Study on the Domain Adaption of Korean Speech Act using Daily Conversation Dataset and Petition Corpus

Youngsook Song
Sionic AI Inc.
Seoul, Korea
song@sionic.ai

Won Ik Cho
Seoul National University*
Seoul, Korea
tsatsuki@snu.ac.kr

Abstract

In Korean, quantitative speech act studies have usually been conducted on single utterances with unspecified sources. In this study, we annotate sentences from the National Institute of Korean Language's Messenger Corpus and the National Petition Corpus, as well as example sentences from an academic paper on contemporary Korean vlogging, and check the discrepancy between human annotation and model prediction. In particular, for sentences with differences in locutionary and illocutionary forces, we analyze the causes of errors to see if stylistic features used in a particular domain affect the correct inference of speech act. Through this, we see the necessity to build and analyze a balanced corpus in various text domains, taking into account cases with different usage roles, e.g., messenger conversations belonging to private conversations and petition corpus/vlogging script that have an unspecified audience.

1 Introduction

People use statements to reveal the intent of a proposition or to express their promises or emotions. However, similar can be applied to questions. Generally speaking, interrogatives are uttered in situations where the speaker does not know the relevant information but assumes that the listener does. To express a question, a speaker would use an interrogative ending and a question mark in written language, or a rising intonation in spoken language. Nonetheless, the use of interrogative endings, question marks, or rising intonation does not necessarily constitute interrogative speech. In this regard, the examples given by Song (2010) and Park (2019) are as follows.

(1) a. Mr. Lee: 바보.. 메주야 넌! (*Fool.. you idiot!*)
Bom: 아휴! 내가 왜 메주야! (*Ahhh! Why am I an idiot!*) (Song (2010): 98)

b. 나라의 운명을 외국의 손에 맡겨서야 되겠습니까 (*Do we hand over the fate of our country to foreigners?*) / 이런 걸 누가 먹겠습니까 (*Who would eat something like this*) (Park (2019): 16)

Example (1a) emphasizes the speaker's negative emotions by utilizing a distinctive speech style, particularly through the use of the interrogative *'why'* by the speaker in the 'Bom' example. (1b) Despite adopting the forms of Yes-No-Questions and Wh-Questions, it is not readily classified as an interrogative speech act because it is used to emphasize the opinion rather than to elicit information. Notably, humans tend to adeptly comprehend the speaker's intention, even when a disparity exists between explicit form and implicit intent. However, artificial intelligence (AI) models may face challenges in such interpretive tasks. Consequently, as exemplified above, speech act annotations could contribute to enhancing the utterance performance of AI models, particularly in instances where the latent meaning of an utterance diverges from its manifest content.

In this study on the Korean speech act, an attempt is made to measure the performance of AI models analyzing distinctions between locutionary and illocutionary force, especially when disparities exist between the two. For the purpose of performance measurement, frequently mispredicted speech acts are typified. For instance, there are cases, such as example (1b), where the emphasis on intention may be misinterpreted as a question because the context is not specified. This is similar to how, without context, it is difficult for humans to categorize 'speech act' into specific categories. In circumstances where distinguishing speech act is possible only if given context, the likelihood of models correctly identifying the answer may become notably low in the case of sentence-level annotated data. Conversely, even without any con-

*Work done after graduation.

Proceedings of the Joint 3rd NLP4DH and 8th IWCLUL, pages 229–234

Statement	*Statement*	Declarative utterances that include or convey proposition
	Future Intention	Utterances that describe the speaker's will or promises
	Sarcasm/Humor	Utterances that convey the speaker's sarcasm or humor towards the object
Suggestion	*Suggestion*	Commands or requests, including short directions
Exclamation	*Exclamation*	Utterances with expressions that display daily emotions
Question	*Yes-No-Question*	Polar and multiple choice questions
	Wh-Question	Open questions that require further answers
	Rhetoric-question	Questions that do not require an answer from the addressee
Greeting	*Greeting*	Conventional greetings including optatives
	Adress term	Addressing others with name or title

Table 1: Speech act annotation criteria.

text provided, if a specific speech act is commonly utilized in a particular discourse situation, anticipations of relatively effortless performance improvements can be posited through the construction of a sufficiently large and diverse corpus. Therefore, this study intends to scrutinize, in detail, various instances such as National Institute of Korean Language (NIKL)'s Messenger Corpus (2022) (which was updated from 2020 NIKL corpora (NIKL, 2020)), excerpts from an academic paper on contemporary Korean vlogging, and the titles of public petitions (those are in oratory style), to identify under which circumstances models incur errors in speech act classification. Initially, after annotating speech acts in conversations within the Messenger corpus, we undergo automatic classification with a widely used pretrained language model (PLM), the bert-base-multilingual-cased model (Devlin et al., 2018).

2 Speech Act Annotation

2.1 Speech Act Theory

Regarding the definition of speech acts, this study adopts Austin and Searle's speech act theories. Austin (1962) categorizes speech acts into Commissives, Verdictives, Exercitives, Behabitives, and Expositives, and describes the speaker's 'utterance intention' as an illocutionary force, and they more adapted in Searle (1976) to a criteria that is widely applicable. Though Stolcke et al. (2000) added rhetorical question as a notable dialogue act among other forty speech act classes, in a more recent and systematic approach, Bunt et al. (2010) encompassed the tripartite classification of questions, namely propositional questions, check questions, set questions/choice questions. In a relatively recent study on Korean, Cho and Kim (2022) distinguished between usual questions and rhetorical questions (denoted as RQ) within the questions, and also within commands; they categorized directives as commands if they solicited a specific action, and otherwise as rhetorical commands (denoted as RC), which is particularly significant in optatives.

In this study, we also deem it necessary to distinguish between the locutionary act, which pertains to the sentence's meaning and directive action, and the illocutionary act, which involves subsequent speech actions such as promises, commands, and coercions. Furthermore, if a developed model can comprehend and generate speech acts based on these distinctions, it could be applied to and utilized across various industrial domains.

2.2 Data Annotation

For the annotation of data that is adopted in the model training, NIKL Messenger Corpus (NIKL, 2020) was utilized by collecting a total of 33,138 sentences from 3,840 files. The source data was collected from free conversation of the participants and is available under application from NIKL online page[1].

In addition to the Messenger Corpus, a more challenging evaluation set with 125 sentences was constructed by extracting examples from a research paper on contemporary Korean vlogging and microblogging (Park, 2022) and bringing titles from public petitions[2]. They are known to characteristically reveal differences between locutionary and illocutionary forces. For instance, Park (2022) claimed that '-e ju-' (to give) has recently been used

[1]The data can be obtained from https://corpus.korean.go.kr/request/reausetMain.do?lang=en and its processing can be conducted with the help of Korpora repository https://ko-nlp.github.io/Korpora/en-docs/corpuslist/modu_messenger.html

[2]Available in https://github.com/lovit/petitions_dataset

	Messenger	Petition/Vlog
# Sentences	6,407	125
Accuracy	85.92	52.87
Macro F1	51.61	22.44

Table 2: Speech act classification evaluation on two test sets of different domain (trained on the messenger dataset).

among Korean language users as a predicate to describe the behavior of the speaker her/himself, dominantly in the context of vlogging and microblogging. Also, owing to conventional pro-drop in Korean, this kind of phenomenon would make it much more difficult for trained models to infer the speech act just given a single utterance. Also, petition titles usually aim to appeal to the readers by using eye-catching phrases that include sarcasm (a representative figurative language where the user intention may differ from the locutionary force) or rhetorical questions, which also contribute to the classification difficulty.

Song (2023) took into account these kinds of language changes in contemporary Korean and addressed new criteria of Korean speech act categorization (Table 1). Speech acts were divided into five major categories following Austin (1962) and Searle (1976): *statement* that corresponds with declaratives, *suggestion* with directives, *exclamation* with exclamatives, *question* with interrogatives, and *greeting* with conventional expressions, with additional subcategories like *sarcasm/humor* and *rhetorical questions* added. We adopt these criteria for the annotation of datasets collected above; that is, we annotate the Messenger Corpus that consists of contemporary Korean colloquial utterances, use it for the model training, and check the model performance using all three types of sentences.

The annotation was conducted by computational linguists who have experience of Korean speech act annotation[3]. Especially for the test set with challenging examples (48 for vlogging expressions and 77 for petition titles), three Korean computational linguists participated in the annotation and obtained the Kappa of 0.715 (Fleiss, 1971)[4].

[3]The data and agreement/label are available only privately upon application due to the policy of the original providers.

[4]Available in https://github.com/songys/DAKoSA-Domain_Adaptation_in_Korean_Speech_Act

2.3 Experiment

For the automatic classification of speech acts, we adopted the bert-base-multilingual-cased model (Devlin et al., 2018) that utilized Wikipedia data for pre-training.

The model classification of speech acts underwent a fine-tuning process, a learning method conventionally used for PLM downstream tasks. The training set consists of randomly selected 25,611 instances (80%) of Messenger Corpus, while the test set incorporates 6,407 instances (20%) of it (batch size 32 with AdamW (Loshchilov and Hutter, 2017) optimizer). Accuracy and Macro F1 scores were used as evaluation metrics.

The classification accuracy for the messenger corpus was 85.92, with the F1 score 52.87 (Table 2). To verify whether the trained model adapts to comparably unseen expressions, a test conducted using the public petition titles and vlogging evaluation set (125 instances). We obtained the accuracy of 51.61 and F1 score 22.44, which implies that the model performance significantly differs from the validation with homogeneous dataset. It displays the discrepancy that comes from the domain difference of both types of sets.

3 Analysis

3.1 Visualization and Error Analysis

To analyze the classification results, error rates among speech act categories were visualized through a heatmap generated via a confusion matrix (Figure 1), for the evaluation with Messenger Corpus (homogeneous to the training corpus). It was notably observed that the misclassification of *statements* as *suggestions* was prevalent, reaching 95 instances, and thus representing the most frequent misprediction. Furthermore, the error of classifying *future intentions* as *statements* was also significant, amounting to 88 instances. Overall, due to the high frequency of statements, the absolute frequency of misprediction involved *statements* being confused with *suggestions* or *future intentions* being misclassified as *statements*. Conversely, while not a high-frequency speech act, *rhetorical questions* demonstrated their trickiness, as the model did not accurately identify any instances, instead incorrectly categorizing them as yes-no questions in 23 instances. This exhibited a relatively high error rate in comparison to cases of accurate identification.

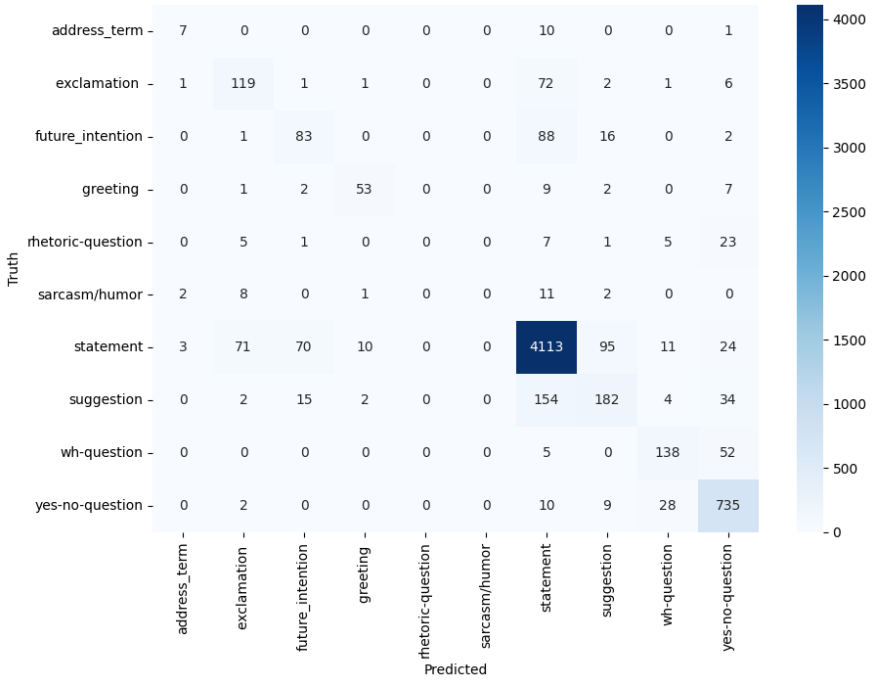

Figure 1: Confusion matrix on the Messenger Corpus.

(2) Speaker 1: 이게 진정한 미식의 길이지 (*This is how a true foodie does it.*)

Speaker 2: ㅋㅋㅋ잘했다 폭식의 길 아닐까 (*Well done! Sounds like a road to gluttony to me.*)

Speaker 1: 조용히 해 줄래? ㅋㅋ (*Could you be quiet? lol*)

In example (2), '조용히 좀 해 줄래? ㅋㅋ (*Could you be quiet? lol*)' was interpreted as a *rhetorical question* by human annotators, but the model classified it as a *yes-no question*. In cases like the aforementioned example, humans might interpret the utterance variously as a *rhetorical question*, a *yes-no question*, or even an imperative, depending on the context. Such errors are presumed to stem from training the model at the sentence level without contextual information. Conversely, in the following example, both humans and the model successfully classified the utterance as a *rhetorical question*.

(3) Speaker 1: 와 가식쟁이다ㅋㅋㅋ (*Wow, what a hypocrite.*)

Speaker 2: 어쩌라고 죽을래? (*What are you gonna do about it, wanna die?*)

In the instance of example (3), responding with '죽을래' ('*wanna die?*') to the term '가식쟁이' ('*hypocrite*') poses a challenge to classify as either a *yes-no question* or an imperative. Thus, in clear contexts like this, both human annotators and the model aptly classified it as a *rhetorical question*, in contrast to situations where context is not provided and where the error rate appeared to be high due to interpretative challenges.

3.2 Further Analysis on RQs

A notable observation from the confusion matrix is that, in the case of *rhetorical questions*, out of 42 questions, 23 were annotated as *yes-no questions*, and 5 as a *wh-question*. It becomes evident that instances like *rhetorical questions*, where the overt sentence form and the underlying semantics differ, present heightened difficulties in classification.

Here, we discuss the case with examples from petition titles in which the model mispredicted a *rhetorical question* as a *yes-no question* (Table 3). Questions concern societally controversial topics in Korea, such as women's military service (which is not mandatory de facto), compensation issues for injuries during the service, and questions on murder and fundamental human rights issues. In these examples, humans annotated a question like "*Is it reasonable not to go to the army just because someone is female?*" not as a question necessitating a binary 'yes' or 'no' answer but as a *rhetorical question*, interpreting it as an emphatic expression. However, the model, probably not having been previously exposed to such types of questions (even in the Messenger Corpus where the sentences are

Example	Human annotation	Model prediction
여성이라는 이유만으로 군대를 안가는게 정상적인가요? (*Is it reasonable not to go to the army just because someone is female?*)	*rhetorical-question*	*yes-no-question*
군대에서 다쳤으면 국가가 보상해야 되지 않나요? (*Isn't it a duty of the nation to compensate for the injury in the army?*)	*rhetorical-question*	*yes-no-question*
부산 여중생 사건 이런 일 정말 반복 안될 수 없을까요? (*Couldn't we stop such a tragedy, like Busan middle schoolgirl incident?*)	*rhetorical-question*	*yes-no-question*
살인을 해야 살인자입니까? (*Do we only call someone a murderer only if he or she commits murder?*)	*rhetorical-question*	*yes-no-question*

Table 3: Petition examples where the model prediction differs from the human annotation.

daily conversation), categorized it as a *yes-no question*. One consideration that needs to be taken into account in a speech act analysis system is that a meticulous analysis of the domain of usage should precede before the inference.

The following example of vlogging text also represents a similar case.

(4) (김치찌개를 끓이는 영상)... 냄비에 채소 먼저 깔아 주고 김치를 반 포기 정도 ①**넣어 줍니다**. ...돼지고기 넣고 푹 ②**끓여 줄게요**. 고기는 목살이에요. (고기가 어느 정도 익은 후에) 먹기 좋은 크기로 ③**잘라 줍니다**. (각종 양념을 넣는다는 설명) 잘 섞어서 오래 ④**끓여 줄게요**....
*(In a video of cooking kimchi stew)... First, put the vegetables in the pot and then ①**add about** half a head of kimchi. ... Add pork and ②**simmer thoroughly**. The meat is pork neck. (After the meat has been cooked to some extent) ③**Cut it into** bite-sized pieces. (Explaining that various seasonings are added) Mix it well and ④**boil for a long time**. ...* (Park, 2022)

Example (4) above highlights a section from a vlog video wherein the speaker, a vlogger, is describing the ongoing process of a cooking activity s/he is engaged in. Notably, the speaker uses the '-어 주-(-e ju-)' expression, as in '넣어 줍니다' (add something) and '잘라 줍니다' (cut something), wherein the agent and the beneficiary of the action reside in the same clause.

So far, in the Korean language, these expressions have not been used by language users to describe the behavior of the speaker her/himself. In this regard, in the experiment using vlogging script, the model predicted 5 out of 6 items as *suggestions* in instances for the pro-drop cases (frequent in Korean spoken language), and predicted as *statements* when the subject was explicitly stated. In other words, the intention of these types of utterances can be determined upon the viewpoint and timestamp of the analysis; the vlogger would have said the utterance with an intention of describing his/her behavior, but the audience of the vlog would interpret it as a suggestion of cooking sequences. This implies that, particularly in pro-drop languages like Korean, a correct understanding of utterance intent may be possible if and only if an accurate and contextual speech act annotation is performed, which reflects the importance of not only domain but also cultural and time-variant characteristics.

4 Conclusion

In this study, speech acts were annotated on the NIKL Messenger Corpus, the titles of public petitions, and vlogging scripts, focusing on the analysis of error items in sentences with discrepancy between locutionary and illocutionary force. Additionally, it turned out that stylistic features used in a specific circumstances also influence the decision of speech acts. Considering different contexts, such as messenger conversations that belong to private dialogue and public petitions or vlogging script that have the nature of having the audience, it is deemed necessary to build and analyze balanced corpora across various domains concerning whether the discourse is public or not and having multiple or anonymous addressee.

Acknowledgments

We thank anonymous reviewers for constructive comments. Authors are also grateful for Jisu Shim for helping the annotation process.

References

J. L. Austin. 1962. *How to do Things with Words*. Oxford University Press, New York. Reprinted 1975.

Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum.

2010. Towards an ISO standard for dialogue act annotation. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Languages Resources Association (ELRA).

Won Ik Cho and Nam Soo Kim. 2022. Text implicates prosodic ambiguity: A corpus for intention identification of the korean spoken language. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(1):1–20.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

NIKL. 2020. Nikl corpora 2020 (v.1.0).

Jinho Park. 2019. Yes-no-questions and rhetorical questions. *Hangughagbo*, pages 1–24.

Mi-eun Park. 2022. About the '-e ju-' construction of that i use to myself –focused on vlog register–. *Korean Semantics*, 78:89–117.

John R Searle. 1976. A classification of illocutionary acts1. *Language in society*, 5(1):1–23.

Sanghoun Song. 2010. Pragmatic usage of wh-elements in korean. *Language Information*, 11:91–113.

Youngsook Song. 2023. *Enhancing AI's Commonsense Reasoning in Conversations through Natural Language Generation*. Ph.D. thesis, University of Kyunghee.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.