PROSODY-SEMANTICS INTERFACE IN SEOUL KOREAN: CORPUS FOR A DISAMBIGUATION OF *WH-* INTERVENTION

Won Ik Cho*†, Jeonghwa Cho*°, Jeemin Kang°, Nam Soo Kim†

Department of Electrical and Computer Engineering and INMC†
Department of English Language and Literature°
Seoul National University
wicho@hi.snu.ac.kr, {jeong9793, bling1104, nkim}@snu.ac.kr

ABSTRACT

In this paper, we construct a corpus that incorporates the recordings of Seoul Korean speech where the scripts are manually generated. The utterances contain *wh-* particles that make the sentences interrogative but sometimes perform as a quantifier. The phonetic property of the corpus concerns not only the sentence-final intonation that differentiates questions from statements, but also the overall prosodies which point out the topic and possibly yield a rhetoricalness. The scripts were generated considering the diversity in the predicate, evidentiality, sentence enders, and the particles regarding politeness. At least two prosodic contours are conveyed from a single script, and it makes the corpus suitable for a multimodal spoken language understanding in the viewpoint of computational linguistics.

Keywords: prosody-semantics, Seoul Korean, corpus, disambiguation, *wh-* particles

1. INTRODUCTION

In analyzing speech, phonetic properties of an utterance are deeply related to semantics. Korean is one of the languages where the sentence meaning highly relies on its prosody, as can be observed in sentence (1) where NOM denotes a nominative case and USE an underspecified sentence ender:

(1) 뭐가 먹고 싶어 **mwe-ka mek-ko siph-e**
what-NOM eat-to want-USE

The sentence type of the utterance is decided upon the sentence-final intonation; an interrogative for a rising tone and a declarative for a falling one. In the viewpoint of speech act, mainly four intentions can be inferred from the given text, namely:

(1a) *What do you want to eat?* (***wh-* Q**)
(1b) *Do you want to eat something?* (**yes/no Q**)
(1c) *Do you really want to eat something?* (**rhetorical Q**)
(1d) *I want to eat something.* (**statement**)

Following the L/M/H marking [10] where each denotes low/middle/high pitch and a single '=' is used to denote the repetition of the previous syllabic pitch, (1a) is expressed by the syllabic intonation contour of LHL==H% or LHML==% (falling), (1b) by L==MLH%, (1c) by LML=HH%, and (1d) by LMLHL=%. To be specific, in (1a) the highest pitch is assigned to the particle '가 (ka)' which is a postposition of *mwe* (*what*) phrase, and the main function of the utterance is to ask and addressee what to eat. In (1b) the high pitch is assigned to the particles '고 (ko)' and '어 (e)' which are respectively the head regarding *eat* and *want*, and here the speaker asks whether the addressee wants to eat something. In (1c), the high pitch is assigned to '싶어 (siph-e)', making the utterance a rhetorical question, where the speaker expresses a negative feeling towards the addressee. In (1d), the high pitch is assigned only to '고 (ko)', conveying that the speaker wants to eat something. This phenomenon fundamentally originates in the *wh-* intervention of '뭐 (mwe)', which corresponds with *what* in English but also can be interpreted as *something* in particular circumstance, making *mwe* an *in-situ* to denote *something edible*.

Despite this variability in prosody-semantics that can emerge from a single utterance, conventional speech corpora usually do not incorporate all the speech act types as in (1a-d). This is the point where we came out with the necessity of a corpus that incorporates the *single text* to *multi prosody/intention* utterances. In this study, a corpus containing the recordings of manually generated Seoul Korean scripts is constructed. The sentences start with *wh-* particles and incorporate a wide variety of predicates and sentence enders. Based on the utterances, all possible prosodic contours are recorded, and the corresponding intention types are annotated with English translation. Note that the utterances were constructed in a simple structure with minimal contextual engagement. Since non-domain-specific words are utilized, the corpus fits with a variety of real-life language analysis that requires disambiguation.

2. RELATED WORK

Studies on the utterances incorporating *in-situ wh-*particles have been done widely in the areas of syntax [16, 1, 9, 13] and prosody-semantics interface [6, 2]. Especially for Korean and Japanese, which are typical *wh-in-situ* languages, the variability of the *wh-* particles was handled within the topic of LF intervention [15]. Also, in [8], it was proposed that a *wh-* particle in embedded self-addressed questions be interpreted as an existential quantifier.

In a slightly different view, in [5], the *wh-* particles associated with negative polarity items were investigated, suggesting the circumstances where the intervention is canceled. In [14], the usage of *wh-*particles as an interrogative and indefinite NP is investigated in a pragmatic view, accompanying the interpretation of gray-zone cases as rhetorical ones. Another thing to note is that it suggests that '왜 (way, *why*)' be interpreted as an exclamation. Taking this into account, we did not generate sentences that include *way*.

In the view of language acquisition, [4] found that L2 Korean learners have difficulties with interpreting *in-situ wh-* particles. This implies the necessity of disambiguation that incorporates syntax, semantics, and phonetics, to which this paper attempts to contribute via a corpus-based approach.

3. CORPUS GENERATION

In generating the corpus script, namely five factors were considered: *wh- particles* that initiate an utterance, *predicates* that convey the content, *reportive particles* that give the utterance evidentiality, *sentence enders* that possess potential to represent various intentions, and *politeness suffixes* which come just after the sentence ender to assign honorific mood to the sentence.

3.1. *wh-* particles

Among the six *wh-* particles, namely '누구 (nwukwu, *who*)', '뭐 (mwe, *what*)', '어디 (eti, *where*)', '언제 (encey, *when*)', '어떻게 (ettehkey, *how*)', and '왜 (way, *why*)', only the first five were utilized in constructing the corpus. This is because *way* is rarely used as a quantifier, except for some cases in child language. Instead of *way*, we used '몇 (meych, *the number of*)', which is widely used as a quantifier for counting. For the purpose of variation, in some cases, nominative (NOM) or accusative cases (ACC) were attached to the *wh-* particles.

3.2. Predicates

Predicates largely depend on the *wh-* particle they are aligned with. For instance, *nwukwu (who)* harmonizes with the verbs that are related to interaction, such as *give* and *receive*. In contrast, *eti (where)* matches with the verbs concerning location, such as *come* and *go*. In selecting the verbs, we referred to the set of 5,800 frequently used lexicons, released by the National Institute of Korean Language (https://www.korean.go.kr/). Depending on the verbs, appropriate particles were agglutinated and the phrases that contain object/complement were inserted. In some circumstances, polarity items such as '좀 (com, *bit*)' or '하나 (hana, *a piece*)' were augmented to modify or restrict the implicature.

3.3. Reportive particles

The reportive particles (RPT) provide utterances with evidential mood. Usually '-대 (tay)', '-래 (lay)', and '-쟤 (cyay)' are used for statements, commands, and hortatives [11]. The particles were selectively added considering the content.

3.4. Sentence enders

The sentence enders (SEs) with various roles are components that influence the sentence type and intention of the utterance. There are mainly two types of SEs; the first type is SEs with a fixed role, e.g., '-다 (ta)' for declaratives and '-니 (ni)' for interrogatives [11]. For these, the sentence type is fixed but the intention can vary regarding *wh-* intervention and rhetoricalness. The second type is the underspecified SEs whose feature is not fixed (e.g., '어 (e)', '지 (ci)'). They have the potential to display various intention types depending on the prosody. Both types of SEs were utilized in the generation.

3.5. Politeness suffix

The politeness suffix (POL), '요 (yo)', can be agglutinated to SEs and in most cases does not affect the functional variability of the sentence, except for rhetoricalness. For some SEs such as '지 (ci)' or '야지 (yaci)', the augmented form is modified to '죠 (cyo)'. On the other hand, the utterances with SEs to which the politeness suffix is not attachable, such as '냐 (nya)', were left without the politeness suffix. An example sentence incorporating the aforementioned concepts (3.1-5) is as follows:

(2) **뭐 좀 먹었대요**	**mwe com mek-ess-tay-yo**
(*statement* or *y/n Q*)	what bit eat-PST-RPT-POL

4. TAGGING INTENTIONS

The substantial feature of this study lies in the annotation of the intentions along with the script and speech. The labels used for the annotation are *statement*, *yes/no question*, *wh- question*, *rhetorical question*, *command*, *request*, and *rhetorical command*, a modified version of the categorization recently suggested in [3].

- **Statement (S)** indicates an utterance that conveys information or the speaker's thought.
- **Yes/no question (YN)** indicates a question where the answer set is limited to yes or no.
- ***Wh*-question (WH)** indicates a question where the answer set is open and variable.
- **Rhetorical question (RQ)** indicates a question whose answer set is in the speaker's mind, usually being adopted to express the thought.
- **Command (C)** incorporates an *order* that corresponds to imperatives in English with a covert subject, *hortative* that indicates an order with a politeness particle (e.g., *please*), and *modal* that indicates a statement with particles which correspond with *should* or *must*.
- **Request (R)** indicates a command expressed in an interrogative form.
- **Rhetorical command (RC)** indicates a command where the to-do-list is not mandatory, usually used as an idiomatic expression.

We list some examples regarding several *wh*-particles, incorporating more than three intention (and prosody) types. The case for *mwe (what)* is explained in the previous section, and the case for *ettehkey (how)* is omitted in this paper since the intention variability is small (two cases at most). *Q* denotes question and *C* denotes command. L, M, H and '=' denote the relative pitches.

(3) 누가 보러 간대 **nwu-ka po-le kan-tay**
who-NOM see-to go-RPT
(3a) *Who will go see it?*
(LHL==H%; ***wh*-Q**)
(3b) *Will sbd go see it?*
(LML==H%; **yes/no Q**)
(3c) *Does anyone say I'm gonna go see it?*
(LMLMLH%; **rhetorical Q**)
(3d) *I heard sbd will go see it.*
(L==HL=%; **statement**)

(4) 어디 가고 싶어 **e-ti ka-ko siph-e**
where go-to want-USE
(4a) *Where do you want to go?*
(LHL==H%; ***wh*-Q**)
(4b) *Do you want to go somewhere?*
(L==MLH%; **yes/no Q**)

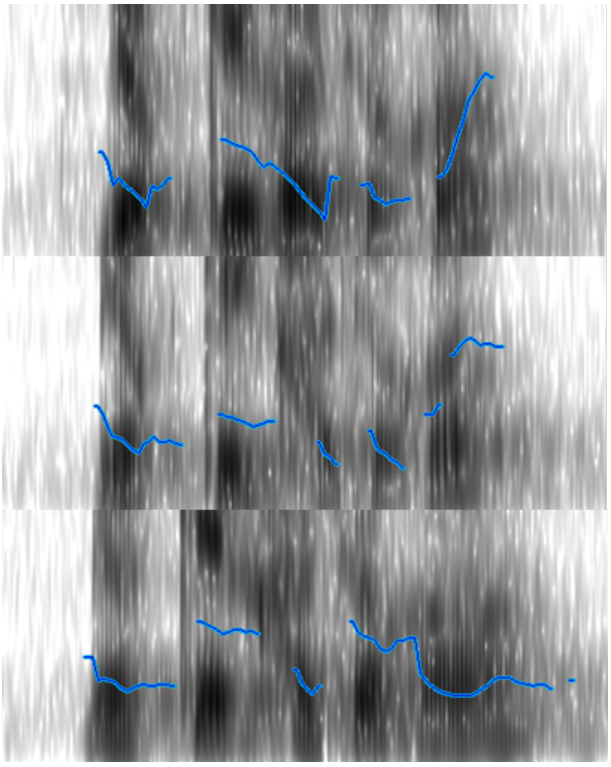

Figure 1: The F0 contour of the guiding voice for 6(a-c) by Praat.

(4c) *I want to go somewhere.*
(L==HL=%; **statement**)

(5) 언제 다시 봐 (보-아) **en-cey ta-si pwa (pw-a)**
when again meet-USE
(5a) *When will we meet again?*
(LHL=H%; ***wh*-Q**)
(5b) *Shall we meet again someday?*
(LML=H%; **yes/no Q**)
(5c) *Let's meet again someday.*
(LMLML%; **rhetorical C**)

(6) 몇 개 가져가 **myech kay ka-cye-ka**
how quantity bring-USE
(6a) *How many shall I take?*
(LHL=H%; ***wh*-Q**)
(6b) *Shall I take some?*
(LML=H%; **yes/no Q**)
(6c) *Take some.*
(LMLML%; **command**)

To aid comprehension, the F0 contours of the guiding voice for 6(a-c) are presented in Figure 1. Note that the relative pitch sequence is displayed in the spectrogram. The sentence-final intonation of (6a,b) implies that they are questions, and the relative pitch of *kay* which follows *myech* distinguishes (6a) *wh-Q* from (6b) *yes/no Q*.

Who	1,895
What	877
Where	199
When	172
How	163
How much	246

<Table 1-a >

Statement	1,085
Yes/no Q	1,047
Wh- Q	849
Rhetorical Q	302
Commands	175
Requests	56
Rhetorical C	38

<Table 1-b >

S, YN, WH	424
S, YN	242
S, YN, WH, RQ	137
S, WH	74
S, YN, RQ	43
YN, WH	33
...	
S, RQ	25
YN, WH, C	25
S, R	24
S, WH, RQ	24
S, YN, WH, C	23
YN, C	23
WH, R	22

··· (total 32 cases)

<Table 1-c >

Table 1: (1-a) describes the statistics on *wh-* particle and (1-b) on the intention types. (1-c) describes some of the possible intention sets that are engaged in a single utterance, here for 1,292 sentences. The code is disclosed with the corpus.

5. DISCUSSION

5.1. Corpus specification

In the corpus construction, the first version of the sentence list was generated by the methodology explained in Section 3, and only the sentences that received the consensus of first three authors (native speakers of Seoul Korean dialect) were taken into account. In total, the corpus contains 3,552 utterances that fall into the seven classes of intention. All the utterances were recorded by two native Koreans, a male and a female. The speech corpus containing a total of 7,104 (= 3,552 * 2) utterances are available on-line (https://www.github.com/warnikchow/prosem) as with the corpus.

The statistics on the corpus is presented in Table 1. Sorting by the *wh-* particles that initiate the sentences, the most were the sentences starting with '누구 (nwugu, *who*)', and the least were the ones starting with '어떻게 (ettehkey, *how*)'. Sorting by the intentions, *S* accounted for the most of the corpus data and *RC* did the least, following the tendency displayed in the recent Korean corpus [3]. Sorting by the possible intentions from a single utterance, the number of cases ranged from 2 to 4. The possible cases are partially listed in Table 1-c, tagged with the quantity. *Wh-* intervention occurs in most cases where interpreting the particles as *wh-* is allowed, but not vice versa.

	S	***YN***	***WH***	***RQ***	***C***	***R***	***RC***
Who	547	544	446	202	112	26	18
What	294	283	186	64	32	14	4
Where	64	64	49	6	11	4	1
When	37	54	40	22	0	4	15
How	59	62	28	8	6	0	0
How much	84	40	100	0	14	8	0

Table 2: A frequency matrix on *wh-* particles and the intention types.

5.2. Analysis

For a more detailed analysis of the corpus, we performed a simple calculation that shows the correlation between the indices (Table 2). It is assumed that *wh-* intervention largely occurs among *how much*-sentences, considering the portion of *wh-* questions within. Also, commands starting with *how* are rare, due to the fact that a to-do-list [12] is usually recommended to convey a specific instruction.

Concerning rhetoricalness, it is notable that *how much* is scarce among the rhetorical sentences since in that case polarity items are accompanied, disambiguating the *wh-* intervention. Consequently, the non-rhetorical directives (*YN*·*WH* and *C*·*R*) dominate *how much* sentences. The portion of rhetorical directives scored the highest in *when-* sentences for both *RQ*/*C*; we roughly assume that *when-* questions have potential to be interpreted as '*have sbd ever ...*' and *when-* commands usually act idiomatically.

6. CONCLUSION

In this paper, we proposed a construction scheme for a corpus that incorporates single-text utterances with multi prosody/intention. In the process, five constituents, namely *wh-* particles, predicates, evidentiality, sentence enders and politeness suffix were considered. We obtained 3,552 utterances from 1,292 sentences, with the major intention types of *statement*, *yes/no question* and *wh- question*. A set of recordings by the native speakers (total 7,104 instances) is disclosed as a pilot research and can be supplemented for industrial purpose.

We suggest the corpus to be used for spoken language understanding systems which require disambiguation of the utterances that may induce *wh-* intervention. Along with the multi-modal approaches as in [7], various statistics- or deep learning-based classification systems may be able to infer a proper intention for a given speech and transcript. Not only for the industry, but this study can also be utilized for the Korean language learners, especially for the acquisition of prosody.

ACKNOWLEDGEMENT

This research was supported by Projects for Research and Development of Police science and Technology under Center for Research and Development of Police science and Technology and Korean National Police Agency funded by the Ministry of Science, ICT and Future Planning (PA-J000001-2017-101). Also, this work was supported by the Technology Innovation Program (10076583, Development of free-running speech recognition technologies for embedded robot system) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea). After all, the authors appreciate the helpful advices provided by Mijeong Song and Minhwa Chung.

7. REFERENCES

[1] Aoun, J., Li, Y.-h. A. 1993. Wh-elements in situ: Syntax or lf? *Linguistic Inquiry* 24(2), 199–238.

[2] Baunaz, L. 2005. The syntax and semantics of wh in-situ and existentials: the case of french. *Leiden Papers in Linguistics* 2(2), 1.

[3] Cho, W. I., Lee, H. S., Yoon, J. W., Kim, S. M., Kim, N. S. 2018. Speech intention understanding in a head-final language: A disambiguation utilizing intonation-dependency. *arXiv preprint arXiv:1811.04231*.

[4] Choi, M. H. 2009. *The acquisition of wh-in-situ constructions in second language acquisition*. PhD thesis Georgetown University.

[5] Choi, Y.-S. 2007. Intervention effect in korean wh-questions: Indefinite and beyond. *Lingua* 117(12), 2055–2076.

[6] Dalrymple, M. Semantics, information structure, and prosody in lfg part i: Semantics in lfg part ii: Information structure in lfg part iii: Prosody, syntax and semantics.

[7] Gu, Y., Li, X., Chen, S., Zhang, J., Marsic, I. 2017. Speech intention classification with multimodal deep learning. *Canadian Conference on Artificial Intelligence*. Springer 260–271.

[8] Jang, Y. 1999. Two types of question and existential quantification. *Linguistics* 37(5), 847–869.

[9] Lai-Shen Cheng, L., Rooryck, J. 2000. Licensing wh-in-situ. *Syntax* 3(1), 1–19.

[10] Lee, H. Y. 1999. An acoustic phonetic study of korean nuclear tones. *Journal of the phonetic society of Korea* 38, 25–39.

[11] Pak, M. D. 2008. Types of clauses and sentence end particles in korean. *Korean Linguistics* 14(1), 113–156.

[12] Portner, P. 2004. The semantics of imperatives within a theory of clause types. *Semantics and linguistic theory* volume 14 235–252.

[13] Soh, H. L. 2005. Wh-in-situ in mandarin chinese. *Linguistic Inquiry* 36(1), 143–155.

[14] Song, S. 2010. Pragmatic usage of wh-elements in korea. *Language Informatio* 10, 95–117.

[15] Tomioka, S. 2007. Pragmatics of lf intervention effects: Japanese and korean wh-interrogatives. *Journal of Pragmatics* 39(9), 1570–1590.

[16] Watanabe, A. 1992. Subjacency and s-structure movement of wh-in-situ. *Journal of east Asian linguistics* 1(3), 255–291.