Generative AI–Human Intuition Gaps in Korean Double Negation: A Big Data-based Empirical Study

Seulkee Park
seulkeepark@khu.ac.kr
Kyung Hee University

Double Negation in Korean

While English double negatives yield a consistent positive meaning through logical cancellation (Huddleston & Pullum 2002), their Korean counterparts exhibit context-sensitive ambiguity.

(1)a. She did**n't not** move. (= *She moved*.)
b. He did**n't** say **nothing**. (= *He said something*.) (Huddleston & Pullum 2002: 847)

(2) Pi-ka **an** o-nun kes-un **an**-i-ya.
rain-NOM not come-ADN thing-TOP not-COP-DECL
'It's not that it is not raining (= *it is raining*).' (Lee 2017: 63)

English double negation typically involves clausal-level logical negation yielding truth-conditional cancellation (¬(¬p) = p), whereas Korean double negation allows scope-dependent interpretations.

What Sparked This Study

Consider the following example, which sparked debate among Korean native speakers on social media. The screenshot below shows the results of a poll conducted on *Threads*:

(3) Yay-nun **an**-ppaci-nun moim-i **eps-**ney.
she-TOP NEG-skip-MOD gathering-NOM exist.not-DECL
'(literal) There is no meeting that s/he does not skip.'



Figure 1. Survey results for the double-negation sentence shared on *Threads* (Meta).

Research Questions

- How do native speakers interpret Korean double negation (DN)?
- Do GenAI models align with human DN interpretations?
- How do syntax and context affect DN interpretation?

Research Hypotheses

1. **Scope**: DN interpretation varies with negation scope.
2. **Acceptability**: Interpretive differences reflect perceived grammaticality.
3. **Lexical Negativity**: Verb semantics is **not** the primary determinant of DN interpretation.

Types of Negation in Korean

Type	Marker	Form	Example
Morphological (an-)	an	SFN	Mimi-ka cip-ey an ka-ss-ta. 'Mimi did not go home.'
	-ci anh-	LFN	Mimi-ka cip-ey ka-ci anh-ass-ta.
Morphological (mot-)	mot	SFN	Mimi-ka cip-ey mos ka-ss-ta. 'Mimi could not go home.'
	ci mos-	LFN	Mimi-ka cip-ey ka-ci mos-hay-ss-ta.
Prohibitive (mal-)	ci mal-	LFN only	Cikak-ha-ci mal-ca. 'Don't be late.'
Copula	ani- 'be not'	—	Haksayng-i ani-ta '(He/She) is not a student.'
Lexical	molu- 'not know'	—	Tap-ul molu-n-ta. '(He/She) does not know the answer.'
	eps- 'not exist'	—	Tap-i eps-ta. 'There is no answer.'

Table 1. Types of Negation in Korean

Logical Double Negation vs. Negative Concord

Do Korean double negation constructions compositionally yield a universal interpretation, or do they instead involve a negative-concord-like interpretation?

- Double negation reading: 'He skips every meeting.'
- Negative concord-like reading: 'He attends every meeting.'

Surface logical structure: Neg × Neg = Double negation = $\neg\exists x\,[\text{meeting}(x) \wedge \neg\text{absent}(x)]$

- ***an-ppaci-***: Neg_1 - morphological negation
- ***eps-***: Neg_2 - lexical negation

Empirical Studies and Results

Task 1: Native Speaker Acceptability Judgment

- Participants & stimuli: 100 Korean native-speaking college students (ages 20–25); 11 *an* + [negative verb] + *eps*- sentences
- Task & measure: Forced choice between DN and NC-like readings; response distribution

#	Sentence	Verb	Gloss	DN Reading	NC-like Reading
1	쟤는 안 빠지는 모임이 없네	*ppaci-*	'skip'	15% (n=15)	85% (n=85)
2	쟤는 안 틀리는 문제가 없네	*thulli-*	'get wrong'	71% (n=71)	29% (n=29)
3	쟤는 안 까먹는 숙제가 없네	*kkamek-*	'forget'	90% (n=90)	10% (n=10)
4	쟤는 안 지각하는 수업이 없네	*cikakha-*	'be tardy'	95% (n=95)	5% (n=5)
5	쟤는 안 늦는 약속이 없네	*nuc-*	'be late'	95% (n=95)	5% (n=5)
6	쟤는 안 미루는 과제가 없네	*milwu-*	'procrastinate'	95% (n=95)	5% (n=5)
7	쟤는 안 우는 영화가 없네	*wul-*	'cry'	100% (n=100)	0% (n=0)
8	쟤는 안 다치는 경기가 없네	*tachi-*	'get hurt'	100% (n=100)	0% (n=0)
9	쟤는 안 태우는 음식이 없네	*tha-*	'burn'	100% (n=100)	0% (n=0)
10	쟤는 안 잃어버리는 물건이 없네	*ilhepeli-*	'lose'	100% (n=100)	0% (n=0)
11	쟤는 안 싸우는 날이 없네	*ssawu-*	'fight'	100% (n=100)	0% (n=0)

Table 2. Native Speaker Acceptability Judgment Results (n=100)

The distribution associated with *ppaci-* differs significantly from that of all other verbs combined (LL = 255.7, df = 1, $p < .001$).

Task 2: Interpreting DN in Generative AI Models

- Models: GPT-5.5, Gemini 3.1 Pro, and Claude Opus 4.5
- Design: 4 prompting conditions × 3 models on one *an-eps* DN sentence
- Measures: DN vs. NC-like choice, 1–7 confidence rating, and comparisons across models and prompting conditions (P1–P4)
- Focus: Effects of CoT and persona prompting on interpretation, and cross-model differences in confidence/ambiguity

With CoT, ChatGPT shifts to logical reasoning, overriding its pragmatic default.

Condition	Reading	Confidence	Note
P1 (Zero-shot)	NC-like	7/7	Recognized DN structure but defaulted to NC-like reading
P2 (Persona)	NC-like	7/7	No effect of persona assignment
P3 (CoT)	DN	7/7	Interpretive shift after stepwise scope analysis
P4 (Persona + CoT)	DN	7/7	CoT effect maintained; no additional persona effect

Table 3. ChatGPT 5.5 Results: Target Sentence *Yay-nun an-ppaci-nun moim-i eps-ney*

Gemini 3.1 Pro shows the same NC-like default, suggesting a general LLM processing tendency.

Condition	Reading	Confidence	Note
P1 (Zero-shot)	NC-like	7/7	No reasoning provided
P2 (Persona)	NC-like	7/7	Recognized DN structure but defaulted to NC-like reading
P3 (CoT)	DN	7/7	Interpretive shift after stepwise scope analysis
P4 (Persona + CoT)	DN	7/7	CoT effect maintained; no additional persona effect

Table 4. Gemini 3.1 Pro Results: Target Sentence *Yay-nun an-ppaci-nun moim-i eps-ney*

Claude consistently preferred the DN reading while recognizing pragmatic ambiguity, with confidence decreasing as context increased.

Condition	Reading	Confidence	Note
P1 (Zero-shot)	DN	6/7	Recognized ambiguity; flagged native speaker confusion
P2 (Persona)	DN	5/7	Confidence dropped; explicitly acknowledged pragmatic trap
P3 (CoT)	DN	4/7	Logical analysis confirmed DN; confidence dropped further
P4 (Persona + CoT)	DN	4/7	Same as P3; flagged literal vs. intended meaning split

Table 5. Claude Opus 4.8 Results: Target Sentence *Yay-nun an-ppaci-nun moim-i eps-ney*

	ChatGPT 5.5 / Gemini 3.1 Pro	Claude Opus 4.8
Default processing	Pragmatic	Compositional / logical
CoT effect	Interpretive shift (NC-like → DN)	Confidence drop only
Ambiguity awareness	Implicit	Explicit
Alignment with native speakers (P1/P2)	✓ (NC-like, 85%)	×
Confidence pattern	Stable (7/7)	Decreasing (6→4)

Table 6. Theoretical Implications: Cross-Model Interpretive Patterns

Implications from LLM Studies on Negation

Vision-language models fail to incorporate negation into their output when prompted with negative instructions (Vrabcová et al. 2025).

- **Negation blindness**: the model fails to register the negation marker altogether
- **NO syndrome**: the semantic effect of "no" is not properly reflected in the model's internal representations

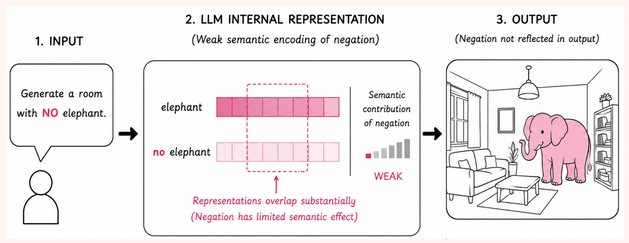


Figure 2. Weak influence of negation tokens on LLM representations. AI-generated and adapted by the author.

Negation tokens have limited representational influence across diverse contexts, weakening their meaning-reversing effect (Singh et al. 2024).

→ ChatGPT and Gemini's NC-like default under P1/P2 **parallels negation blindness**, favoring distributional patterns over compositional scope.

Conclusion

- The NC-like reading dominates in "*Yay-nun an-ppaci-nun moim-i eps-ney*"as a lexeme-specific exception, not a general property of the *an* + *eps* construction.
- Stepwise prompting revealed that logical reasoning yields DN, while pragmatic default yields NC-like across GenAI models.
- This NC-like preference is best attributed to conventionalized intuition or negation blindness (NO syndrome) rather than lexical negativity.

Selected References

- Lee, Chungmin. 2017. Metalinguistic negation vs. descriptive negation: Among their kin and foes. In *The pragmatics of negation*, 63–103. John Benjamins.
- Singh, Jaisidh, Ishaan Shrivastava, Mayank Vatsa, Richa Singh, and & Aparna Bharati. 2024. "Learn "no" to say "yes" better: Improving vision-language models via negations". *arXiv preprint*. *arXiv:2403.20312*.
- Vrabcová, T., Marek Kadlčík, Petr Sojka, Michal Štefánik & Michal Spiegel. 2025. Negation: A pink elephant in the large language models' room? *arXiv preprint arXiv:2503.22395*.