Title

Is an over-polite compliment worse than an impolite insult?:Pragmatic effects of non-normative politeness in Korean

Permalink

https://escholarship.org/uc/item/9k34v1hw

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Shin, Hagyeong

Doyle, Gabriel

Publication Date

2019

Peer reviewed

Is an over-polite compliment worse than an impolite insult?: Pragmatic effects of non-normative politeness in Korean

Hagyeong Shin (hshin@sdsu.edu), Gabriel Doyle (gdoyle@sdsu.edu)
Department of Linguistics, San Diego State University
5500 Campanile Drive, San Diego, CA USA 92182

Abstract

Honorifics in Korean appear as verbal inflections and have been considered as markers of politeness. This study investigates the pragmatic effects of honorifics, and suggests that honorifics can contribute to the semantic interpretation of verb phrases in complex ways. Native Korean speakers reported different inferred meanings of "did very well" and "did very poorly" based on the normative or non-normative honorific forms. We found significant effects of non-normative honorifics in positive assessments: over-polite honorifics brought negative interpretations. This suggests that pragmatic listeners interpret utterances based on the interaction between literal meanings, honorifics, and the normativity of the honorifics within a relationship context, to obtain an estimate of the speaker's intended meaning. This is inconsistent with the previous explanations of honorific usage as discernment or volitional politeness. We suggest that non-literal meaning inferences reflect listeners treating the honorifics as signals to potential communicative goals.

Keywords: pragmatics; semantics; politeness; honorifics; pragmatic inference; Korean

Introduction

Languages have many ways of expressing politeness. Some languages explicitly mark politeness with *honorifics*: grammaticalized or lexicalized forms for politeness. Honorifics are prevalent in languages such as Japanese, Javanese, Hindi, and the subject of this investigation: Korean. Because appropriate honorifics depend on the speaker-listener relationship, they primarily function as a reflection of social norms (*discernment politeness*). However, speakers may strategically deviate from the normative form in certain contexts, such as when making requests (*volitional politeness*, Hill et al., 1986).

In this paper, we investigate whether such deviations integrate more generally into the pragmatic inference process that listeners undertake when interpreting a message. Specifically, we look at whether a speaker's choice of honorific forms influences how a listener assesses the speaker's true opinion. We carry out judgment experiments to compile data on Korean listeners' interpretations. Our result shows that the inferred meaning of the message changes with honorifics in a complex manner that cannot be adequately explained by either strictly normative or strictly strategic use of honorifics. Instead, honorifics could be used as pragmatic signals to the meaning depending on the context.

Overall, this suggests that despite grammaticalized forms seeming to be low in semantic content, they can still significantly influence the inferred meaning of the message. We argue that a full understanding of honorific use will require their incorporation into frameworks of pragmatic inference, such as the Rational Speech Act framework (RSA, Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013).

Table 1: Honorific inflections of the past tense of "do" (*ha-*) in Korean. Honorification decreases from top to bottom and left to right. *-ess* is past tense suffix.

Speech Level	Honorific Suffix *-sy*	
	Present	Absent
DEF (deferential)	ha-**sy**-ess-*supnita*	ha-ess-*supnita*
POL (polite)	ha-**sy**-ess-e-*yo*	ha-ess-e-*yo*
INT (intimate)	ha-**sy**-ess-*e*	ha-ess-*e*
PLN (plain)	ha-**sy**-ess-*ta*	ha-ess-*ta*

Honorifics in Korean

Honorifics in Korean have two main realizations: honorific lexical items and honorific inflections. This study focuses on verbal honorific inflections, specifically speech levels and the *-sy* suffix. Table 1 demonstrates some of the honorific inflections that are available for the verb "do" (*ha-*). The speech level appears at the end of the verb phrase, and reflects the relationship between the speaker and the addressee. Four levels presented in Table 1 are tested in this study, and their perceived honorification decreases from top to bottom. The presence of honorific suffix *-sy* increases the honorification toward a subject of a sentence or a referent of the verb. This study examines cases where the subject of the sentence is the addressee, thus both the suffix and the speech level refer toward the listener. Honorific inflections, therefore, can be generally defined as stylistic features reflecting the speaker and listener's position within a social hierarchy, not the truth-conditional meanings (Sohn, 1999).

In colloquial Korean, speakers must choose some level of honorific inflection to form valid verb phrases; there is no default form or level. In most cases, the appropriate honorifics can be determined by the speaker-listener relationship, as honorifics were mentioned as relationship-acknowledging devices (Matsumoto, 1988). Honorifics are grammaticalized and conventionalized in relation to the speaker-listener relationship. Speakers using appropriate honorific forms assigned by the relationship context will therefore stay aligned with the normative use of honorifics. This type of honorific use can be summarized as *discernment politeness* (Hill et al., 1986; Ide, 1989; Koo, 1995).

Besides the normative use of honorifics, they can also be used more strategically. Politeness Theory (Brown & Levinson, 1987) has explained strategic honorific use through *negative politeness*, a politeness strategy for minimizing threats to the listener's *negative face*—the desire not to be imposed

upon.[1] This form of politeness is distinguished from *positive politeness*, a strategy used to minimize threats to *positive face*—the desire to be liked or approved. In the Politeness Theory perspective, speakers use honorifics largely to mitigate the potential face threats existing in the utterance. This type of honorific use can be summarized as *volitional politeness* (Hill et al., 1986; Ide, 1989; Koo, 1995)

These explanations for honorifics' uses are well-supported, but such general politeness strategies may represent only a subset of how honorifics are actually used. We argue that deviations from normative politeness levels can function as a pragmatic signal to the listener about the intended meaning of an utterance. We suggest that honorific use ties to a more general pragmatic behavior than previously described, providing pragmatic information beyond mitigating face-threats and potentially signaling a speaker's communicative goals.

Hypotheses

Based on the above discussion, we consider three hypotheses for the potential effects of honorifics on pragmatic inference. These span from a null pragmatic effect (if honorifics mainly express the speaker-listener relationship) to a monotonic relationship between inferred meaning and levels of honorifics (if honorifics mainly manage face-threat) to a complex relationship between honorifics and inferred meaning (if honorifics provide cues about the speaker's communicative goals).

To test this, we examined listeners' inferences of values for scalars: speakers' statements that a listener had done "well" or "poorly" on a test. We first described the speaker-listener relationship, then provided assessment sentences with eight honorific inflections from Table 1, and asked participants to estimate the exam score based on the assessment. More details are in the next section, but our hypotheses and the predictions they make follow.

Hypothesis 1: Honorifics are primarily about *discernment politeness*. **Changes in levels of honorifics will have no significant effects on pragmatic interpretation of scalars.**

Under this hypothesis, the speaker-listener relationship determines the appropriate honorifics, and forms that deviate from the normative standard would be similar to errors of subject-verb agreement—they could affect the perceived acceptability of a sentence, but not the meaning. If Hypothesis 1 is correct, we should not see differences in the listener's interpretations depending on the honorific forms used within the relationship context. This hypothesis is consistent with traditional analyses of the Korean honorifics, as in Sohn (1999).

Hypothesis 2: Honorifics primarily serve to mitigate face threat through *volitional politeness*. **As the utterance becomes more honorific, inferred values of scalars will be monotonically decreased.**

Under this hypothesis, speakers would use higher levels of honorifics to offset the negativity of an honest assessment. Therefore, we can expect to see a monotonic decrease in listeners' inferred values of scalars as the honorific level increases, with the "poorly" condition possibly showing a larger effect due to the more explicit face-threatening assessment. Being over-polite or being under-polite (relative to normative forms) should show opposite effects on the inferred meaning. This hypothesis is similar to the threat-management account of Politeness Theory (Brown & Levinson, 1987), or the social utility addition (Yoon et al., 2016) to the RSA framework explaining listeners' discounting of compliments when they thought the speaker was being polite.

Hypothesis 3: Honorifics, in addition to their discernment or volitional use, also can signal cues that influence the listener's interpretations in complex ways. **Effects of honorific levels will differ by the relationship context and the literal meanings of the utterance.**

Under this hypothesis, there will be a significant but non-monotonic effect of honorific levels. Unlike Hypothesis 2, here we do not necessarily expect under- versus over-polite messages (again, relative to normative forms) to have different effects on the listener's inference. Instead, deviations from normative honorifics could signal that the speaker is indicating different meanings or goals, for example, being hyperbolic or sarcastic. This hypothesis is similar to the QUD (Question Under Discussion) addition (Kao & Goodman, 2015) to the RSA explaining ironic interpretations.[2]

Experiment 1: Literal interpretations

Method

Design The purpose of Experiment 1 was to establish the literal baseline interpretations of the phrases "did very well" and "did very poorly". Each question in the experiment started with a vignette describing a conversation and a relationship context: a speaker is asked to tell a listener how the listener did on an exam, when the listener does not know of his own exam score. The speaker's assessments of the listener's exam score were then presented. Each participant rated 8 assessment sentences: 2 valences (positive, negative) in 4 relationship settings. Participants were asked to guess the listener's exam score in a number between 0 and 100.

Relationship settings were explicitly stated. In the Friend-Friend setting, the speaker and the listener were defined as friends who were in the same year at college. In the Upperclass-Underclass setting, the speaker was a student senior than the listener. In the Professor-Student setting, the speaker was a professor and the listener a student. In the Underclass-Upperclass setting, the speaker was a student junior than the listener. These four settings were chosen to have normative honorifics that allowed for a range of under-

---

[1]Despite the term, *negative politeness* is still a way of being polite; it is the "do no harm" counterpart to the "do good" sense of *positive politeness*.

[2]More details on the hypotheses can be found in the Open Science Foundation preregistration page: http://osf.io/s8nfu/register/5771ca429ad5a1020de2872e.

and over-polite forms by varying the honorific forms. In all settings, both the speaker and the listener had male Korean names to keep gender differences from influencing the result. **Stimuli** Participants saw the speaker's description of the listener's score presented as indirect quotes (i.e., [Speaker] said [Listener] did very well/poorly on the exam), so that participants would not see what honorific inflections the speaker used and thus would respond with their baseline inference in the absence of honorifics.

Participants Experiment 1 was posted on the online crowdsourcing website Dooit Survey (`http://www.dooit.co.kr`) based in South Korea. A total of 67 adult native Korean speakers completed the experiment for a small cash-value reward.

Result

Baseline scores In Experiment 1, literal interpretations of positive and negative phrases "did very well" and "did very poorly" were measured within each relationship setting. The mean of the scores in each condition were then treated as baseline scores representing literal interpretations in further analyses, since they represented the participants' estimates in the absence of honorifics. Baseline scores in each setting and condition are presented by horizontal dashed line in Figure 1. Participants reported mean baseline scores of 85.60 for the positive and 47.92 for the negative phrases. There was no significant differences according to *t*-tests between the relationship settings within the positive or negative valence, suggesting that participants viewed all four relationship settings having similar expected literal meanings.

Experiment 2: Inferences from honorific use

Design Experiment 2 followed the same basic idea of Experiment 1, but participants were asked to infer scores based on direct quotes, with honorific inflections. Deviations between the literal baselines from Experiment 1 and the inferences in Experiment 2 should therefore reflect pragmatic interpretations guided by the honorifics. Each participant rated a total of 16 sentences: 2 valences (positive, negative), each with 8 honorific inflections (4 speech levels × -*sy* present/ absent), in one of the 4 relationship settings (Friend-Friend, Upperclass-Underclass, Professor-Student, Underclass-Upperclass). After presenting the relationship context and the speaker's assessment, participants were again asked to infer the listeners exam score with a number between 0 and 100.

Stimuli Each assessment sentence was presented as a direct quote (i.e., [Speaker] said the following sentence: "[Address of the listener] did very well."). The presence of a direct quote meant that the sentence included one of the eight honorific inflections from Table 1, and therefore could influence participants' inferences accordingly. A sample vocative address of the listener by the speaker was included in these sentences to reinforce the normative honorifics for each relationship. In the Friend-Friend and Upperclass-Underclass setting, where the speaker was in an equal or higher position to the listener, the speaker addressed the listener with a plain "you". In the

Underclass-Upperclass setting, the honorific addressee term *senbay-nim* was used. In the Professor-Student setting, no addressee term was presented, because the speaker is on a much higher social rank than the listener and could in principle use any of the honorific inflections. Below shows the assessment sentences given in the Friend-Friend setting, with -*sy* and the deferential speech level[3].

neo	*cham*	*cal*	*ha-sy-ess-supnita*
You	very	well	do.AH.PST.DEF

Positive: "You did very well."

neo	*cham*	*mos*	*ha-sy-ess-supnita*
You	very	poorly	do.AH.PST.DEF

Negative: "You did very poorly."

Participants Experiment 2 was also posted on Dooit survey. Unlike Experiment 1, we asked each participant to answer for only one relationship type, to avoid any confusion about the speaker-listener relationships. 81 adult Korean participants were collected in total, with 20 participants in three settings and 21 participants in the Underclass-Upperclass setting. The participants in Experiment 1 and 2 were recruited separately.

Results

Baseline scores vs. Normative scores The baseline scores from Experiment 1 were then compared against the "normative" scores from Experiment 2. These normative scores are the mean of the inferred scores in each setting obtained from sentences with normative honorific forms established for that setting. Normative scores are presented by black points in Figure 1. For example, in the Friend-Friend setting, normative scores were calculated from the scores reported on sentences with intimate or plain speech level without an honorific suffix -*sy* (INT, PLN, black triangles in Figure 1). The Professor-Student setting did not include a specific addressee term to establish normative forms, because of the high social position the speaker was in. To calculate the baseline scores, we considered deferential and polite speech levels used as teacher's classroom register (Sohn, 1999). In Figure 1, normative forms (black points) aligned closely to the baseline scores (dashed line), showing that participants treated utterances with normative honorific forms similarly to the literal interpretations. This is confirmed by the regression below.

Baseline scores vs. Non-normative scores Scores reported on non-normative honorific forms are presented by the red points in Figure 1. In the Friend-Friend setting, participants reported the baseline score of 82.74 for the positive condition and 46.68 for the negative. When positive assessments appeared with non-normative -*sy* (red circles), regardless of the following speech levels, participants reported scores far lower than the baseline score (*sy*+DEF: 54.25, −28.49 score difference from the baseline score; *sy*+POL: 50.10, −32.64; *sy*+INT: 46.00, −36.74; *sy*+PLN: 56.25, −26.49)[4]. This is

---

[3]AH: Addressee honorification, PST: Past tense suffix, DEF: Deferential speech level

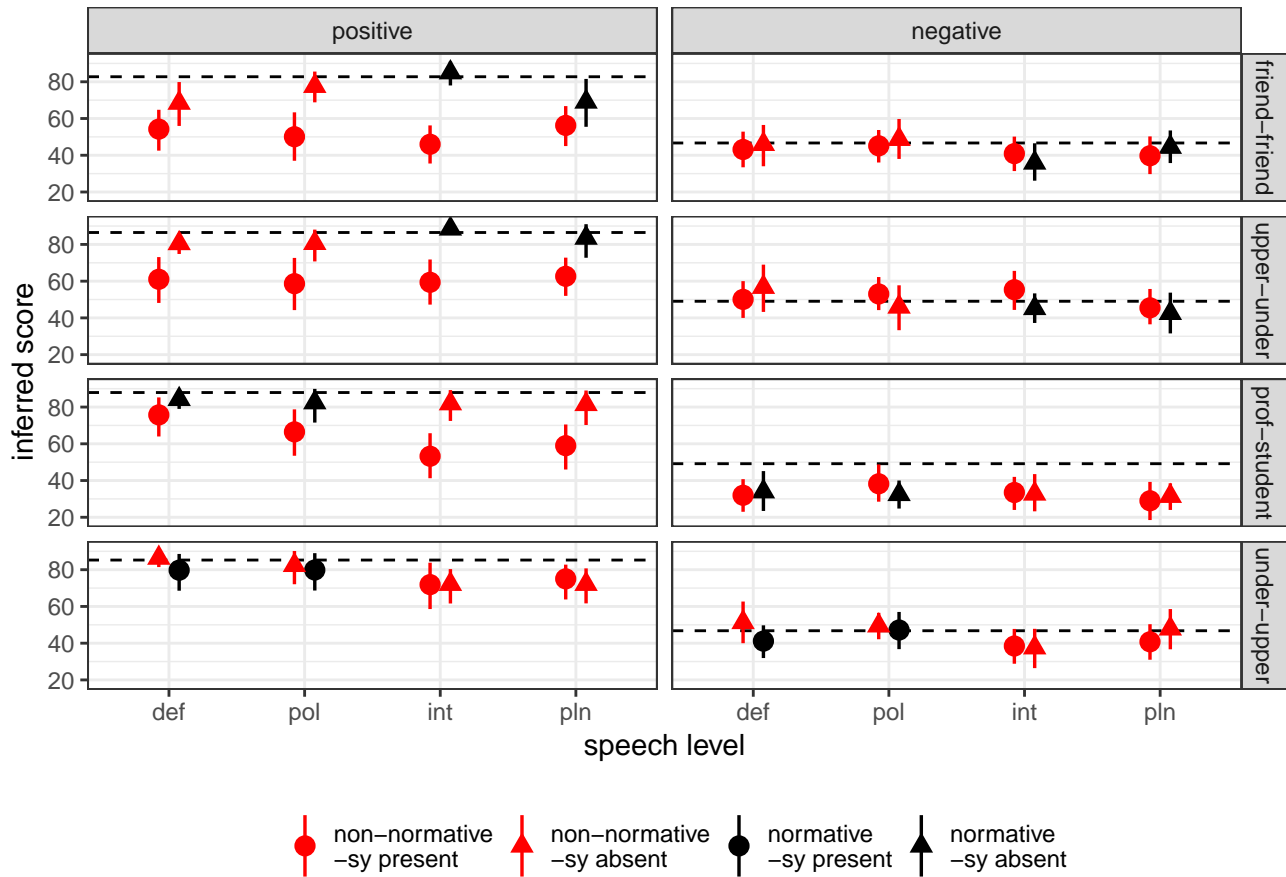[4]We test for significance on these values in the following regression model.

Figure 1: Participants inferred exam scores under valence (positive, negative), honorific suffix -*sy* (-*sy* present, -*sy* absent), and speech level (DEF, POL, INT, PLN) conditions in each relationship settings (friend-friend, upper-under, prof-student, under-upper). Circles and triangles indicate the -*sy* suffix being present or absent. The colors show the normativity of the forms, red being non-normative and black being normative. The dashed line in each condition shows baseline scores (from Expt 1). Vertical lines in each score point show 95% confidence interval calculated from 5000 bootstrap samples.

our first piece of evidence that over-polite forms can induce large pragmatic effects that substantially reduce the estimates of the test scores.

In the Upperclass-Underclass setting (Upperclass speaker), participants reported the baseline score of 86.49 for the positive condition and 49.05 for the negative. Normative forms were defined as intimate or plaivel without the -*sy* suffix (INT, PLN, black triangles). When positive utterances appeared with the -*sy* suffix (red circles), participants reported scores below the baseline score across the speech levels (*sy*+DEF: 61.05, −25.44; *sy*+POL: 58.65, −27.84; *sy*+INT: 59.40, −27.09; *sy*+PLN: 62.65, −23.84). Again, over-polite forms caused participants' pragmatic inferences to substantially drop.

Professor-Student setting showed the similar result. The baseline score was 87.91 for the positive condition and 49.17 for the negative. Normative forms in this setting were the deferential or polite speech level without the -*sy* suffix (DEF, POL, black triangles). Participants reported lower scores

when the professor's positive feedback were given with non-normative -*sy* (*sy*+DEF: 75.75, −12.60; *sy*+POL: 66.50, −21.41; *sy*+INT: 53.25, −34.66; *sy*+PLN: 59.00, −28.91). Under-polite forms (the intimate/plain speech levels) do not show any increase over the normative forms, as Hypothesis 2 would have predicted.

In the Underclass-Upperclass setting, the baseline score was 85.26 for the positive condition and 46.76 for the negative. The normative forms were defined as deferential or polite speech level with the -*sy* suffix (*sy*+DEF, *sy*+POL, black circles). This relationship setting showed the least amount of score variance among all settings. One explanation could be that non-normative forms in this setting produced outright socially unacceptable sentences. Not coincidentally, this is the one setting where the speaker is of a lower social standing than the listener. Since a lower-standing speaker speaking in under-polite forms violates social norms in a great degree, participants might have been confused with those sentences and have failed to properly reason on the meaning.

Table 2: Estimated effect sizes in the linear regression with random by-participant intercepts. Default values for Valence, Speech Level and Setting are negative, deferential and Underclass-Upperclass, respectively.

	β
(Intercept)	-0.29
Valence	
Positive	3.15
Speech Level	
Intimate	-4.40
Plain	-4.10
Polite	0.77
-*sy* suffix	
Present	-0.63
Setting	
Friend-Friend	-1.14
Professor-Student	-13.39 ***
Upper-Under	3.12
Valence × Speech Level	
Positive × Intimate	0.27
Positive × Plain	0.12
Positive × Polite	-2.30
Valence × -*sy* suffix	
positive × -*sy* present	-15.88 ***
Valence × Setting	
Positive × Friend-Friend	-10.46 **
Positive × Professor-Student	6.67
Positive × Upper-Under	-9.62 **

$^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$

Overall, the presence of the over-polite and non-normative honorific -*sy* suffix when talking to an equal or lower-standing listener signaled participants that the positive feedback could not be taken literally, and participants substantially reduced their estimates of the test scores. This is strongly at odds with Hypothesis 1. At the same time, non-normative scores (red points) were mostly at or below the normative scores (black points) in the positive valence cases, regardless of the non-normative forms being over- or under-polite. This argues against both Hypothesis 1 and Hypothesis 2.

Linear mixed-effect model

We fit a linear mixed-effect model predicting the score differences (inferential score from Expt 2 minus the baseline score from Expt 1) with random by-participant intercepts. The fixed-effect variables were Valence × Speech level, Valence × -*sy* suffix, Valence × Setting. Regression coefficients (β) are reported in Table 2. The default values were set to the scores obtained in the Under-Upper setting, from negative sentences with no -*sy* suffix and the deferential speech level.

Starting from the top of Table 2, the intercept term confirmed that there were no significant differences between inferred scores and baseline scores in the default setting of the model. Among the main effects of the Setting, only the Professor-Student setting showed significantly lowered scores. This was because the default condition (no -*sy* suffix, deferential speech level) of the model was a normative form in the setting, thus brought literal (thus, more negative) meaning. Overall, the main effects were largely small and non-significant, with the exception of one setting. This confirmed our initial anticipation that honorifics' meanings should be considered relative to the context, including the literal meanings of the message and the speaker-listener relationship.

Moving onto the interaction terms, Valence × -*sy* suffix had a large effect, lowering the inferred scores by 15.88 points from the baseline ($p < 0.001$). This shows that positive sentences with the -*sy* suffix in the three settings where the speaker was at least equal in social standing to the listener (Friend-Friend, Professor-Student, Upperclass-Underclass) showed lower scores than in the one where the speaker was of higher social standing. Those three settings shared the normativity context that the -*sy* suffix was an over-polite form, and the inferred scores dropped as a result.

Valence × Setting term also reflected the result shown in Figure 1. Friend-Friend and Upperclass-Underclass showed lowered scores in positive sentences with deferential speech level (this was default setting of the model, which was a non-normative form in the relationship). The Professor-Student setting did not show significant differences from Underclass-Upperclass setting, again because the deferential speech level was the normative form in the relationship. This showed that the Friend-Friend and Upperclass-Underclass setting behaved similarly, as higher honorifics became non-normative and over-polite forms. We could see that Professor-Student and Underclass-Upperclass behaved similarly as well. These two settings shared deferential and polite speech levels (higher honorifics) as normative forms.

In both the numeric values in Figure 1 and the regression coefficients in Table 2, we see a few patterns. First, participants' inferred scores varied substantially based on the honorifics, contrary to Hypothesis 1, which considered the honorifics to primarily serve as an agreement to the relationship. This suggests that listeners assume that speakers have made volitional choices in their honorific inflections when they deviate from normativity. Second, looking at the regression model, we see strong evidence of an effect of the -*sy* suffix in non-normative context, generally lowering the score differences (inferred score − baseline score). But there is no consistent effect with the speech levels. This runs counter to the expectation of Hypothesis 2; there is no monotonic relationship between the honorific levels and the inferred scores. Instead, we see a complex pattern that is driven largely by the normativity of the forms, rather than their relative politeness. In the next section, we discuss the implications of these results and sketch a possible explanation for the phenomenon.

Discussion

The experiment results showed clear differences on listeners' interpretations, depending on the honorific inflections, the speaker-listener relationship, and the valence of the verb phrases. Contrary to Hypothesis 1, we saw significant effects of honorifics on the pragmatically-inferred values even within a given relationship setting. Contrary to Hypothesis 2, we did not see a monotonic effect of politeness levels; deviation from the normative form generally decreased (or maintained) the inferred value of positive verb phrases regardless of whether the non-normative form was more or less polite than expected. Furthermore, in the negative valence, normative and non-normative forms showed no significant difference in their interpretation. These results suggest that while there appears to be a volitional component to the speaker's choice of honorific forms, the choice extends beyond straightforward face-threat mitigation.

Hypothesis 3 therefore appears to be the best fit to our data, though in some sense it reflects a less specific explanation at present. How can we further expand our hypothesis to explain the observed variance in pragmatic inferences? A promising direction is to build on the Rational Speech Act (RSA, Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013) framework. The core idea of the RSA is that speakers and listeners each know that the other is trying to communicate in an efficient, rather than a literal, manner. The listener considers the speaker's choice of utterance as a rational decision over the set of available alternatives. The speaker prefers utterances that maximize the expected conversational utility (such as maximizing the listener's probability of inferring the intended message). The listener then uses a recursive inference process to determine the most likely meaning.

Two particular extensions of RSA inference may be relevant for unravelling the inferences that result from honorifics. Yoon et al. (2016) proposed that the speaker not only has a desire to provide epistemic utility in their utterance (giving the listener an accurate representation of the world) but also social utility (such as minimizing listener's face-threats). Honorifics, especially in the Politeness Theory (Brown & Levinson, 1987) framework, can supply social utility alongside the epistemic utility of the message itself. This fits with, for instance, speakers' selective use of honorifics when making requests. In our data, however, we see that the same honorific forms lead to substantially different inferences based on the setting. Even if we view social utility relative to the speaker-listener relationship, with over-polite forms adding social utility and under-polite forms reducing it, this would still not be sufficient to explain the variation in Figure 1.

By combining the idea of social utility with a goal- or QUD-based approach (Kao & Goodman, 2015) in RSA, though, we may be able to capture the pragmatic effects of honorifics. A Goal/QUD framework says that when a speaker produces a message that seems to violate the listener's expectations, the listener may instead interpret the message with a different goal in mind. For example, if a speaker complains that they paid an unbelievably high cost for some object, the listener may infer that the speaker's epistemic utility is not coming from conveying the literal cost but rather an affective interpretation of the cost (i.e., hyperbole).

Building on these extensions to the RSA framework, we suggest that Korean honorifics may be modelled as an interaction between the relationship context r, shared knowledge of normative honorifics k, an intended meaning s, and a goal g. The speaker's choice of utterance can be broken down into the semantic content of the word stem c and the honorific inflections m:

$$P_{speaker}(c, m | s, r, k, g) \qquad (1)$$

This expresses the idea that a speaker chooses c and m jointly to deliver their intended meaning s, conditioned on the relationship context r and normativity k for the honorifics, as well as their communicative goal or QUD g. If we assume that the listener has no uncertainty about the relationship r or normativity k, we can express the listener's inference process as Bayesian inference, marginalized over the potential goals of the speaker:[5]

$$P_{listener}(s | c, m, r, k) \propto \sum_g P_{speaker}(c, m | s, r, k, g) P(s) P(g) \quad (2)$$

This joint distribution over c and m gives the model the flexibility to capture the complex patterns in our results in a way that a basic social utility term alone cannot. Being overly polite may come from the speaker signalling their ironic intentions by violating normative expectations of the honorific inflections. When such a deviation from the norms is slight, or consistent with a goal of mitigating face-threat, the listener merely tweaks their interpretation. When the deviation from the honorific norms is large (as when a student is overly polite to their friend, or the professor talks to the student with the honorific -*sy* suffix), the listener assumes the speaker's goal has changed. In this way, honorifics signal cues to the meaning similar to the inferred product prices in Kao et al. (2014); a small deviation from expectations retains an approximately literal interpretation, while a large deviation triggers an ironic interpretation. This argument could be verified by a follow-up experiment measuring inferred goals.

Conclusion

We have examined honorific inflections and their effect on pragmatic inference. Contrary to discernment or volitional politeness accounts, we find complex interactions between honorifics and a listener's pragmatic interpretation. We propose that this result may be explained with an extended RSA framework with jointly-distributed content and honorifics that can both provide social utility and serve to signal a speaker's communicative goals.

---

[5]Of course, the listener may want to update their belief about their relationship with the speaker based on the speaker's choice of honorifics! If so, the listener could marginalize over r and k.

Acknowledgements

References

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Hill, B., Ide, S., Ikuta, S., Kawasaki, A., & Ogino, T. (1986). Universals of linguistic politeness: Quantitative evidence from Japanese and American English. *Journal of pragmatics*, *10*(3), 347–371.

Ide, S. (1989). Formal forms and discernment: Two neglected aspects of universals of linguistic politeness. *Multilingua-journal of cross-cultural and interlanguage communication*, *8*(2-3), 223–248.

Kao, J. T., & Goodman, N. D. (2015). Let's talk (ironically) about the weather: Modeling verbal irony. In *Proceedings of the 36th conference of the cognitive science society* (pp. 1051–1056).

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Koo, J. S. (1995). *Politeness Theory: Universality and Specificity*. Unpublished doctoral dissertation, Harvard university.

Matsumoto, Y. (1988). Reexamination of the universality of face: Politeness phenomena in Japanese. *Journal of Pragmatics*.

Sohn, H. M. (1999). *The Korean Language*. Cambridge: Cambridge University Press.

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2771–2776).