

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

Youngsook Song<sup>1</sup> and Won Ik Cho<sup>2</sup>

<sup>1</sup>Sionic AI Inc., Seoul, Korea

<sup>2</sup>Seoul National University\*, Seoul, Korea

Corresponding author: Won Ik Cho, [tsatsuki@snu.ac.kr](mailto: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.

## Keywords

Korean, Speech Act, Domain Adaptation, Daily Conversation, Petition Corpus

## I INTRODUCTION

People use statements to reveal the intent of a proposition or to express their promises or emotions. However, a similar principle 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

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

Table 1: Speech act annotation criteria.

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 acts, an attempt is made to measure the performance of AI models distinguishing locutionary and illocutionary force, especially when disparities exist between the surface form and the semantics. In this light, frequently mispredicted speech acts are primarily 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, without context, how 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 context 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]), the titles of public petitions (those are in oratory style), and excerpts from an academic paper on contemporary Korean vlogging, to identify under which circumstances models incur errors in speech act classification. After annotating speech acts in conversations within the Messenger corpus, we undergo automatic classification with a widely used pretrained language model (PLM), multilingual BERT [Devlin et al., 2018].

## II SPEECH ACT ANNOTATION

### 2.1 Speech Act Theory

On 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 ‘intention’ as an illocutionary force, which was 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 within the questions, and also within commands; they categorized directives as commands if they solicited a specific action, and otherwise as rhetorical commands, 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.

## 2.2 Speech Act Annotation for Modern Korean Corpora

Recent studies on Korean speech act annotation are variation of general speech act studies above, but Song [2023] took into account language changes in contemporary Korean and addressed new criteria of Korean speech act categorization (Table 1).

Speech acts are organized into five major categories following 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. Before finalizing the categories, various types of question such as self-addressed questions [Ginzburg et al., 2013] and tag questions [Kalouli et al., 2021] were additionally considered. However, self-addressed questions could be categorized as either yes-no question, wh-question, or rhetorical question, and tag questions like "You are coming, aren't you?" could be considered as a type of rhetorical question. Other kinds of rhetorical questions like "How can I love and respect someone who doesn't love and respect herself?", which frequently appeared on English Twitter [Paul et al., 2011], were also commonly observed in Korean messenger corpora, positioning as one of the main categories. Below we add a detailed characteristics of each speech act considered in the annotation phase.

- **Statement** - *Statement*
  - Aims at widening common ground between the speaker and the addressee
  - Most commonly observed among statement categories (less priority in the annotation)
- **Statement** - *Future intention*
  - Aims at conveying the speaker's future plan involved with his/her intention
  - More priority in the annotation (between statement categories)
- **Statement** - *Sarcasm/Humor*
  - Aims at conveying the speaker's sarcastic stance towards the situation or the speaker's will to entertain the addressee
  - More priority in the annotation (between statement categories)
- **Question** - *Yes-no question*
  - Requires yes/no answer of the addressee
  - Also includes alternative (multiple-choice) questions
- **Question** - *Wh-question*
  - Requires open answer of the addressee
  - Also includes indirect expressions that can be interpreted as questions
- **Question** - *Rhetorical question*
  - Questions that does not require the answer of the addressee
  - Aims at conveying or emphasizing the speaker's emotion or thought, using the surface form of question
  - Usually conveys the speaker's astonishment, anger, reprimand, etc.
- **Suggestion** - *Suggestion*

	<b>Messenger (Train)</b>	<b>Messenger (Test)</b>	<b>Petition (Test)</b>	<b>Vlogging (Test)</b>
<i>Statement</i>	18,160	4,540	1,206	24
<i>Future Intention</i>	673	169	5	-
<i>Sarcasm/Humor</i>	138	35	19	-
<i>Yes-No Question</i>	2,409	603	14	-
<i>Wh-Question</i>	1,248	313	22	-
<i>Rhetoric Question</i>	507	127	117	-
<i>Suggestion</i>	966	242	4,936	24
<i>Exclamation</i>	712	179	2	-
<i>Greeting</i>	357	90	5	-
<i>Address term</i>	68	17	10	-
<i>Unknown</i>	372	93	72	-
	25,610	6,408	6,408	48

Table 2: Statistics of the annotated corpora. Unknown denotes the sentences that more than half of the taggers found difficulty in the annotation.

- Assigns a to-do-list to the addressee in the form of command, request, suggestion, advice, prohibition, etc.
- Usually effective when such to-do-list is not optatives or idiom
- Counts when addressee is not specified
- **Exclamation** - *Exclamation*
  - Includes exclamative expressions that display human emotion such as astonishment or embarrassment
  - Less priority in the annotation (since the portion of utterances only consisting of exclamation is small)
- **Greeting** - *Greeting*
  - Daily expressions including greetings, appreciation, optatives, etc.
- **Greeting** - *Address term*
  - Utterances that aim at calling the addressee with the name or title

We adopt these criteria for the annotation of datasets we adopt in this study; that is, we annotate corpora with contemporary Korean utterances including colloquial and web-style texts, use it for the model training, and check the model performance using all three types of sentences.

### 2.3 Data Annotation

For the annotation, NIKL Messenger Corpus (NIKL [2020], hereafter messenger corpus) was utilized by collecting a total of 32,018 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<sup>1</sup>. We split the messenger corpus into 25,610 sentences (training and validation) and 6,408 sentences (test).

In addition to the Messenger Corpus, for more challenging evaluation we collected two more types of text: 6,408 titles from public petitions<sup>2</sup> (hereafter petition corpus) and 48 example sentences from a research paper on contemporary Korean vlogging and microblogging (Park

<sup>1</sup>The data can be obtained from <https://corpus.korean.go.kr/request/reasetMain.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](https://ko-nlp.github.io/Korpora/en-docs/corpuslist/modu_messenger.html)

<sup>2</sup>Available in [https://github.com/lovit/petitions\\_dataset](https://github.com/lovit/petitions_dataset)

[2022], hereafter vlogging corpus). They characteristically reveal differences between locutionary and illocutionary forces from conventional spoken languages; for instance, Park [2022] claimed that ‘-e ju-’ (to give) has recently been used among Korean language users as a predicate to describe the behavior of the speaker her/himself, dominantly in the context of vlogging and microblogging. 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.

Above three types of corpora were all annotated under the same criteria of Section 2.2. The vlogging corpus was annotated manually by three Korean computational linguists and the final label was determined via majority voting. However, to cover the large-scale data, the messenger and the petition corpus were annotated by workers of a crowdsourcing company<sup>3</sup>, with the worker education conducted by the moderators. In detail, the moderators recruited a total of ten workers from the distinguished worker pool (who are all L1 Korean speakers and biological gender/age evenly distributed) and educated them with a polished guideline with example sentences<sup>4</sup>. After the pilot annotation of five days using 20% of the dataset (5,000 messenger sentences, 1,400 petition sentences) and the feedback phase, the main annotation lasted for ten days. The pilot annotation was conducted with four annotators per sentence (five including the original annotation of the first author) and the main annotation with two annotators per sentence (three including the original annotation of the first author). The final label was determined with majority voting, referring the author’s annotation as priority one in the case of conflict. For the messenger, petition, and vlogging corpus, we obtained Fleiss’ Kappa [Fleiss, 1971] of 0.561, 0.376, 0.665 for three annotators tagging, respectively, which implies a moderate agreement for messenger and vlogging corpus, and slightly lower one for the petition. This difference is assumed to be dominantly influenced by the oratorical style of the petitions, while other two share the similar daily conversation style albeit the vlogging corpus is monologue instead of dialogue. As a summary, we add the statistics of the annotated corpora (Table 2).

### III EXPERIMENT

#### 3.1 Implementation

Since this study aims at observing the transferring ability of fine-tuned language models regarding domains with different speech act tendency, we trained a PLM using the messenger corpus and validate the performance with the same test set, the petition corpus, and the vlogging corpus. For model, we adopted the bert-base-multilingual-cased model [Devlin et al., 2018] that utilized mainly Wikipedia and book corpus for pre-training<sup>5</sup>.

We adopted a learning method conventionally used for PLM downstream tasks provided by Huggingface Transformers library [Wolf et al., 2020]. The batch size used in the training was set to 32, with AdamW optimizer [Loshchilov and Hutter, 2017], weight decay 0.01, and warm-up steps 500.

---

<sup>3</sup><https://deepnatural.ai/>

<sup>4</sup><https://deepnatural.notion.site/v1-5-91dcfde367174a029acea43188a02818>

<sup>5</sup>Other Korean PLMs tend to utilize petition corpus in the pretraining, which may harm the fairness of the evaluation.

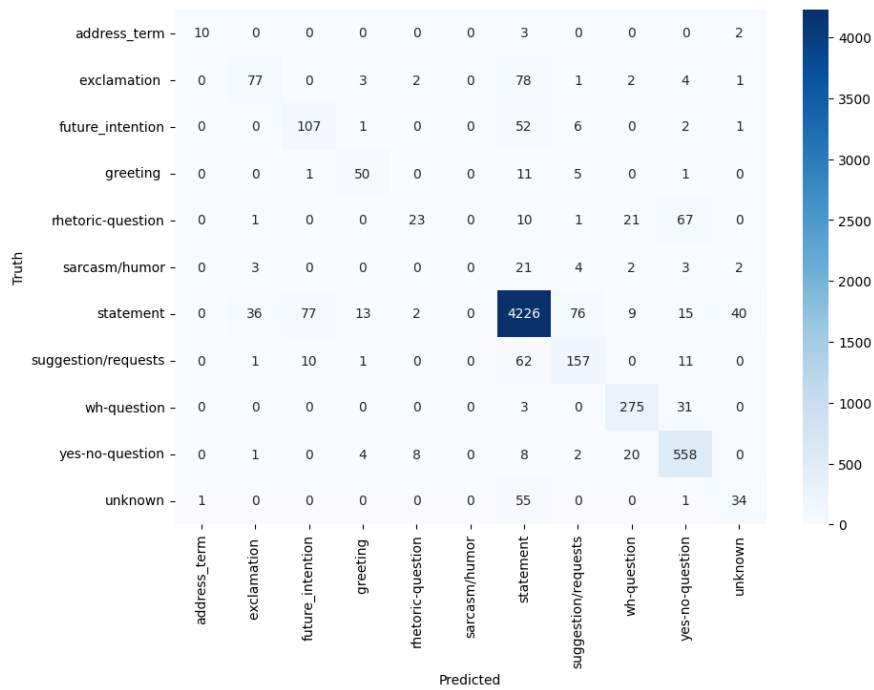


Figure 1: Confusion matrix on the Messenger Corpus.

The training set consists of 25,611 instances (80%) of Messenger Corpus, while there are a total of three test sets with 1) 6,408 messenger instances (20%), 2) 6,400 petition titles, and 3) 48 excerpts from vlogging video transcript. Accuracy and Macro F1 scores were used as evaluation metrics.

### 3.2 Result

The classification accuracy for the messenger corpus was 87.36, with the F1 score 0.5994 (Table 3). As an in-domain and in-distribution data, the model inference showed high accuracy especially for dominant categories such as questions and statement.

To verify whether the trained model adapts to comparably unseen expressions, a test was conducted using 6,408 public petition titles and 48 vlogging script excerpts. For the petition title, We obtained the accuracy of 49.35 and F1 score 0.2510, and for vlogging scripts, the accuracy 60.42 and F1 score 0.4544. The result implies that the model performance significantly differs from the validation with homogeneous dataset, especially in the petition title. It displays the discrepancy that comes from the domain difference of both types of sets, which includes the difference in the distribution of speech act types and the text style.

	<b>Messenger</b>	<b>Petition</b>	<b>Vlogging</b>
<i># Test</i>	6,408	6,408	125
<i>Accuracy</i>	87.36	49.35	60.42
<i>F1</i>	0.5994	0.2510	0.4544

Table 3: Speech act classification evaluation on the three test sets of different domain (trained on the messenger dataset).

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

Table 4: Petition examples where the model prediction differs from the human annotation.

## IV ANALYSIS

### 4.1 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). Overall, due to the high frequency of *statements*, the frequency of prediction errors was high, with the model often mispredicting *exclamations* as *statements* (78) or misclassifying *statements* as *future intentions* (77). Conversely, while not a high-frequency speech act, *rhetorical questions* demonstrated their trickiness, with the model misclassifying them as *yes-no questions* in 67 cases, a relatively high error rate compared to the cases it correctly identified.

- (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.*)  
 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 가식쟁이다 ㅋㅋㅋ (*Wow, what a hypocrite.*)  
 Speaker 2: 어찌라고 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.

### 4.2 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 4). 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 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 ①넣어 줍니다. ... 돼지고기 넣고 푹 ②끓여 줄게요. 고기는 목살이에요. (고기가 어느 정도 익은 후에) 먹기 좋은 크기로 ③잘라 줍니다. (각종 양념을 넣는다는 설명) 잘 섞어서 오래 ④끓여 줄게요....

(In a video of cooking kimchi stew)... First, put the vegetables in the pot and then ①add about half a head of kimchi. ... Add pork and ②simmer thoroughly. The meat is pork neck. (After the meat has been cooked to some extent) ③Cut it into bite-sized pieces. (Explaining that various seasonings are added) Mix it well and ④boil 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.

## V 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<sup>6</sup>. 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

<sup>6</sup>Further details of our work can be found online.

[github.com/songys/DAKSA-Domain\\_Adaptation\\_in\\_Korean\\_Speech\\_Act](https://github.com/songys/DAKSA-Domain_Adaptation_in_Korean_Speech_Act)



is public or not and having multiple or anonymous addressee.

## ACKNOWLEDGMENTS

We thank anonymous reviewers for constructive comments. Authors are also grateful for Eyoung Ko and Soyoung Jeon (of DeepNatural AI), and Jisu Shim for helping the annotation process.

## References

- J. L. Austin. *How to do Things with Words*. Oxford University Press, New York, 1962. 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. Towards an ISO standard for dialogue act annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), 2010.
- Won Ik Cho and Nam Soo Kim. 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, 2022.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.
- Jonathan Ginzburg, Raquel Fernández, and David Schlangen. Self-addressed questions in disfluencies. In *DiSS 2013: The 6th Workshop on Disfluency in Spontaneous Speech*, 2013.
- Aikaterini-Lida Kalouli, Rebecca Kehlbeck, Rita Sevastjanova, Oliver Deussen, Daniel A. Keim, and Miriam Butt. Is that really a question? going beyond factoid questions in nlp. In *International Conference on Computational Semantics*, 2021. URL <https://api.semanticscholar.org/CorpusID:239992804>.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- NIKL. Nikl corpora 2020 (v.1.0), 2020. URL <https://corpus.korean.go.kr>.
- Jinho Park. Yes-no-questions and rhetorical questions. *Hangughagbo*, pages 1–24, 2019.
- Mi-eun Park. About the ‘-e ju-’ construction of that i use to myself –focused on vlog register–. *Korean Semantics*, 78:89–117, 2022.
- Sharoda A. Paul, Lichan Hong, and Ed H. Chi. What is a question? crowdsourcing tweet categorization. In *CHI 2011, Workshop on Crowdsourcing and Human Computation*, 2011. URL <https://api.semanticscholar.org/CorpusID:7085566>.
- John R Searle. A classification of illocutionary acts1. *Language in society*, 5(1):1–23, 1976.
- Sanghoun Song. Pragmatic usage of wh-elements in korean. *Language Information*, 11:91–113, 2010.
- Youngsook Song. *Enhancing AI's Commonsense Reasoning in Conversations through Natural Language Generation*. PhD thesis, University of Kyunghee, 2023. URL [https://www.riss.kr/search/detail/DetailView.do?p\\_mat\\_type=be54d9b8bc7cdb09&control\\_no=7e5dea0eb8aff1bffffe0bdc3ef48d419&keyword=%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5%EC%9D%98%20%EC%9D%BC%EC%83%81%20%EB%8C%80%ED%99%94%20%EC%B6%94%EB%A1%A0%20%EB%8A%A5%EB%A0%A5%20%ED%96%A5%EC%83%81%EC%9D%84%20%EC%9C%84%ED%95%9C%20%EC%9E%90%EC%97%B0%EC%96%B4%20%EC%83%9D%EC%84%B1%20%EC%97%B0%EA%B5%AC](https://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=7e5dea0eb8aff1bffffe0bdc3ef48d419&keyword=%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5%EC%9D%98%20%EC%9D%BC%EC%83%81%20%EB%8C%80%ED%99%94%20%EC%B6%94%EB%A1%A0%20%EB%8A%A5%EB%A0%A5%20%ED%96%A5%EC%83%81%EC%9D%84%20%EC%9C%84%ED%95%9C%20%EC%9E%90%EC%97%B0%EC%96%B4%20%EC%83%9D%EC%84%B1%20%EC%97%B0%EA%B5%AC).
- Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373, 2000.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45, 2020.