI compared to DA as well.

To confirm the frontal distribution of the P600 in HI and distinguish it from the posterior syntactic P600, we conducted a topographic analysis comparing anterior (LA, MA, RA) with posterior (LP, MP, RP) ROIs in the 450–600 ms window (Figure 2). The Anteriority  $\times$  Condition interaction using repeated measures ANOVA was not significant ( $F(1, 27) = 0.74, p = .399$ ). However, individual ROI analyses revealed that the HI positivity relative to DA was significant at anterior sites MA ( $t(27) = 2.50, p = .019$ ) and RA ( $t(27) = 2.30, p = .029$ ), while absent at all posterior sites (all  $ps > .06$ ) as summarized in Table 4. This topographic asymmetry confirms that the HI effect reflects an anterior pragmatic P600 rather than an extension of the

posterior syntactic P600, consistent with prior reports of the pragmatic positivity in irony and indirect request processing (Coulson and Lovett 2010; Regel et al. 2011).

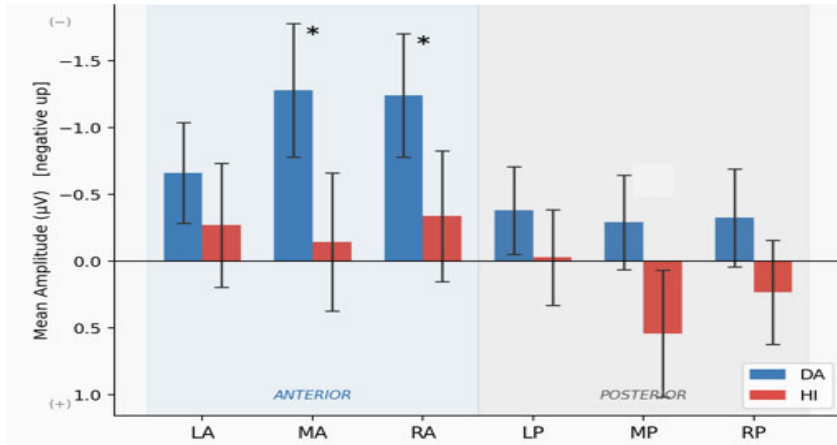


Figure 2. Topographic Analysis for DA and HI comparison - Anterior (LA, MA, RA) vs. Posterior (LP, MP, RP)

Table 4. Paired t-test for DA and HI comparison per ROI

ROI	DA (M ± SE)	HI (M ± SE)	t(27)	p
LA	-0.662 ± 0.377	-0.270 ± 0.463	-0.873	0.390
MA	-1.279 ± 0.500	-0.144 ± 0.518	-2.498	0.019*
RA	-1.243 ± 0.462	-0.338 ± 0.490	-2.300	0.029*
LP	-0.380 ± 0.327	-0.029 ± 0.358	-1.055	0.301
MP	-0.291 ± 0.354	0.542 ± 0.474	-1.900	0.068
RP	-0.325 ± 0.366	0.231 ± 0.390	-1.380	0.179

## 2) Late left-anterior negativity in UA (DA vs UA comparison)

In the late window (600–800 ms; exploratory), the lateral analysis revealed a significant main effect of Posteriority ( $F(1, 27) = 4.880, p < .05$ ) in UA compared to DA, and a significant three-way Sentence Type × Laterality × Posteriority interaction ( $F(1, 27) = 5.630, p < .05$ ). Follow-up paired t-test confirmed a significant negativity at Left Anterior (LA;  $t(27) = 3.218, p < .01$ ). This late left-anterior negativity was absent in MI and HI. An earlier DA vs. UA comparison in the 300–450 ms window also yielded a marginal three-way interaction ( $F(1, 27) = 3.170, p = .086$ ) and a significant

negativity at LA ( $t(27) = 2.458, p < .05$ ), though this earlier effect did not reach omnibus significance.

## 5. Discussion

The present study characterized the neural time course of relevance-based conversational implicature processing in Korean. The norming data confirmed a graded appropriateness gradient (DA > MI > HI > UA), validating the experimental materials. In the ERP experiment, acceptability judgment latencies and ERP components revealed complementary aspects of online pragmatic processing: behavioral data confirmed that implicit answers incur additional processing costs, while ERP data showed the specific neural computations at each phrase position. The following subsections address each of the five research questions in turn.

### 5.1. Graded appropriateness gradient in offline norming (RQ1)

RQ1 asked whether the manipulation of relevance-based implicitness would produce a graded appropriateness gradient in an offline norming task. Mean appropriateness ratings decreased monotonically with increasing inferential demand (DA = 6.43 > MI = 4.69 > HI = 3.92 > UA = 1.68), and all pairwise differences relative to DA reached significance (all  $ps < .001$ ). Crucially, both MI and HI ratings fell above the scale midpoint of 3.5, confirming that both indirect answer types are judged as pragmatically acceptable under Korean high-context norms, albeit less so than direct answers. The significant MI–HI difference further confirmed that the presence versus absence of a lexical bridge noun is perceived as a meaningful distinction in pragmatic clarity by native Korean listeners. These results validate the experimental materials as instantiating a reliable, graded pragmatic continuum, and provide the essential normative foundation for interpreting the online ERP and behavioral data reported below.

### 5.2. Acceptability judgment latencies: Cost of pragmatic inference (RQ2)

RQ2 asked whether online acceptability judgment latencies would reflect the processing

costs of pragmatic inference. The RT pattern maps directly onto the implicitness hierarchy established in the norming procedure. Significantly longer RTs for MI and HI than for DA confirm that listeners sustain additional processing costs when deriving pragmatic meaning from indirect answers—a finding that replicates and extends Jang et al. (2013) to a phrase-final evaluation paradigm. The graded RT pattern ( $DA \approx UA < MI < HI$ ) is consistent with Holtgraves' (1999) behavioral evidence that indirect replies impose measurable processing costs proportional to their degree of indirectness. Importantly, although Holtgraves found that indirect reply meanings were activated as rapidly as direct answers in a priming paradigm, the present judgment task requires more explicit evaluation of appropriateness, which may amplify observable RT differences.

The near-identical RTs for UA and DA are particularly informative with respect to RQ2. Although the ERP data show that UA triggers online semantic expectation violation at Phrase 2, the binary appropriateness decision was made as rapidly as for DA, consistent with listeners quickly flagging the answer as irrelevant without investing in extended pragmatic inference (cf. Noveck and Reboul 2008). This dissociation between online ERP processing and post-utterance judgment latency demonstrates the value of collecting both measures within the same paradigm.

### **5.3. N400 at phrase 1 in HI: Lexical accessibility and the role of bridge words (RQ3)**

RQ3 asked whether the HI condition would elicit an early N400 effect at the first content word of the answer, reflecting referential integration cost in the absence of a lexical bridge noun. This prediction was confirmed. The enhanced N400 for HI at Phrase 1 reflects reduced lexical accessibility: in DA, MI, and UA, the first noun (e.g., 교수님) appeared in the prior question and is therefore a given referent whose processing is facilitated by contextual priming (Gernsbacher 1990; Burkhardt 2006). In HI, the question referred to the same individual using a different noun (e.g., 강연자 [lecturer]), making the first noun of the answer (교수님 [professor]) a new referent requiring an additional referential linking step. This increased integration cost is indexed by the larger N400, consistent with its established sensitivity to referential givenness in discourse (Nieuwland and Van Berkum 2006; Kutas and Federmeier

2011).

The early onset of this effect further supports evidence that pragmatic discourse processing begins incrementally from the first content word of the answer (Egorova et al. 2014; Gisladdottir et al. 2015), well before the sentence-final predicate at which implicature derivation is completed. Importantly, and as predicted in RQ3, neither MI nor UA elicited an N400 at Phrase 1, confirming that the effect is specifically attributable to the absence of a shared lexical bridge word in HI, rather than to pragmatic implicitness per se.

#### **5.4. N400 at phrase 2 in UA: Semantic expectation violation (RQ5-1)**

RQ5 asked whether the UA condition would produce a distinct ERP pattern reflecting pragmatic failure. The N400 observed at Phrase 2 (자동차 [car]) in UA constitutes the first component of this distinct profile. In the UA condition, the question (e.g., 지도교수님 전공이 뭐예요? [What is your supervisor's major?]) creates no contextual basis for a response about a car, generating a semantic integration cost when that word is encountered and eliciting the classic N400 signature of contextual mismatch (Kutas and Hillyard 1980; Kutas and Federmeier 2011). This UA N400 at Phrase 2 was not followed by a P600, indicating that listeners did not subsequently reinterpret the answer as pragmatically meaningful. This pattern distinguishes UA from HI: in HI, initial referential integration difficulty at Phrase 1 is eventually resolved via implicature derivation (indexed by the P600 at Phrase 4), whereas in UA, semantic mismatch at Phrase 2 cannot be resolved because the answer is genuinely irrelevant. The contrast between HI (N400 at P1 → P600 at P4) and UA (N400 at P2 → no P600) thus provides direct evidence that the HI P600 reflects true pragmatic enrichment rather than a generic response to processing difficulty.

#### **5.5. Frontal P600 at phrase 4 in HI: Implicature derivation (RQ4)**

RQ4 asked whether the HI condition would produce a sentence-final P600 effect indexing implicature completion and, if so, whether it would exhibit a distinctively frontal rather than posterior topographic distribution. The frontal P600 for HI at the sentence-final predicate (maximal at midline anterior and right anterior sites and

peaking 450–600 ms post-stimulus onset) is topographically distinct from the posteriorly distributed syntactic P600 (Osterhout and Holcomb 1992; Hagoort et al. 1993) and aligns with the pragmatic P600 reported in studies of irony (Regel et al. 2011), indirect requests (Coulson and Lovett 2010), and metaphor (De Grauwe et al. 2010). The topographic analysis comparing anterior and posterior ROIs ruled out the possibility that the HI positivity reflects posterior syntactic reanalysis rather than frontal pragmatic enrichment.

We interpret this frontal P600 as reflecting the cognitive effort required to complete implicature derivation at sentence-final position: having received the full answer, the listener integrates it with prior context and world knowledge (professor's car is outside → professor must be present), completing a multi-step inference that could not be fully resolved at earlier phrase positions. The timing of this effect (450 ms) supports architectures in which pragmatic enrichment unfolds incrementally and in parallel with semantic processing, rather than strictly after semantic processing is complete (Levinson 2000; Bezuidenhout 2017).

The absence of a P600 in MI, despite MI also requiring pragmatic inference, is theoretically informative with respect to RQ4. The lexical bridge noun shared by the MI question and answer (e.g., 교수님) allows the implicature to be resolved incrementally as the answer unfolds, without accumulating sufficient processing load at any individual phrase to yield a detectable neural deflection at sentence-final position. This result is consistent with Holtgraves' (1999) finding that surface-relevant indirect replies are processed as efficiently as direct replies, and with Burkhardt's (2006) account of how given information facilitates referential integration. From an ERP perspective, MI is neurophysiologically equivalent to DA when a lexical bridge is present, a finding that underscores the central role of lexical scaffolding in pragmatic inference.

The frontal topography implicates prefrontal cortex involvement, consistent with the fMRI evidence of Jang et al. (2013) and with the broader literature on mPFC and TPJ involvement in ToM-based pragmatic inference (Van Ackeren et al. 2012; Bašnáková et al. 2014; Schurz et al. 2021). The convergence between the present ERP finding of a frontal P600 and Jang et al.'s neuroimaging finding of anterior temporal lobe activation in the HI condition supports a unified account in which both the temporal dynamics (ERP) and spatial localization (fMRI) of implicature derivation reflect engagement of fronto-temporal networks associated with discourse integration

and mentalizing.

### **5.6. Late negativity at phrase 4 in UA: Failed pragmatic repair (RQ5-2)**

Returning to RQ5, the late left-anterior negativity (600–800 ms) for UA at sentence-final predicate (absent in MI and HI) completes the ERP profile of pragmatic failure, distinguishing it not only from successful implicature (HI) but also from pragmatic challenge without failure (MI). This late negativity suggests an attempted but ultimately unsuccessful pragmatic repair: having encountered a semantically unexpected element at Phrase 2, listeners may engage pragmatic inference mechanisms (analogous to those that succeed in HI) in an attempt to salvage a coherent interpretation, but fail because the answer is genuinely irrelevant. Late left-anterior negativities have been associated with working-memory demands during discourse integration and with difficulty incorporating incoherent information (Kaan and Swaab 2003; Van Berkum et al. 2007).

In direct answer to RQ5, the combination of a Phrase 2 N400 (expectation violation), the absence of a P600 (no successful reanalysis), and the late negativity at Phrase 4 (failed repair) jointly distinguishes UA from all other conditions as a case of outright pragmatic failure rather than pragmatic challenge. This three-component pattern is consistent with Holtgraves' (1999) characterization of unrelated replies as prompting inference attempts that ultimately fail due to the absence of any viable relevance relation, and extends that behavioral account with a neural time course of the failure process itself.

### **5.7. Implications for models of pragmatic processing**

Taken together, the present results support a processing architecture in which pragmatic inference operates incrementally and engages dissociable neural mechanisms depending on the nature and degree of inferential demand. The N400 indexes the cost of referential integration at the word level—specifically, the mismatch between a predicted referent (primed by the question) and the encountered word—while the frontal P600 indexes the completion of higher-order implicature derivation at the discourse level. This dissociation parallels findings from the figurative language

literature (De Grauwe et al. 2010; Bambini et al. 2016) and extends them to relevance-based conversational implicature.

The finding that MI yields no ERP differentiation from DA despite requiring pragmatic inference has implications for computational and psycholinguistic models of implicature. Models that treat implicature as a costly post-semantic operation (e.g., Grice 1975; Levinson 2000) would predict processing costs for MI relative to DA; the absence of such costs—in both ERP and RT data—suggests instead that lexical scaffolding enables rapid, incremental resolution of relevance-based implicature without accumulating measurable neural cost. This is consistent with constraint-based models that emphasize the role of contextual support in modulating inference difficulty (Breheny et al. 2013; Degen and Tanenhaus 2015) and with the prediction-based account of N400 effects, in which prediction errors—rather than inference per se—drive electrophysiological responses.

The present study also highlights the importance of appropriate control conditions in ERP research on pragmatic inference. The UA condition, which has not been included in prior Korean studies of conversational implicature, proved essential for establishing that the HI P600 reflects genuine implicature computation rather than a surface-level response to shared lexical content with the question. Future studies should adopt similarly fine-grained baseline conditions to disentangle lexical and pragmatic contributions to ERP effects.

## 6. Conclusion

The present study provides the first electrophysiological characterization of relevance-based conversational implicature in Korean, combining a materials norming procedure and an ERP experiment in which both acceptability judgment latencies and EEG were collected from the same participants within a single session. The principal findings were: (1) longer acceptability judgment RTs for MI and HI than for DA and UA, confirming additional processing costs for pragmatic inference; (2) an N400 at Phrase 1 in HI, reflecting the cost of an unbridged referent; (3) an N400 at Phrase 2 in UA, indexing semantic expectation violation; (4) a frontal P600 at Phrase 4 in HI, indexing implicature derivation; and (5) a late left-anterior negativity at Phrase 4 in UA, reflecting failed pragmatic repair. The absence of ERP effects in

MI relative to DA confirms the facilitating role of lexical bridge words in scaffolding implicature derivation.

These results support a model in which conversational implicature processing engages both semantic integration mechanisms (N400) and higher-order meaning construction mechanisms (frontal P600), with distinct temporal profiles depending on the availability of lexical scaffolding and the degree of inferential demand. The topographic dissociation between the frontal P600 in HI and the posterior syntactic P600 provides electrophysiological evidence that relevance-based implicature derivation recruits prefrontal resources associated with discourse integration and speaker intention inference, converging with neuroimaging evidence from Jang et al. (2013) and Bašnáková et al. (2014). Future research should examine whether these neural signatures generalise across languages and pragmatic contexts, investigate the contribution of individual differences in pragmatic ability and working memory, and explore whether the lexical bridge advantage observed here extends to spoken dialogue processing.

## References

- Bambini, Valentina. 2010. Neuropragmatics: A foreword. *Italian Journal of Linguistics* 22(1): 1-20.
- Bambini, Valentina and Bruno G. Bara. 2012. Neuropragmatics. In Jan-Ola Östman and Jef Verschueren (eds.), *Handbook of pragmatics*, 1-21. Amsterdam: John Benjamins Publishing Company.
- Bambini, Valentina, Chiara Bertini, Walter Schaeken, Alessandra Stella, and Francesco Di Russo. 2016. Disentangling metaphor from context: An ERP study. *Frontiers in Psychology* 7: 559.
- Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley. 2001. The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31(1): 5-17.
- Bašnáková, Jana, Kirsten Weber, Karl Magnus Petersson, Jos van Berkum, and Peter Hagoort. 2014. Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex* 24(10): 2572-2578.
- Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.
- Bezuidenhout, Anne. 2017. Contextualism and semantic minimalism. In Jonathan Jenkins

- Ichikawa (ed.), *The Routledge handbook of epistemic contextualism*, 3-12. London: Routledge.
- Breheny, Richard, Heather J. Ferguson, and Napoleon Katsos. 2013. Taking the epistemic step: Toward a model of on-line pragmatic inference. *Cognition* 126(3): 423-440.
- Burkhardt, Petra. 2006. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language* 98(2): 159-168.
- Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- Coulson, Seana and Christopher Lovett. 2010. Comprehension of indirect requests: Contrasting effects of conventionality and pragmatic plausibility in event-related brain potential studies. *Brain and Language* 112(2): 93-104.
- De Grauwe, Sophie, Abigail Swain, Phillip J. Holcomb, Tali Ditman, and Gina R. Kuperberg. 2010. Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia* 48(7): 1965-1984.
- De Zubicaray, Greig and Niels O. Schiller (eds.). 2019. *The Oxford handbook of neurolinguistics*. Oxford: Oxford University Press.
- Degen, Judith and Michael K. Tanenhaus. 2015. Processing scalar implicature: A constraint-based approach. *Cognitive Science* 39(4): 667-710.
- Egorova, Natalia, Friedemann Pulvermüller, and Yuri Shtyrov. 2014. Neural dynamics of speech act comprehension: An MEG study of naming and requesting. *Brain Topography* 27(3): 375-392.
- Ferstl, Evelyn C. and D. Yves von Cramon. 2001. The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research* 11(3): 325-340.
- Frith, Christopher D. and Uta Frith. 2006. The neural basis of mentalizing. *Neuron* 50(4): 531-534.
- Gernsbacher, Morton Ann. 1990. *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Geurts, Bart. 2010. *Quantity implicatures*. Cambridge: Cambridge University Press.
- Gisladottir, Rosa S., Dorothee J. Chwilla, and Stephen C. Levinson. 2015. Conversation electrified: ERP correlates of speech act recognition in turns of talk. *PLoS ONE* 10(3): e0120068.
- Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole and Jerry Morgan (eds.), *Syntax and semantics, vol. 3: Speech acts*, 41-58. New York: Academic Press.
- Hagoort, Peter. 2013. MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology* 4: 416.
- Hagoort, Peter, Colin M. Brown, and Jolanda Groothusen. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes* 8(4): 439-483.
- Hagoort, Peter, Jos Van Berkum, Giosuè Baggio, and Pienie Zwitserlood. 2009. Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1521): 1261-1269.

- Holtgraves, Thomas. 1999. Comprehending indirect replies: When and how are their conveyed meanings activated? *Journal of Memory and Language* 41(4): 519-540.
- Hunt, Lyndsey, Stephen Politzer-Ahles, Leah Gibson, Utako Minai, and Robert Fiorentino. 2013. Pragmatic inferences modulate N400 during sentence comprehension: Evidence from picture-sentence verification. *Neuroscience Letters* 534: 246-251.
- Jang, Giyoung, Shin-ae Yoon, Soo-Eun Lee, Hyelin Park, Jieun Kim, Joong Hwi Ko, and Hae-Jeong Park. 2013. Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *NeuroImage* 81: 61-72.
- Kaan, Edith and Tamara Y. Swaab. 2003. Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience* 15(1): 98-110.
- Katsos, Napoleon and Dorothy V. M. Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120(1): 67-81.
- Ko, Ha Young, Won-Hee Lee, Eun-Kyung Won, Ji Jeong Ban, Da Eun Jung, and Yeni Kim. 2018. The reliability and validity of the Korean version of the Autism-Spectrum Quotient. *Psychiatry Investigation* 15(8): 783-789.
- Kuperberg, Gina R., Tatiana Sitnikova, and Balaji M. Lakshmanan. 2008. Neuroanatomical distinctions within the semantic system during sentence comprehension: Evidence from functional magnetic resonance imaging. *NeuroImage* 40(1): 367-388.
- Kutas, Marta and Kara D. Federmeier. 2011. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology* 62: 621-647.
- Kutas, Marta and Steven A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207(4427): 203-205.
- 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.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Nam, Yunju, Hyenyeong Chung, and Upyong Hong. 2023. Language artificial intelligences' communicative performance quantified through the Gricean conversation theory. *Cyberpsychology, Behavior, and Social Networking* 26(12): 919-923.
- Nieuwland, Mante S. and Jos J. A. Van Berkum. 2006. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18(7): 1098-1111.
- Nieuwland, Mante S., Tali Ditman, and Gina R. Kuperberg. 2010. On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language* 63(3): 324-346.
- Noveck, Ira A. and Anne Reboul. 2008. Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences* 12(11): 425-431.
- Osterhout, Lee and Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic

- anomaly. *Journal of Memory and Language* 31(6): 785-806.
- Politzer-Ahles, Stephen and Robert Fiorentino. 2013. The realization of scalar inferences: Context sensitivity without processing cost. *PLoS ONE* 8(5): e63943.
- R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Regel, Stefanie, Thomas C. Gunter, and Angela D. Friederici. 2011. Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience* 23(2): 277-293.
- Schurz, Matthias, Joaquim Radua, Matthias G. Tholen, Lara Maliske, Daniel S. Margulies, Rogier B. Mars, Jerome Sallet, and Philipp Kanske. 2021. Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin* 147(3): 293-327.
- Sperber, Dan and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
- Van Ackeren, Markus J., Daniel Casasanto, Harold Bekkering, Peter Hagoort, and Shirley-Ann Rueschemeyer. 2012. Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience* 24(11): 2237-2247.
- Van Berkum, Jos J. A. 2009. The neuropragmatics of 'simple' utterance comprehension: An ERP review. In Uli Sauerland and Kazuko Yatsushiro (eds.), *Semantics and pragmatics: From experiment to theory*, 276-316. Basingstoke: Palgrave Macmillan.
- Van Berkum, Jos J. A., Aart W. Koornneef, Marlies Otten, and Mante S. Nieuwland. 2007. Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research* 1146: 158-171.
- Van Selst, Mark and Pierre Jolicoeur. 1994. A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology* 47(3): 631-650.

**Yunju Nam**

Associate Professor  
Department of German Language and Literature  
Hanyang University  
222 Wangsimni-ro, Seongdong-gu,  
Seoul, 04763, Korea  
E-mail: yjnam05@hanyang.ac.kr

**Hyeyeong Chung**

Instructor  
Department of Media and Communication  
Konkuk University

620 Yunju Nam · Hyenyeong Chung · Shin-ae Yoon

120 Neungdong-ro, Gwangjin-gu,  
Seoul, 05029, Korea  
E-mail: chn4050@konkuk.ac.kr

**Shin-ae Yoon**

Instructor  
Department of Media and Communication  
Konkuk University  
120 Neungdong-ro, Gwangjin-gu,  
Seoul, 05029, Korea  
E-mail: sculy82@gmail.com

Received: 2026. 03. 14.

Revised: 2026. 05. 20.

Accepted: 2026. 05. 24.