2021 Winter LSK Young Scholar Symposium and Winter Conference Handbook

Dates: December 9 (Thu) - December 11 (Sat), 2021
Venue: Zoom
Host & Organizer: The Linguistic Society of Korea
Sponsor: National Research Foundation of Korea
General Inquiries: 2021youngscholarlsk@gmail.com

한국언어학회
The Linguistic Society of Korea

Symposium/Conference Program and Organizing Committee

President

Minhaeng Lee (Yonsei Univ.)

Program committee chair

Jong-Bok Kim (Kyung Hee Univ.)

Program committee members

Kyungchul Chang (Pusan National Univ.), Taegoo Chung (Korea Univ.)
Se-Eun Jhang (Korea Maritime and Ocean Univ.), Heejeong Ko (Seoul National Univ.)
Iksoo Kwon (Hankook Univ. of Foreign Studies), Chung-Hoon Lee (Sogang Univ.)
Hanjung Lee (Sungkynkwan Univ.), Hye-Kyung Lee (Ajou Univ.)
Jungmee Lee (Seoul National Univ.), Mira Oh (Chonnam National Univ.)
Bum-Sik Park (Dongguk Univ.), Hyopil Shin (Seoul National Univ.)
Jeong-Ah Shin (Dongguk Univ.), Gwangrak Son (Kyungpook National Univ.)

Symposium organizing committee chair

Sanghoun Song (Korea Univ.)

Conference organizing committee chair

Hanjung Lee (Sungkynkwan Univ.)

Organizing committee members

Keeseok Cho (Cyber Hankuk Univ. of Foreign Studies), Youngju Choi (Chosun Univ.)
Seungwan Ha (Kyungpook National Univ.), Myupyo Hong (Sungkyunkwan Univ.)
Seungshik Kang (Kookmin Univ.), Ilkyu Kim (Kangwon National Univ.)
Rhanghyeyun Kim (Korea Univ.), Sang-Geun Lee (Korea Univ.)
Hyehyun Nam (Yonsei Univ.), Yunju Nam (Konkuk Univ.)
Soyoung Park (Pusan National Univ.), Seok-Chae Rhee (Yonsei Univ.)
Hong Joon Um (Keimyung Univ.)

Contents

1. Symposium and Conference Program

2021 Winter LSK Young Scholar Symposium and Winter Conference

Dates: December 9-11, 2021
Online Venue: Zoom (hosted by the Linguistic Society of Korea)

Young Scholar Symposium Day 1 (December 9, Thursday)

09:20-09:30	**Opening Ceremony** Welcome Speech Minhaeng Lee (President, Yonsei University) (https://khu-ac.zoom.us/j/3380168854)	
09:30-10:30	**Invited Talk 1** Title: Statistical analyses of learner corpus data: Pitfalls and recommendations Speaker: Stefan Th. Gries (University of California, Santa Barbara & JLU Giessen) Moderator: Hye-Won Choi (Ewha Womans University) (https://khu-ac.zoom.us/j/3380168854)	
10:30-10:40	**Break**	
	Session 1 (Corpus Linguistics) Moderator: Rhanghyeyun Kim (Korea University) Commentator: Stefan Th. Gries (University of California, Santa Barbara & JLU Giessen) (https://khu-ac.zoom.us/j/3380168854)	**Session 2 (Syntax, Deep Learning)** Moderator: Sang-Hee Park (Duksung Women's University) Commentator: Rui P. Chaves (University at Buffalo) (https://khu-ac.zoom.us/j/6472196084)
10:40-11:10	Language use between men and women: A multifactorial approach Ji-Hye Kim & Yong-hun Lee (Hannam University & Chungnam National University)	Semantic role labeling: A layer-wise analysis of KRBERT Hye-Jin Seo (Dongguk University)
11:10-11:40	Exceptive constructions: A corpus-based perspective Geonhee Lee (Kyung Hee University)	How does BERT address polysemy of Korean adverbial postpositions *-ey, -eyse*, and *-(u)lo*? Seongmin Mun (Chosun University)
11:40-12:10	Comparing Korean and the U.S. newspaper articles on Covid-19: A corpus-based analysis Junghyun Lee & Sanghoun Song (Korea University)	Syntactic priming by LSTM L1/L2 language models Sunjoo Choi (Dongguk University)
12:10-14:30	**Lunch**	
14:30-15:30	**Invited Talk 2** Title: Constructional change and distributional semantics Speaker: Martin Hilpert (Université de Neuchâtel) Moderator: Jungsoo Kim (Kyung Hee University) (https://khu-ac.zoom.us/j/3380168854)	
15:30-15:40	**Break**	
	Session 3 (Cognitive Linguistics, Construction Grammar, Language Variation and Change) Moderator: Iksoo Kwon (Hankuk University of Foreign Studies) Commentator: Martin Hilpert (Université de Neuchâtel) (https://khu-ac.zoom.us/j/3380168854)	**Session 4 (Language Processing, Semantics, Pragmatics)** Moderator: Soyeon Yoon (Incheon National University) Commentator: Stefan Engelberg (Leibniz-Institut für Deutsche Sprache, Mannheim) (https://khu-ac.zoom.us/j/6472196084)
15:40-16:10	*Carefully does it, reckless does not*: A construction grammar approach to the *X does it* construction Hanbeom Jung (Hankuk University of Foreign Studies)	L1 and L2 processing of lexical bundles Hyojin Jeong (Sungkyunkwan University)
16:10-16:40	*Much less* construction in English: A direct interpretation approach Jiseop Hyeon (Kyung Hee University)	The electro-physiological evidence on the processing of the Korean conversational implicature Shin-ae Yoon, Hyenyeong Chung, & Yunju Nam (Konkuk University)
16:40-17:10	"Call me by my name": Names, address, and identities of Korean women Eldin Milak (Sungkyunkwan University)	Teachers' questions repairing in classroom Songdan Piao (Sungkyunkwan University)

Young Scholar Symposium Day 2 (December 10, Friday)

09:30-10:30	**Invited Talk 3** Title: Just what do you think you're doing, BERT? Speaker: Rui P. Chaves (University at Buffalo) Moderator: Myung-Kwan Park (Dongguk University) (https://khu-ac.zoom.us/j/3380168854)	
10:30-10:40	**Break**	
	Session 5 (Corpus Linguistics) Moderator: Eunkyung Yi (Ewha Womans University) Commentator: Stefan Th. Gries (University of California, Santa Barbara & JLU Giessen) (https://khu-ac.zoom.us/j/3380168854)	**Session 6 (Syntax, Deep Learning)** Moderator: Jungmee Lee (Seoul National University) Commentator: Rui P. Chaves (University at Buffalo) (https://khu-ac.zoom.us/j/6472196084)
10:40-11:10	Use of English intensifiers by gender and text genre Yoon-Young Jeon & Eugene Chung (Korea University)	How can 'what'-questions be interpreted as 'why'-questions? Okgi Kim (University of Wisconsin-Milwaukee)
11:10-11:40	Ditransitive construction in English: A corpus-based study Jennifer Pearl (Kyung Hee University)	Processing the dynamicity of events in language Sarah Hye-yeon Lee & Elsi Kaiser (University of Southern California)
11:40-12:10	Korean EFL learners' underuse of the English perfect forms Kang San Noh & Sanghoun Song (Korea University)	The dative puzzle in Chinese revisited Hyunjun Park (Dongguk University)
12:10-14:30	**Lunch**	
13:40-15:30	**Invited Talk 4** Title: Metaphors: Morphosyntax and pattern building Speaker: Stefan Engelberg (Leibniz-Institut für Deutsche Sprache, Mannheim) Moderator: Eugene Chung (Korea University) (https://khu-ac.zoom.us/j/3380168854)	
15:30-15:40	**Break**	
	Session 7 (Cognitive Linguistics, Construction Grammar, Language Variation and Change) Moderator: Youngju Choi (Chosun University) Commentator: Martin Hilpert (Université de Neuchâtel) (https://khu-ac.zoom.us/j/3380168854)	**Session 8 (Language Processing, Semantics, Pragmatics)** Moderator: Hongoak Yun (Jeju National University) Commentator: Stefan Engelberg (Leibniz-Institut für Deutsche Sprache, Mannheim) (https://khu-ac.zoom.us/j/6472196084)
15:40-16:10	Two types of semi-modal V-TO in English Lehan Yang (Pusan National University)	Progressives with dense and discrete interval verbs in English Minji Kang (Kyungpook National University)
16:10-16:40	A comparative analysis of certain double object constructions in Spanish and Korean Nakyung Yoon (Korea University)	Projecting presupposition "not again" in Korean Kihyo Park & Yunju Nam (Konkuk University)
16:40-17:10	Emergent constructional knowledge of the masked language model: A corpus-based approach to do-be construction Un-sub Shin & Sanghoun Song (Korea University)	Forward-looking effects in pronoun interpretation Jina Song & Elsi Kaiser (University of Southern California)
17:10-17:20	**Closing Remarks** Jong-Bok Kim (President-elect & Program committee chair, Kyung Hee University) (https://khu-ac.zoom.us/j/3380168854)	

Winter Conference (December 11, Saturday)

Theme: Linguistics in the Age of Transition: New Directions in Theory and Application

ZOOM link: https://korea-ac-kr.zoom.us/j/81698786130?pwd=dENwdmdqUkZVekVJTlFEYnZVY0w5QT09

09:20-09:25	**Opening Remarks** Hanjung Lee (Organizing committee chair, Sungkyunkwan University)
09:30-10:30	**Plenary Talk 1 [in English]** Title: Explain me this: Choosing good-enough constructions Speaker: Adele Goldberg (Princeton University) Moderator: Hye-Won Choi (Ewha Womans University)
10:30-10:40	**Break**
10:40-11:40	**Invited Talk 1 [in Korean]** Title: A marcopragmatic approach to societal communication Speaker: Sungbom Lee (Sogang University) Moderator: Il-Kyu Kim (Kangwon National University) Discussant: Haeyun Lee (Hankuk University of Foreign Studies)
11:40-12:40	**Lunch**
12:40-13:40	**Invited Talk 2 [in Korean]** Title: New trends and major issues in deep-learning based approaches to language processing Speaker: Seungshik Kang (Kookmin University) Moderator: Hyun-Jung Son (Yonsei University) Discussant: Munpyo Hong (Sungkyunkwan University)
13:40-13:50	**Break**
13:50-14:50	**Invited Talk 3 [in Korean]** Title: How we talk to infants and why it matters: Insights from Korean infants and their caregivers Speaker: Eon-Suk Ko (Chosun University) Moderator: Hae-Il Park (Kyung Hee University) Discussant: Eunhae Oh (Konkuk University)
14:50-15:00	**Break**
15:00-16:00	**Plenary Talk 2 [in English]** Title: The layers of assertion: Propositions, judgements, commitments, acts Speaker: Manfred Krifka (Leibniz-Zentrum Allgemeine Sprachwissenschaft & Humboldt-Universität zu Berlin) Moderator: Minpyo Hong (Myongji University)
16:10-16:40	**LSK General Meeting**
16:40-16:45	**Closing Remarks** Minhaeng Lee (President, Yonsei University)

2. Symposium Information

Sessions with Remarks by Invited Speakers

Session 1 & Session 5
Stefan Th. Gries (University of California, Santa Barbara & JLU Giessen)
Corpus Linguistics

Session 2 & Session 6
Rui P. Chaves (University at Buffalo)
Syntax, Deep Learning

Session 3 & Session 7
Martin Hilpert (Université de Neuchâtel)
Cognitive Linguistics, Construction Grammar, Language Variation and Change

Session 4 & Session 8
Stefen Engelberg (Leibniz-Institut für Deutsche Sprache, Mannheim)
Language Processing, Semantics, Pragmatics

Biography of Invited Speakers

Stefan Th. Gries (University of California, Santa Barbara)

Talk Title: Statistical analyses of learner corpus data: Pitfalls and recommendations

Stefan Th. Gries is (full) professor of linguistics in the Department of Linguistics at the UC, Santa Barbara and Chair of English Linguistics (Corpus Linguistics with a focus on quantitative methods, 25%) at the Justus-Liebig-Universität Giessen. Gries is a quantitative corpus linguist at the intersection of corpus linguistics, cognitive linguistics, and computational linguistics, who uses a variety of different statistical methods to investigate linguistic topics such as syntax, semantics, and corpus-linguistic methodology, as well as first and second/foreign language acquisition.

Rui P. Chaves (University at Buffalo)

Talk Title: Just what do you think you're doing, BERT?

Rui P. Chaves is Associate Professor of Linguistics at the University at Buffalo, the State University of New York. His work focuses on how linguistic knowledge interfaces with cognition, and in particular with probabilistic information that shapes linguistic behavior. He has specialized in formally explicit construction-based models of grammar.

Martin Hilpert (Université de Neuchâtel)

Talk Title: Constructional change and distributional semantics

Martin Hilpert is Professor of English Linguistics at the University of Neuchâtel. He holds a PhD from Rice University. He did postdoctoral research at the International Computer Science Institute in Berkeley and at the Freiburg Institute for Advanced Studies. He is interested in cognitive linguistics, language change, construction grammar, and corpus linguistics.

Stefan Engelberg (Leibniz-Institut für Deutsche Sprache Mannheim)

Talk Title: Metaphors: Morphosyntax and pattern building

Stefan Engelberg is head of the "Department of Lexical Studies" and vice director of the "Leibniz Institute of the German Language". He studied general linguistics, German and Slavic philology at the universities of Frankfurt and Münster (MA thesis on morphology in HPSG, 1990) and did his PhD on "Verbs, Events, and the Lexicon" (1998) and his habilitation on "Lexical and structural aspects of the constitution of sentence meaning" (2005) at the University of Wuppertal.

Abstracts of Invited Talks

Statistical analyses of learner corpus data: Pitfalls and recommendations

Stefan Th. Gries
(UC Santa Barbara & JLU Giessen)

In this talk, I will provide a didactically motivated selective discussion of the timeline of (quantitative) corpus-linguistic approaches in learner corpus research, from early-ish over-/under-use frequencies to the advent of regression-based approaches to new developments using in particular the multi-step MuPDAR(F) procedure. I will discuss a handful of studies with the aims of (i) raising awareness of potential shortcomings and problems of some traditional kinds of analyses and (ii) illustrate the advantages that a stepwise refinement of our quantitative corpus tools can provide.

Constructional change and distributional semantics

Martin Hilpert
(Université de Neuchâtel)

In recent years, many innovations have re-shaped how linguists use corpora to investigate diachronic developments in language. This talk will focus on one such innovation, namely the use of distributional semantic techniques, which try to operationalize meaning in terms of co-occurrence frequencies of linguistic items in corpus data. While distributional techniques have had their place in synchronic corpus linguistics for some time already, applying it to diachronic corpus data comes with its own set of challenges, but also with its own unique opportunities. This talk will explore how distributional techniques can be used for analyses that are framed against the background of Diachronic Construction Grammar. I will present examples and case studies that show how constructional change is reflected in shifting distributional patterns that can be observed in diachronic corpus data, and I will flesh out the theoretical conclusions that can be drawn on the basis of these observations.

Just what do you think you're doing, BERT?

Rui P. Chaves
(University at Buffalo)

Natural language processing is now dominated by large neural language models, and there is some evidence that models like BERT have learned rich syntactic and semantic representations. This presumably explains why such models work so well. In particular, BERT has been claimed to be able to cope with filler-gap constructions (like those observed in wh-questions, topicalization, and relatives) where there are morphosyntactic, semantic and pragmatic dependencies that can span an arbitrary number of clauses. Such constructions are traditionally among the most challenging syntactic phenomena for computational language models, precisely because they involve syntactically unbounded linkages. But are BERT's rich syntactic and semantic representations actually used by BERT during sentence processing, or are they mere artifacts, created by the probing procedure? I will discuss evidence that BERT learns syntactic dependencies only very imperfectly, including trivial subject-verb agreement, despite being trained on massive amounts of data, and I will argue that BERT archives high accuracy on standard test sets because it likely over-relies on linguistically shallow heuristics to cope with complex tasks. Such heuristics happen to yield high accuracy because the test sets contain random idiosyncrasies that neural models opportunistically exploit. In this talk I will highlight the value of linguistically-informed testing of such models, in order to better evaluate them, by using more adversarial evaluation methodologies capable of exposing the fragility of such models.

Metaphors: Morphosyntax and pattern building

Stefan Engelberg
(Leibniz-Institut für Deutsche Sprache, Mannheim)

Metaphors are not the result of unrestrained linguistic creativity, but are dependent of the patterns, rules and restrictions of language and cognition, as metaphor theories of the last decades have shown. However, in most of these theories, metaphors are mainly considered from a purely semantic or general conceptual perspective. In contrast, the approach presented here focusses more on morphosyntactic and lexical structures. The basic ideas are the following: (i) *Feature extraction*: metaphorical readings of an expression arise from interpretive conflicts that are resolved by feature extraction from the source meaning of the expression. (ii) *Local binary structures*: local binary structures (esp. genitive constructions, compounds, adjective-noun combinations) and the head-modifier distribution in these structures play a prominent role in conflict activation and resolution. (iii) *Syntagmatic interpretation patterns*: the different types of local binary structures are associated with syntagmatic interpretation patterns (e.g., explicative genitive, part-whole compound meaning, relational adjective interpretation), which also limit the possible metaphorical readings. (iv) *Paradigmatic metaphorical patterns*: the words within a word field often show tendencies to base their metaphorical interpretation on the extraction of the same feature type, e.g., in body part names, the position and shape of the body part (examples from German): *Lampenfuß*, lamp foot, 'lampstand'; *Bergrücken*, mountain back 'mountain ridge'; *Tischbein*, table leg 'leg of the table'; *Felsnase*, rock nose, 'rock outcrop'). (v) *Bound metaphors*: simplex words usually show tendencies or restrictions in their metaphorical readings with respect to the local binary structures in which they occur; in this respect, they are bound to certain morphosyntactic structures (e.g., *Lachs auf Spinatbett* / **Lachs auf Bett des Spinats* 'salmon on a bed of spinach').

Abstracts of General Talks

Language use between men and women: A multifactorial approach

Ji-Hye Kim and Yong-hun Lee
(Hannam University · Chungnam National University)

It is known that men's language use is different from women's, and there have been a lot of studies on the gender differences in language use (Holmes and Meyrehoff, 2003; Baker, 2014). These studies investigated the differences between men and women, and they examined the differences theoretically or with a corpus-based approach. This study also employed a corpus-based approach, but took a multifactorial approach where various linguistic factors were analyzed simultaneously.

On the other hand, there have been several studies in corpus linguistics which demonstrated there are some differences in the use of *can* and *may*, even within native speakers and between native speakers and non-native speakers (Deshors, 2010; Deshors, 2014; Deshors and Gries, 2014). Those studies manually encoded various linguistic factors and adopted multifactorial approaches to examine how the linguistic factors influenced the choice of *can* and *may*, within native speakers and between native speakers and non-native speakers.

Our study started from these corpus-based and multifactorial analysis methods. We supposed that, if there are some differences in the use of *can* and *may* within native speakers, then men and women might use the modal differently and modals could be used for distinguishing the gender differences. Accordingly, we tried to examine if these two simple modals (*can* and *may*) could also be used to distinguish the gender differences.

For this purpose, we took the BNC64 corpus (Brezina, 2013; 1.6 million word tokens). The corpus contained 64 files (32 files for males and 32 for females), whose data came from the British National Corpus (BNC) and included gender information. Then, we extracted 6,043 sentences which included *can* and *may*. We adopted multifactorial analyses in previous studies (Gries, 2003; Deshors, 2010; Deshors, 2014; Deshors and Gries, 2014) and encoded more than twenty linguistic factors. Then, we statistically analyzed the data with a mixed-effects model, after a multicollinearity analysis.

It was found that several linguistic factors were statistically significant, which implied that men and women used these two modals (*can* and *may*) differently. For example, the person, gender, and animacy type of the subject were statistically significant. As for verbs, mood, transitivity, Vendler's classification were statistically significant. At the sentence level, sentence type, clause type, and sentence length were statistically significant. We will show the (gender) differences with the effect plots. We will also illustrate the results of the random-effects analysis and demonstrate that how the differences were distinguished depending on the gender and files.

In sum, this paper took a corpus-based multifactorial approach and analyzed the corpus data with a mixed-effects model. As a result, this study clearly showed that two simple modals (*can* and *may*) could be used to distinguish the gender differences even within the native speakers.

References

Baker, Paul. 2014. *Using corpora to analyze gender*. London: Bloomsbury.

Brezina, Vaclav. 2013. BNC64. http://corpora.lancs.ac.uk/bnc64

Deshors, Sandra. 2010. *Multifactorial study of the uses of* May *and* Can *in French-English Interlanguage*. Ph.D. dissertation, University of Sussex.

Deshors, Sandra. 2014 Constructing meaning in L2 discourse: The case of modal verbs and sequential dependencies. In Glynn, Dylan and M. Sjölin (eds.) *Subjectivity and epistenicity: Stance strategies in discourse and narration*, 329-348. Lund: Lund University Press.

Deshors, Sandra and Stefan Gries. 2014. A case for the multifactorial assessment of learner language: The Uses of *May* and *Can* in French-English interlanguage. In Glynn, Dylan and Justyna Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 179-204. Amsterdam: John Bejamins.

Gries, Stefan. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum Press.

Holmes, Janet. and Meyrehoff Miriam. 2003. *The handbook of language and gender*. 2nd Edition. Oxford: Blackwell.

Exceptive constructions: A corpus-based perspective

Geonhee Lee
(Kyung Hee University)

Exceptive refers to a type of construction that expresses exclusion to a generalization. According to Hoeksema (1995) and others, there are two types of exceptives. Connected exceptives have a close relation to restricted quantifiers as in (1a) and free exceptives function as clausal modifier that is located adjacently to a clause as in (1b):

(1) a. Every linguistics professor except Jones drives a Mercedes.
b. Except Jones, every linguistics professor drives a Mercedes.

We use the term exceptive phrase (EP) to refer to phrases that consist of a exception marker *except* and its following XP is called remnant. The DP that has a direct relation to EP refers to the term associate (Moltmann 1995).

Exceptive constructions mainly have been dealt with in semantics based on the idea that an exceptive phrase introduces a DP that is interpreted as a set. However, this analysis does not wholly explain properties of the exceptive constructions as there are some instances where exceptives do not have a strong association to DP, rather they indicate a clausal source that the counter examples from exceptives in (2) show (Vostrikova 2019):

(2) a. I got no present except from my mom.
b. Every girl danced with every boy except Eva with Bill.

In (2a), the PP 'from my mom' follows after except which is difficult to be accounted for by traditional semantic approach as the exceptive phrase do not have a connection to a set of things mentioned in the associate. Also, there are some instances that the remnant of the EP consists of multiple constituents which cast doubt on semantic approach and support the idea of clausal source as shown in (2b).

This paper performs a corpus investigation of exceptive constructions using the corpus COCA (Corpus of Contemporary American English). A total of 97,409 tokens are listed from the search string "except". After manually excluding irrelevant examples, 200 random samples are collected.

The identified data raise nontrivial issues that cannot be accounted for with semantic view. For instance, some evidence is found in COCA that the exceptives can be considered as elliptical phenomenon as the EPs do not form a single constituent seen in (3a). The category mismatch between the EP and its associate is pivotal as it supports the EP in ellipsis shown in (3b). Also, note that the syntactic category of the exceptive is not restricted to DP in (3c):

(3) a. She never seemed to go any-where, except out with Matt at night. (COCA 1990 FIC)
b. We can't figure out why, but it doesn't work anywhere else except at your farm. (COCA 2015 MOV)
c. Percival and Ernesto don't live here except on the weekends. (COCA 2002 FIC)

Observing such data challenging the previous studies of exceptive construction, this paper sketches a syntactic analysis of exceptive constructions. In particular, this study suggests that rather than postulating a theoretical set of DPs, a syntactic perspective on exceptive construction provides a clear

view that the EP is elided from its linguistic antecedent, thus having a clausal source.

Selected References

Hoeksema, Jacob. 1995. The semantics of exception phrases. In Jaap Van der Does and Jan Van Eijk (eds.). *Quantifiers, logic and language*. 145-177. Stanford: CSLI Publications.

Moltmann, Friederike. 1995. Exception sentences and polyadic quantification. *Linguistics and Philosophy* 18(3). 223-280.

Potsdam, Eric. 2018. Exceptives and ellipsis. In *Proceedings of the 48th Meeting of the North East Linguistic Society (NELS)* 48.

Stockwell, Richard and Deborah Wong. 2020. Sprouting and the structure of except-phrases. In *Proceedings of the 50th Meeting of the North East Linguistic Society (NELS)* 50. 169-183.

Vostrikova, Ekaterina. 2019. Compositional analysis for clausal exceptives. In *Proceedings of the 29th Semantics and Linguistic Theory Conference (SALT)* 29. 420-440.

Comparing Korean and the U.S. newspaper articles on COVID-19: A corpus-based analysis

Junghyun Lee and Sanghoun Song

(Korea University)

This study makes comparative analysis on the corpuses which consist of the COVID-19-related newspaper articles published in South Korea and the United States by employing natural language processing. The period of analysis is from January 2020 to December 2020. Since the COVID-19 has unprecedentedly impacted the whole world, it is important to examine how other countries have reacted in order to gain better insights into the novel virus and to prepare for the future pandemic. Language is powerful means which reflects one society. Especially the newspaper articles deliver crucial information and they influence the public perception. Adopting natural language processing on the corpuses, we expect to discover different aspects of these two countries in terms of COVID-19 in 2020. South Korea and the United States show distinct topic trends in 2020. The most frequently used words and the words used with the main keywords in the newspapers are also distinguished from each other. By employing quantitative analysis with corpus data and objective examination with natural language processing, this study finds out that South Korea and the United States have reacted to COVID-19 with different point of view.

References

Asif, M., Zhiyong, D., Iram, A., & Nisar, M. (2021). Linguistic analysis of neologism related to coronavirus (COVID-19). Social sciences & humanities open, 4(1), 100201. https://doi.org/10.1016/j.ssaho. 2021. 100201

Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998–2009. Applied Linguistics, 34, 255-278.

Basirat, A., Allassonnière-Tang, M. & Berdicevskis, A. (2021). An empirical study on the contribution of formal and semantic features to the grammatical gender of nouns. Linguistics Vanguard, 7(1), 20200048. https:// doi.org/10.1515/lingvan-2020-0048

Choi, S., Lee, J., Kang, M. G., Min, H., Chang, Y. S., & Yoon, S. (2017). Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks. Methods (San Diego, Calif.), 129, 50–59. https://doi.org/10.1016/j.ymeth.2017.07.027

Ghasiya, P., & Okamura, K. (2021). Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach. IEEE Access, 9, 36645-36656.

Van Aelst, P., Toth, F., Castro, L., Štětka, V., de Vreese, C. H., Aalberg, T., Cardenal, A. S., Corbu, N., Esser, F., Hopmann, D. N., Koc-Michalska, K., Matthes, J., Schemer, C., Sheafer, T., Splendore, S., Stanyer, J., Stępińska, A., Strömbäck, J., & Theocharis, Y. (2021). Does a crisis change news habits? A comparative study of the effects of COVID-19 on news media use in 17 European countries. Digital Journalism. https://doi.org/10.1080/21670811.2021.1943481

Voigt, R., Jurgens, D., Prabhakaran, V., Jurafsky, D., & Tsvetkov, Y. (2019). RtGender: A corpus for studying differential responses to gender. In H. Isahara, B. Maegaard, S. Piperidis, C. Cieri, T. Declerck, K. Hasida, H. Mazo, K. Choukri, S. Goggi, J. Mariani, A. Moreno, N. Calzolari, J. Odijk, & T. Tokunaga (Eds.), LREC 2018 - 11th International Conference on Language Resources and Evaluation (pp. 2814-2820). (LREC 2018 - 11th International Conference on Language Resources and Evaluation). European Language Resources Association (ELRA).

Shen W., Wang L., & Kim J. (2020). A critical discourse analysis of Korea-China newspaper articles on COVID-19 - Focused on medical staff and confirmed patients. Korean Language Research, 57, 97-129.

Semantic role labeling: A layer-wise analysis of KRBERT

Hye-Jin Seo
(Dongguk University)

Background and Purpose: Deep neural networks provide an influential method to model language. Recently developed Bidirectional Encoder Representations from Transformers (BERT; Devlin et al., 2018) show state-of-the-art outcomes in various natural language tasks such as sentence classification and sequence labeling (Shi & Lin, 2019). Even though their performances are remarkably enhanced than other deep learning models such as Recurrent Neural Networks (RNN), understanding of their internal black box is also necessary to interpret deep neural networks for natural language processing (NLP). To investigate how it works, this study uses semantic role labeling (SRL). Semantic role labeling becomes fundamental task in natural language understanding. The task of SRL is to recover the predicate-argument structure of a sentence, determining "who did what to whom", "when", "where", etc (Shi & Lin, 2019; See Figure 1). The ability for extracting features of SRL in sentence is also useful in a variety task such as question answering (Shen & Lapata, 2007) and open information extraction (Fader et al., 2011). The deep learning language models for this task generally depend on linguistic features such as part-of-speech tags (Marcheggiani et al., 2017) and syntactic trees (Roth & Lapata, 2016; Zhang et al., 2018; Li et al., 2018). More specifically, Roth and Lapata (2016) suggest that syntactic features are required to achieve high performance in dependency-based SRL. However, some studies suggested that BERT trained without linguistic features have shown impressive state-of-the-art results in SRL task (Shi & Lin, 2019; Bae et al., 2020). Two aims are settled in this study. First, the present study investigates whether the fine-tuned BERT model without linguistic features performs well in SRL task of Korean. Second, this study examines which hidden-state of layer in BERT model works best in SRL task.

Dataset and model: We conduct experiment on SRL task by using a released corpus, which was made up for a competitive exhibition hosted by NAVER and Changwon National University. The SRL dataset consists of about 0.3 million sentences. Each word in sentence was tagged according to its semantic role. In the current experiment, SNU KR-BERT is used because this pretrained model was known to be carefully trained by considering the linguistic features of Korean. There are two types of tokenization method: character tokenization and sub-character tokenization. Character tokenization splits texts into characters. On the other hand, sub-character tokenization splits the piece of textual data into sub words. In this study, sub-character tokenization method is utilized because this approach is more robust against out of vocabulary (OOV). This model's vocabulary size, parameter size, and data size were 12,367, about 96 M, 20M sentences, respectively. The fine-tuned BERT-BiLSTM-CRF model is used (See Figure 2).

Figure 1. The example of SRL task

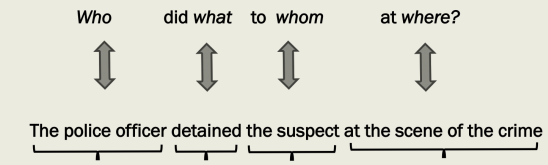
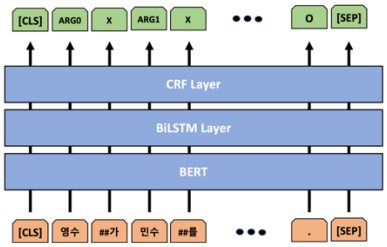


Figure 2. The fine-tuned model for SRL task



BERT model is suitable to improve the semantic representation of sentences. Bidirectional Long Short-Term Memory (BiLSTM) layer makes the fine-tuned model possible to utilize the overall sentence information before deciding the semantic label for each word. Lastly, Conditional Random Field (CRF) layer refers to adjacent semantic label information to determine the semantic label belong to the current word (Bae et al., 2015; Bae & Lee, 2017).

Analysis and Results: We analyzed the performance of model in two ways: First, the performance of the fine-tuned BERT model was examined. In this model, the hidden state of the last hidden layer in KR-BERT model was conveyed into BiLSTM-CRF layers. As shown in Table 1, F1-score of the model was 66.4%. Second, the performance was compared to whether the performance varies depending on which hidden state of the model was selected. KR-BERT was composed of 12 encoder layers (or blocks). Each encoder block has hidden state of the corresponding layer. Specifically, the current study examined three choices (compared to the performance of the last hidden layer). When last four layers were concatenated, F1-score was 67.9%. When the last four hidden states were summed, F1-score was 65.3%. When the 11^{th} hidden state was selected, F1-score was 68.1%.

Table 1. The performance of model depending on the layers

Layer of the fine-tuned BERT model	**F1-score**	Layer of the fine-tuned BERT model	**F1-score**
Last hidden layer	66.4%	Sum last four layers	65.3%
Concatenated last four layers	67.9%	Second-to-last layer	68.1%

Discussion: The performance of model was somewhat low because the provided dataset contains only the most basic information for SRL task. In the dataset, the semantic label of one sentence was indicated, but it was not specified which verb the semantic label was associated with. Therefore, there was a limit in improving the performance of the model. In this study, we focused more on the performance of various layers rather than the performance of last hidden layer. Interestingly, the performance varied depending on which hidden state of layers was chosen and the way to convey hidden state of layers into BiLSTM-CRF layers. The performance of the fine-tuned model was better when it was trained by the hidden state of second-to-last layer and the concatenated last four layers, which eventually outperformed the last hidden layer. In particular, 11^{th} layer outperforms comparing with other combinations of layers in SRL task. However, in general, the last hidden state of layer is only used to train deep learning models. The current study suggests that it is necessary to examine which layer is more appropriate to improve the performance of model according to task.

References

배장성, 이창기, 임수종, & 김현기. (2020). BERT 를 이용한 한국어 의미역 결정. *정보과학회논문지*, *47*(11), 1021-1026.

Bae, J., & Lee, C. (2015). End-to-end Learning of Korean Semantic Role Labeling Using Bidirectional LSTM CRF. In *Proc. of the KIISE Korea Computer Congress 2015* (pp. 566-568).

Bae, J., & Lee, C. (2017). Korean Semantic Role Labeling using Stacked Bidirectional LSTM-CRFs. *Journal of KIISE*, *44*(1), 36-43.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fader, A., Soderland, S., & Etzioni, O. (2011, July). Identifying relations for open information extraction. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 1535-1545).

Li, Z., He, S., Cai, J., Zhang, Z., Zhao, H., Liu, G., ... & Si, L. (2018). A unified syntax-aware framework for semantic role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 2401-2411).

Marcheggiani, D., Frolov, A., & Titov, I. (2017). A simple and accurate syntax-agnostic neural model for dependency-based semantic role labeling. arXiv preprint arXiv:1701.02593.

Roth, M., & Lapata, M. (2016). Neural semantic role labeling with dependency path embeddings. arXiv preprint arXiv:1605.07515.

Shen, D., & Lapata, M. (2007, June). Using semantic roles to improve question answering. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)* (pp. 12-21).

Shi, P., & Lin, J. (2019). Simple bert models for relation extraction and semantic role labeling. *arXiv preprint arXiv:1904.05255*.

Zhang, Y., Qi, P., & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. *arXiv preprint arXiv:1809.10185*.

How does BERT address polysemy of Korean adverbial postpositions *-ey*, *-eyse*, and *-(u)lo*?

Seongmin Mun
(Chosun University)

Polysemy, one type of ambiguity, occurs when one form delivers multiple, and yet related, meanings/functions (Glynn and Robinson, 2014). In this regard, Bidirectional Encoder Representations from Transformer (BERT; Devlin et al., 2018) which is a contextualized word-embedding model—word vectors that are sensitive to the context in which they appear—has yielded significant improvements on the task of polysemy interpretation (Yenicelik et al., 2020).

I pose a question as to how BERT applies to the polysemy of a function word such as a postposition in Korean, a language typologically different from the major Indo-European languages that have been investigated for this task. I report a computational simulation that explores how BERT account for polysemy of Korean adverbial postpositions *-ey*, *-eyse*, and *-(u)lo*, which interprets as eight functions for *-ey*, two functions for *-eyse*, and six functions for *-(u)lo* (Shin, 2008). Suppose the following sentence involving the postposition *-(u)lo* as a marker of DIR (direction) as in (1).

(1) pemin-un etwuwun kolmok-ulo talana-ss-ta.
criminal-NOM dark alley-DIR flee-PST-DECL
'The criminal fled into a dark alley.'

For this purpose, I used the Sejong corpus (Kim et al., 2007; 90% for training and 10% for testing), with semantic annotation of this corpus cross-verified by three native speakers of Korean ($\kappa = 0.95$). For model training, I devised the BERT model by employing KoBERT as a pre-trained model in order to obtain high accuracy of outcomes (Jeon et al., 2019). I then fine-tuned the pre-trained model 50 times (i.e., 50 epochs) by using the training set. In each epoch (i.e., learning step), model performance was measured by comparing the intended function of the postposition in each test sentence with the classified function of the postposition via the BERT model. In addition, I developed a BERT-based visualization system to understand how a BERT model simulates human interpretation of word-level polysemy of Korean adverbial postpositions.

I note two major findings of the current study. First, there is an inverse relation between the classification accuracy and the number of functions of each postposition. Second, the model can identify the intended functions of a postposition as the epoch progresses, even though the corpus size of a function is small. However, despite these findings, the BERT model still seems to be affected by semantic closeness between the items, limiting its performance in the given task to some extent. The findings of this study should be further verified by incorporating more postposition types that have similar degrees of polysemy that three adverbial postpositions demonstrate, which I plan to pursue next. I believe our visualization system will contribute to extending the current understanding of how BERT works for language tasks (particularly in non-English settings).

References

David Yenicelik, Florian Schmidt and Yannic Kilcher. 2020. How does BERT capture semantics? A closer look at polysemous words. Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, pages 156-162

Heewon Jeon, Donggeon Lee and Jangwon Park. 2019. Korean BERT pre-trained cased (KoBERT). Retrieved from https://github.com/SKTBrain/KoBERT

Hyo-pil Shin. 2008. The 21st sejong project: with a focus on selk (sejong electronic lexicon of korean) and the knc (korean national corpus). In The 3rd International Joint Conference on Natural Langauge Processing.

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.

Zellig S. Harris. 1954 Distributional Structure. WORD, 10(2-3): 146-162.

Byoungsoo Kim, Yong-Hun Lee, and Jong-Hyeok Lee. 2007. Unsupervised semantic role labeling for Korean adverbial case. Korean Institute of Information Scientists and Engineers, pages 32–39.

Syntactic priming by LSTM L1/L2 language models

Sunjoo Choi
(Dongguk University)

Introduction: Language models (LMs) based on recurrent neural networks (RNNs) have been shown to track abstract properties of the sentence and acquire human-like linguistic knowledge (van Schijindel et al. 2018). Recently, drawing on the syntactic priming paradigm assumed as part of abstract linguistic knowledge, Prasad et al. (2019) establish a gradient similarity metric between structures and apply this method to reconstruct the organization of the LM's syntactic representational space. In so doing, they demonstrate that LSTM LM's representations of different types of sentences are organized hierarchically in a linguistically interpretable manner, suggesting that the LMs track abstract properties of the sentence. In this research, given the LM's ability to track abstract properties of the sentence, we refer to the previous study to investigate how L2 LSTM LMs process the syntactic priming paradigm, compared to L1 human performances. In other words, in line with previous work, we examine how much L2 LSTM LM can attain the syntactic priming performance of native speakers.

Background: Syntactic priming occurs when processing of a target sentence is facilitated following processing of a prime sentence that has the same syntactic structure (Bock, 1986). For example, after encountering sentence like (1) (the prime), readers expect sentence like (2a), which shares syntactic structure with the prime sentence than (2b).

(1) The boy threw the dog the ball.
(2) a. The chef made the guest some excellent pizza.
b. The chef made some excellent pizza for the guest.

If (1) primes (2a) than it primes (2b), we can infer that the representations of (1) are more similar to that of (2a) than to that of (2b). The effects of syntactic priming can be cumulative and long-lasting: sentences with a shared structure Sx become easier to process when preceded by *y* sentences with the same structure Sx than when preceded by *n* sentences with a different structure Sy (Kaschak et al., 2011). Applying cumulative priming paradigm, Prasad et al. (2019) demonstrate that 75 of the LSTM LM's representations of different types of sentences with relative clauses(RC) are organized in a linguistically interpretable manner: sentences with a particular RC were most similar to other sentences with the same type of RC.

Models: We design 9 of the LSTM language models employing the model architecture proposed by Gulrodava et al. (2018); these LMs varied in the number of hidden units per layer (100, 200, 400) and the number of tokens they were trained on (7 million, 10 million, 13 million). For each training corpus, we collect all the sentences extracted from EBS-CSAT English Prep Books and English textbooks for L2ers published in 2016~2018. The process of calculating the similarity (the adaptation effect) is schematized in Figure 1.

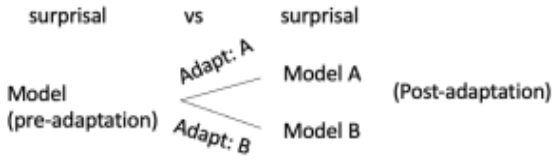


Figure 1: Adaptation and calculating the adaptation effect

We measure the change in surprisal after adaptation when the L2 LM is tested either on sentences with the same structure or sentences with different structures. The surprisal values were averaged across the entire sentence.

Syntactic structures: We follow the same factorial approach and code of Prasad et al. and all experimental items are taken from Prasad et al., as illustrated in table below. We use five experimental lists and each adaptation set involved 20 sentences and each test set involved 50.

Structure	Example
Unreduced Object RC	The teachers that my shy partner admired last year moved the stone remorsefully.
Reduced Object RC	The teachers my shy partner admired last year moved the stone remorsefully.
Unreduced Passive RC	The teachers that were admired by my shy partner last year moved the stone remorsefully.
Reduced Passive RC	The teachers admired by my shy partner last year moved the stone remorsefully.
Subject RC	My shy partner that admired the teachers last year stirred the water.

Discussion: We employ a novel technique to analyze how the representations of various syntactic structures are organized in L2 LSTM LMs. Applying this technique, we gain insight into the representations of sentences with relative clauses in L2 LSTM LMs. Under the adaptation-as-priming paradigm, sentences that share the same specific structure are more similar to each other than lexically matched sentences that do not share the structure. Consequently, experiments with controlled materials show that L2 LMs are capable of tracking abstract syntactic constructions, as native speakers do.

References

Bock, J. Kathryn. 1986. Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 575.

Gulordava, Kristina., Bojanowski, Piotr., Grave, Edouard., Linzen, Tal., & Baroni, Marco. 2018. Colorless green recurrent networks dream hierarchically. *arXiv preprint arXiv:1803.11138*.

Prasad, Grusha., Van Schijndel, Marten., & Linzen, Tal. 2019. Using priming to uncover the organization of syntactic representations in neural language models. *arXiv preprint arXiv:1909.10579*.

Kaschak, Michael P., Kutta, Timothy. J., & Jones, Jones. L. 2011. Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic bulletin & review*, *18*(6), 1133-1139.

Van Schijndel, Marten., & Linzen, Tal. 2018. A neural model of adaptation in reading. *arXiv preprint arXiv:1808.09930.*

Carefully does it, reckless does not: A construction grammar approach to the X *does it* construction

Hanbeom Jung
(Hankuk University of Foreign Studies)

The aim of this study is to provide a Construction Grammar analysis of the $X_{ADV/ADJ}$ *does it* construction in English. Specifically, it explores the construction's idiosyncratic formal properties and non-compositional functional properties by investigating attested data obtained from corpus. Furthermore, the study defines the X *does it* construction as the INSTRUCTIONAL MANUAL construction, which is a genre-based construction, rather than a kind of imperative.

This study initially focuses on the fact that what *easy does it* conveys cannot be accounted for by the lexical meanings of its components or general syntactic rules. In this regard, previous examinations classified *easy does it* as an idiomatic expression that does not accord with general syntactic rules but serves a function of giving directions just as imperative clauses.

However, this study argues that the target construction is not a construct of the English imperative construction. To begin with, the X *does it* construction is not congruent with the formalism of the English imperative construction (Stefanowitsch 2003:2); for example, its verb does not appear in a bare form and its grammatical subject seemingly is not a second person. Based on these formal differences, the study does not regard the construction in focus as a type of the English imperative construction. Furthermore, the speaker of *easy does it* intends to let addressees know how to execute a course of action, rather than to force them to. In other words, the directive illocutionary force of the X *does it* construction is weaker than that of the English imperative construction. Such a functional difference qualifies the X *does it* construction to be an independent construction. Comparing the X *does it* construction with the English imperative construction, this study analyzes the phenomenon in focus within the framework of Construction Grammar.

Attested data obtained from iWeb corpus shows that not only *easy* but other adverbs/ adjectives, such as *gently* and *slow*, can also appear in the initial position of the construction. This tells us that *easy does it* should be understood as a construct of a more schematized structure. The attested data was categorized into four different groups and the constructional properties of each group are analyzed. In addition, a consultation with English speakers is conducted, which backs up the idea that the X *does it* construction is differentiated from the English imperative construction both formally and functionally. Moreover, the corpus search reveals that the X *does it* construction is frequently used in a certain genre of contexts, which is an instructional manual. Based on this observation, this study argues that the construction can be labeled as the INSTRUCTIONAL MANUAL construction. This study is significant in that it added the INSTRUCTIONAL MANUAL construction to the inventory of the genre-based constructions of Ruppenhofer and Michaelis (2010).

References

Croft, William. and Cruse, D. Alan. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.

Hilpert, Martin. 2014. *Construction Grammar and Its Application to English* (Edinburgh Textbooks on the English Language). Edinburgh: Edinburgh University Press.

Ruppenhofer, Josef and Laura A. Michaelis. 2010. A constructional account of genre-based argument omissions. Constructions and Frames 2(2), 158-184.

Stefanowitsch, Anatol. 2003. “The English imperative: A construction-based approach,” Paper presented at the Workshop Form and Function of Sentence Types at the Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Ludwig-Maximilians-Universität München, Ferbruary 2003.

Much less Construction in English: A Direct Interpretation Approach

Jiseop Hyeon
(Kyung Hee University)

The so-called much less construction is superficially similar to a coordinate structure, involving *and*, *or* and *for*. See Carlson and Harris (2018)'s example.

(1) a. John can't eat *caviar* and **veal**.
b. John can't eat *caviar*, much less **veal**.

In (1), two constituents in each sentence are conjoined by *and* in (1a) and *much less* in (1b). In case of the *much less* construction, these constituents are the correlate *caviar* in the antecedent clause, and the remnant *veal* after much less. That is, the *much less* construction coordinates two constituents in terms of their syntactic category, and their match of semantic representation and types (Partee and Rooth 1983). From a syntactic perspective, the examples (1) seem to behave quite alike, but differ in many respects. Their differences mainly arise from the fact that the *much less* construction can have a word or a phrase, but not a full clause. Accordingly, the *much less* construction has an elliptical phenomenon.

To pose a problem for the movement and deletion approach, there are some variables which the analysis requires. The first is to categorize both a matrix and embedded environment of a correlate in an antecedent clause. The second is to find the data, violating island constraint in terms of syntactic structure. In addition, another key point is to investigate whether or not the construction has an overt correlate in terms of sematic and pragmatic representation related to the correlate. Through this analysis, we will resolve the ellipsis of the *much less* construction.

There is little previous research on the construction, and their research has just mainly focused on parallelism (Harris and Carlson 2019). At first, I investigated the authentic uses of the *much less* construction in English using the corpus, COCA (Corpus of Contemporary American English). I analyzed the data (1,297 tokens) of the *much less* construction in genres SPOK and FIC, and filtered out the data (100 tokens) like comparatives, and the total is 1,397 tokens.

Through the analysis, I found its syntactic category match case (1,210 tokens), and mismatch case (87 tokens), which involves sprouting case. The mismatch case varies in perspectives of syntactic category mismatch and tense mismatch like the below examples.

(2) a. Until a few months ago, finding a man or woman in the park *in trouble*, **much less dead**, would be a rare thing. (COCA 2017 FIC)
b. The more I dug and the more that I saw, it was just absolutely incredible that this young boy *had ever went to jail*, **much less is still in jail** . . . (COCA 1999 SPOK).
c. Nobody had ever seen this group of people get together, **much less like onstage and working with each other like that**. (COCA 2016 SPOK)

The *much less* construction has an elided phrase, and this ellipsis phenomenon can be supported by two

approaches: the movement and deletion approach and a direct interpretation approach. Merchant (2004) argues an ellipsis construction is a kind of syntactic phenomenon, and it involves movement and deletion. If we use this approach, the *much less* construction can be compared to fragment in that it only takes and conjoins a word or a phrase, not a full clause. However, there are some examples which cannot be explained by the movement and deletion approach as seen in (3).

(3) a. It wasn't the moment, Schnooty decided, to tell *her anything*, **much less both of them that he'd be gone in a few days, perhaps for good if things went badly**. (COCA 2013 FIC)
b. I don't know how happily *the story of our colony on Green* will end, **much less the story of my own life**. (COCA 2009 FIC)

The examples (3) show that some data cannot be explained in terms of the movement and deletion approach. The example (3a) shows the remnant has double objects *both of them that he'd be gone in a few days*, and the example (3b) violates Ross (1967)'s island constraints. To account for the authentic grammatical properties, I attempt to adopt a direct interpretation approach which makes use of semantics and pragmatics as well as syntax (Kim and Abeillé 2019).

References

Carlson, Katy and Jesse A. Harris. 2018. Zero-Adjective contrast in much-less ellipsis: the advantage for parallel syntax. *Language, Cognition and Neuroscience* 33(1): 77-97.

Harris, Jesse A. and Katy Carlson. 2019. Correlate not optional: PP sprouting and parallelism in "much less" ellipsis. *Glossa* 4(1): 83-107.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27: 269-289.

Kim, Jong-Bok and Anne Abeillé. 2019. Why-stripping in English: A corpus-based perspective. *Linguistic Research* 36(3): 365-387.

Partee, Barbara and Mats Rooth. 1983. Generalized conjunction and type ambiguity. In R. Bäuerle, C. Schwarze, and A. von Stechow (eds.), *Formal semantics: The essential readings*. Walter de Gruyter, Berlin.

Ross, John Robert. 1967. *Constraints on Variables in Syntax*. Ph.D. dissertation, MIT.

“Call me by my name”: Names, address, and identities of Korean women

Eldin Milak
(Sungkyunkwan University)

The complexity of selecting an appropriate form of address in Korean cannot be overstated. While the process of negotiating the terms usually requires only a few brief exchanges, the entanglements of power, status, intimacy, and expectations which transpire within those moments require high social cognizance, cultural sensitivity, and personal awareness that take years to be socialized into. The tenor that is set in the early stages of an interaction often persists long into the relationship, and some speakers may continue using the initially selected terms of address despite possible changes in status over time. Within frames of this demanding social practice, there is no form of address more risky and potentially more rewarding than that of personal names.

Personal names in South Korea are subject to avoidance and restrictions in use grounded int the asymmetric relations of power and age that constitute the social and linguistic ideologies of the country (Brown 2015; King 2006). At the same time, as lexical items whose positioning within the Korean referential hierarchy allows them to assume more directly referential roles, similar to pronouns, while retaining a conventional denotation characteristic of symbols, such as kinterms, names have the unique reformational power to change normative social practice (Fleming & Slotta 2015; Fleming 2011). Their mobility within the referential hierarchy enables forms of intentional and reflexive practice situated in spatiotemporal settings of speech events (Ball 2018). In turn, the socio-pragmatic boundaries derived from such a setting impose constraints on the movement of names within the system of address, ultimately delimiting their performance. The resulting boundaries are not readily permeable but are also not impenetrable.

As this study aims to show, the affordances provided by the unique status of names in the Korean referential hierarchy allow speakers to (re)negotiate the parameters of the relationship indexed by a particular speech event and reposition themselves in the wider chronotopic setting (Agha 2015; Bakhtin 1981) through performance on the smaller scale of interaction. This study focuses primarily on female speakers, since negotiation of this sort is often at odds with the process by which Korean women are linguistically socialized (Duff 2010) into extant systems of address emblematic of social institutions, such as the family, and liminal life events, such as adulthood, marriage, and motherhood. The selection and usage of names over other available and approved alternatives is interpreted as movement of Korean women towards greater visibility and subject-positions in both microscale interactions and macrosocial spatiotemporal settings and institutions.

This phenomenon is explored through interviews and open ethnographic observation (Copland & Creese 2017) of thirty Korean women from different social and regional background, as they navigate the thin line between normative social propriety and personal identity performance and maintenance within the spatiotemporal setting of Korea’s compressed modernity (Chang 2010). The study findings suggest that the practice of usage and avoidance of names is undergoing transgenerational change to reflect the equally changing position of women in the country, simultaneously altering the internal structure of the Korean system of address. Although the study traverses the intersection of gender and power relations, its ultimate focus is sociolinguistic in nature, aiming to circumscribe a unique linguistic phenomenon unfolding within a unique social setting.

Keywords: names, address, women, identity, South Korea.

References:

Agha, A. (2015). "Chronotopic formulations and kinship behaviors in social history." *Anthropological Quarterly*, 88(2): 401–415.

Ball, C. (2018). "Language of Kin Relations and Relationlessness". *Annual Review of Anthropology*, 47: 47-60.

Bakhtin, M. (1981). *The Dialogic Imagination*, Holquist, M. (Ed.), University of Texas Press.

Brown, L. (2015). "Honorifics and Politeness". In *The Handbook of Korean Linguistics*, Brown, L. & Yeon, J. (Eds.), 303-319. Wiley Blackwell Publishing.

Chang, K-S. (2010). "The second modern condition? Compressed modernity as internalized reflexive cosmopolitization." *British Journal of Sociology*, 61(3): 444-464.

Copland, F. and Creese, A. (2017). *Linguistic Ethnography: Collecting, Analysing and Presenting Data*. SAGE Publications Ltd.

Duff, P. (2010). "Language Socialization." In *Sociolinguistics and Language Education*, Hornberger, N. and McKay, S. L. (Eds.), 427-452. Bristol: Multilingual Matters.

Fleming, L. and Slotta, J. (2015). "Named relations: A universal in the pragmatics of reference within the kin group". *Chicago Linguistics Society*, 165-179.

Fleming, L. (2011). "Name Taboos and Rigid Performativity". *Anthropological Quarterly*, 84(1): 141-164

King, R. (2006). "Korean Kinship Terminology." In *Korean Language in Culture and Society*, Ho-Min Sohn (Ed.). University of Hawaii Publishing.

L1 and L2 processing of lexical bundles

Hyojin Jeong
(Sungkyunkwan University)

Multiword sequences such as *on the contrary* had been largely assumed to be stored and processed as whole units in the mental lexicon. This assumption is clear from the terms often used to describe such sequences, such as *prepackaged* (Clark,1974), *ready-made* (Raupach,1984), and *unanalyzed* (Fillmore, 1979; Hickey, 1993). It is also evident in the definition of such multiword sequences. For example, in Wray's (2002, p. 9) definition, formulaic sequences (e.g., *to begin with, in other words)* are those "stored and retrieved whole from memory at the time of use". According to the most traditional holistic view (Peters, 1983; Weinert, 1995; Wray, 2002), these formulaic sequences are represented as single units and retrieved directly from the mental lexicon as wholes. This holistic view is also in line with the usage-based theories of languages. To date, the holistic hypothesis has been put to empirical test with different types of multiword sequences in focus. Previous studies provided supporting evidence for the holistic hypothesis with respect to formulaic sequences (e.g., Jiang & Nekrasova, 2007). However, findings are inconclusive with regard to lexical bundles (e.g., *that there is no, the presence of a)*. Thus, the present study examined the representation and processing of lexical bundles in second language (L2) learners of English and native English speakers.

To examine the processing of lexical bundles, 28 L2 learners and 30 native speakers were tested using a visual word monitoring task. The critical test materials include 30 lexical bundles and 30 non-lexical bundles in addition to 30 formulaic sequences and 30 nonformulaic sequences. In the visual word monitoring task, the participants were presented with a target word (the ending word of the critical expressions or their controls) first followed by a string of words that include either the critical items or their controls. The participants were instructed to first remember the target word and then monitor the following string of words for the target word. They were asked to press the YES button on the keyboard as quickly as possible when they saw the target word.

The results show that both groups responded significantly faster to the ending words of the formulaic sequences than to the same target words in the nonformulaic controls, while no such processing advantage was found for the lexical bundles over the controls for both groups. The observed processing advantage for formulas is consistent with the findings from earlier research (e.g., Arnon and Snider 2010; Jiang & Nekrasova, 2007). More importantly, the observed processing advantage for formulas in the word monitoring task suggests that the task provides an adequate means for examining the processing issue of multiword sequences. The lack of a processing advantage for LBs, however, was also compatible with the results of some previous studies. Schmitt et al. (2004), for example, reported no clear evidence for holistic representation and processing of multiword sequences including LBs in both L1 and L2 speakers. Tremblay and Tucker (2011) also found that phrase type (whether the phrase can stand alone or not) was a significant predictor in influencing the participants' performance in a reading-aloud task. Overall, the results seem to suggest that a high frequency of occurrence may not be the only factor that determines holistic processing of multiword units. Possible explanations will be discussed with respect to the differences between formulaic sequences and lexical bundles in mental representation.

Selected References

Arnon, I. and Snider, N. 2010. "More than words: Frequency effects for multi-word phrases," *Journal of memory and language, 62*(1), 67-82.

Jiang, N.A. and Nekrasova, T.M. 2007. "The processing of formulaic sequences by second language speakers," *The Modern Language Journal, 91*(3), 433-445.

Schmitt, N., Grandage, S. and Adolphs, S. 2004. “Are corpus-derived recurrent clusters psycholinguisticallyvalid,” in Norbert Schmitt ed., *Formulaic sequences: Acquisition, processing and use,* 127-151. Philadelphia: John Benjamins Publishing Company

Tremblay, A. and Tucker, B.V. 2011. “The effects of N-gram probabilistic measures on the recognition and production of four-word sequences,” *The Mental Lexicon, 6*(2), 302-324.

Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

The electro-physiological evidence on the processing of the Korean conversational implicature

Shin-ae Yoon, Hyenyeong Chung, and Yunju Nam
(Konkuk University)

The “conversational implicature”. It has not drawn much attention other than pragmatic phenomena in the neurolinguistic perspectives. Jang et al. (2013) ’s fMRI study investigated the functional neuroanatomy during conversational implicature processing by varying the degree of implicity caused by violating the maxims of relation. The results revealed language-related regions in both moderately and highly indirect answers than in the direct one. Interestingly, the activation in the left anterior temporal lobe in the highly indirect condition compared to the moderate indirect condition suggests the central role in inference, such as semantic integration. However, whether the degree of the implicitness or the intensity of the activation in the anterior temporal lobe, caused by longer involvement modulates the involvement of the anterior temporal lobe so that whether we could tell the highly implicit answer are distinctively different from the moderate implicit one is not clear.

To gain more apparent evidence on those questions, we conducted an ERPs experiment with the same conditions of the study above, adding one condition (e.g., unrelated answer). With its high temporal resolution, this study can be expected to clarify the vagueness in inference processing in verbal communication.

160 adjacency paired utterances were used, comprising interlocuter A’s question and interlocutor B’s answer to that. The implicity of the same target utterance (interlocutor B’s utterance) was manipulated for 4 conditions (Table 1). Items were distributed evenly in four lists in the Latin-square and counter-balanced. 34 participant’s brain responses were recorded with BrainAmp DC amplifiers using 32 channels ActiCap in 250 Hz.

The behavioral data and brain response data from 28 Korean native speakers (Mean age = 22.3, 13 male) were analyzed. We conducted separate analyses on every four phrases of Answer utterance to investigate the time course of conversation implicature processing. EEGs were epoched from -100ms to 900ms after the onset of each phrase and averaged using the -100 ms to baseline correction. 300-450ms (for N400) and 450-550ms (for P600) time windows for the statistical analysis were selected based on the visual inspection. Preliminary, the ERP data in the midline central area (i.e. average value of the Cz, CP1, CP 2) and the anterior midline area (i.e. average value of the Fz, FC1, and FC2) were analyzed to capture the critical effect. We used the Linear Mixed Effect Regression Model to analyze the behavioral data using R software, and the repeated measures ANOVA and compared t-test for the ERP data. Only accurate trials in behavior data were included for the statistical analysis.

[Table 1] Example of experimental materials

Condition	Question	Answer
Direct Answer(DA)	교수님 자동차 어디에 있어요? Where is the professor’s car?	교수님 / 자동차 / 밖에 / 있던데요. The professor’s car is out there.
Moderlately Implicit (MI)	오늘 교수님 나오셨어요? Did the professor come here?	
Highly Implicit (HI)	오늘 강연자 도착했나요? Did today’s lecturer arrive?	
Unrelated Answer(UA)	지도교수님 전공이 뭐예요? What is the major of your supervisor professor?	

Across all conditions, participants showed over 99 percent accuracy, meaning that the task itself was not challenging. The statistical analysis showed that the degree of the implicity significantly increased in reaction time, which is likely to say that participants had more difficulties processing the implicit answer than no implicity (i.e., DA and UA) (Table 2).

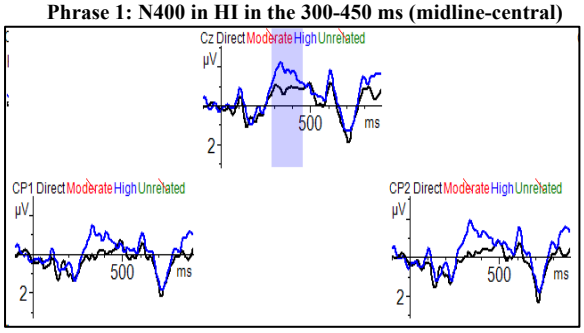
[Table 2] The result of the LMER analysis on the reaction times for each condition

	Estimates	S.E	t-value
Intercept	380.67	48.44	7.858
Moderlately Implicit (MI)	236.77	35.53	6.665***
Highly Implicit (HI)	320.58	30.69	10.446***
Unrelated Answer(UA)	18.80	26.39	0.712

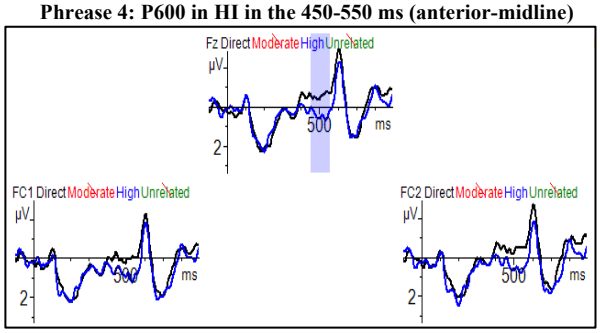
***refer to the significant role of the factor at a < 0.001

At the first phrase position, there was a significant main effect of the condition in 300-450 ms time window ($F(3,81)=5.849$, $p<.01$) in the midline central area. The compared t-test between conditions confirmed that the N400 effect was revealed in the HI condition (-1.749 μV) compared to DA(-0.662 μV, $p<.05$), MI(-0.427 μV, $p<.01$), and UA(-0.227 μV, $p<.001$). This effect would be due to the lexical effect since we used the same or related words to DA in MI or UA conditions to manipulate the implicity. Critically, at the last word (sentence-final predicate), the positive deflection in HI (0.208 μV) condition compared to DA (-0.910 μV, $p<.05$) and UA(-0.849 μV, $p<.05$) was increased in 450-550 ms time-window in the anterior midline area. Considering the topographical distribution and the time window, we could interpret this effect as P600 reflecting the higher cognitive load for the inference because of the high implicity in HI condition. However, we could not find any effect in the MI condition compared to the DA condition.

Phrase 1: N400 in HI in the 300-450 ms (midline-central)



Phrase 4: P600 in HI in the 450-550 ms (anterior-midline)



With this ERP evidence, we could conclude that conversational implicature requires the cognitive processing in line with the fMRI study. Moreover, since there was not any significant difference in HI condition compared to MI condition, they are not distinctively different from each other. No difference between MI condition and DA or UA condition, however, is needed the further discussion.

References

Jang, G., Yoon, S. A., Lee, S. E., Park, H., Kim, J., Ko, J. H., & Park, H. J. 2013. Everyday conversation requires cognitive inference: neural bases of comprehending implicated meanings in conversations. *Neuroimage* 81: 61-72.

Teachers' question repairing in classroom

Songdan Piao
(Sungkyunkwan University)

Classroom conversation is a kind of institutional interaction. It is a one-sided transfer process between two parties: teachers and students. Different from ordinary conversations, a classroom conversation aims to transfer knowledge. For this kind of highly purposed conversation, cooperation between the parties is essential in order to accomplish the common goals of both participants. Interaction between teachers and students can provide an atmosphere of synergy for classroom conversation and help both sides think in sync (Brown, 2001). One of the ways to increase interaction is question. In a classroom conversation, which is a traditional institutional interaction, teachers have the right to ask questions freely. By making successful questions (i.e., questions with students' answers), they can not only encourage students to think and keep involved, but also promote classroom interaction. Moreover, teachers can measure student's knowledge and comprehension.

This article aims to clarify teachers' motivation of using question repairs in a classroom conversation by applying conversation analysis method. Two English classroom videos were selected from YouTube websites. All of these data are English instructions in Chinese high schools. One video is an English course at a high school in Huangshan city. The second video is an English course in NO.1 Middle School Affiliated to Central China Normal University. After the transcript, the number of sequences containing questions was counted. The sequences are distinguished according to the topic of the question. All of the questions made by teacher are calculated and the target of this paper is repaired questions. There are 345 questions in the collected video data. The most common question types are Wh-questions (145) like *why it is impossible*, and the second most frequently used question is Yes/No question: a total of 79. The reason teachers use Wh-questions more frequently is that the answer to these questions is not short-answered, and students must understand the learning content in order to answer. In contrast, although Yes/No question is easy for students to understand and answer, it takes longer to ask questions than to answer, which may result in too little oral expression by students and too much oral expression by teachers. Most of the questions shows a typical IRE pattern, a sequence like teacher's question - student's response - teacher's feedback (Johnson, 1995).

Besides them, there are 48 repaired sequences. By observing previous segments and following segments of repairing sequences, we determined the factors that induce repair. The results showed that the teachers modified their own question because they failed to get immediate responses after the question, and the sequence was question – repair 1 – repair 2 – repair 3 – response; sometimes teachers modified his/her own question without any interval. These repairs indicate that teachers strive to simplify their own questions by applying different types of question (e.g., Yes/No question – Wh-question – Uncompleted statements; Yes/No question – Yes/No question – Uncompleted statements), or making the questions more specific, by doing so, teachers help students keep thinking during the class and have more chance to express themselves in the classroom conversation.

References

Brown, H. D. (2001). Teaching by principles: an interactive approach to language pedagogy. New York. In: Pearson. Language assessment principle and classroom.

Johnson, K. E. (1995). Understanding communication in second language classrooms: Cambridge University Press.

Use of English intensifiers by gender and text genre

Yoon-Young Jeon and Eugene Chung
(Korea University)

In the processes of communication, discourse participants intentionally use intensifier to emphasize their opinions or express their emotions. The intensifier chosen for such purposes serve to more clearly convey the meaning that the speaker wants to express. At the same time, intensifiers, which are adverbs of degree, play a significant role in controlling the mood and strength of the entire sentence.

Lakoff (1975) proposed characteristic differences in the pattern of language use according to the gender of the speaker. He claimed that the frequency of using of adverbs of degree such as intensifiers was higher in women than in men as a characteristic of women's language. Many subsequent studies have focused on the frequency of intensifier use in research on the characteristics of women's language. Based on the spoken and written data of the International Corpus of English-Great Britain (ICE-GB), Chung (2018) analyzed the frequency of intensifier use according to gender and educational background. Jeon & Choe (2009) extracted data from the ICE-GB to confirm differences in the use of adverb according to gender. They showed that the key adverbs for women are *really, definitely, honestly, never, just, that*, and *too*, while the key adverbs for men are *certainly, clearly, entirely, right*, and *simply*.

This study goes beyond the presentation of gender differences in the frequency of intensifier use. The purpose of this study is to identify the pragmatic implications of the high frequency of intensifier use in women based on usage patterns in the discourse context as well as to measure the frequency of intensifier use according to gender and text genre. This study identifies the purposes and functions of intensifiers and examines the differences in usage frequency through statistical analysis of text genre in the ICE-GB.

This study uses discourse analysis to demonstrate how intensifiers are used to reveal the speaker's intentions as well as how they play a role in Face Threatening Acts (FTA) (Culpeper, 2011) by either indicating the degree of impolite speech or by helping to avoid an FTA in order to protect oneself or others. This study also identifies the pragmatic aspects of intensifiers by analyzing intensifier usage according to the context in the ICE-GB. The use of intensifiers by gender is analyzed by dividing spoken data from into private and public discourse genres. We then determine whether there is a genre difference in the use of intensifiers by gender.

References

Culpeper, J. 2011. *Impoliteness Using Language to Cause Offense*. Cambridge, MA: Cambridge University Press.

Chung, S. 2018. Gender, education background, and the use frequency ratios of tag questions and intensifiers. *Discourse and Cognition*, 25(3): 69-99.

Jeon, J. and J. Choe. 2009. A key word analysis of English intensifying adverbs in male and female speech in ICE-GB. *Proceedings of the 23rd Pacific Asia Conference on Language, Information, and Computation*, 210-219.

Jeon, J. 2010. *A Key Word Analysis of Gender-specific Vocabulary in English*. Doctoral dissertation, Korea University.

Lakoff, R.1975. *Language and Women's Place*. New York: Harper & Row.

Ditransitive construction in English: A corpus-based study

Jennifer Pearl

(Kyung Hee University)

A ditransitive construction is basically a construction which consists of a verb which has two objects - where the verb is called a ditransitive verb, an agent argument which is the subject, a recipient-like argument (first object), and a theme argument (second object) (cf. Malchukov et. al 2007).

(1) a. I brought you some soup. (COCA 2018 MOV)

b. I wrote Daddy the letter. (COCA 1998 FIC)

The examples of ditransitive construction mentioned above tend to vary according to different syntactic and semantic properties. For example, the concept of *dative shift* or *dative alternation* has a strong connection with ditransitive constructions as mentioned by Dryer (1986) and also the term *Indirect Object Advancement* introduced by him plays a major role in describing the functions in the examples below.

(2) a. John gave a book to Mary.

S DO IO

b. John gave Mary a book.

S IO DO

The above example set is a classical dative analysis of the English ditransitive sentences and therefore, it shows that (2a) with a PP, is accounted to be basic, involving a single syntactic level whereas (2b) is non-basic, involving two levels – an initial level, at which the grammatical relations are identical to that of (2a) and a final level, at which the initial IO is the final DO. Therefore, (2b) takes the rule of *Indirect Object Advancement* where the IO advances to become the DO (cf. Dryer 1986).

An interesting fact of the ditransitive constructions is that they also have the ability to introduce an understanding of such sentences in (3), which is quite beyond the realistic sense or giving out a figurative use, compared to the literal understanding.

(3) a. The medicine brought him relief.

b. The music lent the party a festive air.

Keenly being noted, *the medicine* and *the music* in (3a) and (3b) respectively, means that they were the cause of something which is followed by in those examples. This may lead to CAUSE-MOTION/CAUSE-RECEIVE effect and is considered a conceptual metaphor in the sense of Lakoff and Johnson (1980).

Considering the properties of a common ditransitive construction, a lot of research and study has been made until now which explains the stability of its syntactic and pragmatic existence in the language. This study performs a corpus investigation of the construction using the corpus COCA (Corpus of Contemporary American English) based on 312 tokens which were selected from almost 5000 tokens with the formats PRONOUN + VERB + NOUN/PRONOUN + DET/ART + NOUN, including *to* variant forms like PRONOUN + VERB + DET/ART + NOUN + to + NOUN/PRONOUN. More specifically, they were chosen according to the meaning differences in the NP objects being concrete/abstract which had general verbs and also with different assumption of meanings arising from a single verb. They were observed to have a slight variation from a simple ditransitive construction i.e., properties involving idiomatic expressions, the animacy of the objects, and dative shift possibilities. In

addition, the selection was also based on the frequently occurring verbs - i.e., the *give*-type (*give, hand, lend, etc.*) and the *send*-type (*send, mail, forward, etc.*) verbs appearing in great numbers with this particular construction found in COCA.

The investigation revitalizes and refreshes some of the concepts that have been discussed earlier and also provides an extension towards them which may not have been discussed so far. For example, some conceptual understanding in sentences like (4), were one of the main attentions for this study.

(4) …when I was 18, one of my friends brought his guitar, and **he showed me some licks**

and played some chords… (COCA 2016 ACAD)

One may not understand what *licks* mean in this sentence unless he is aware of the different methods in playing guitars. These types of examples found in COCA have a specific target in explaining the whole depth of the sentence.

In addition, COCA gave the sentences like (5a) and (6a) with verbs such as *promise* which do not always undergo dative shift, having the *to* variant. Rappaport Hovav and Levin (2008) had also discussed properties similar to the one below, where they state a need of further study which would explain why one variant is sometimes appropriate, while the other one is not.

(5) a. …if **you promise me the world** and then turn your back… (COCA 2017 TV)

b. …if **you promise the world to me** and then turn your back…

(6) a. Check your e-mail and read the attachment. **I promise you no virus**…no Trojan

horse. (COCA 2014 FIC)

b. *Check your e-mail and read the attachment. **I promise no virus to you**…no Trojan

horse.

Further, the study discusses briefly about the ditransitive construction concepts including idiomatic expressions accepted by the real world such as *I gave you my word (COCA 1990 FIC)* which means *I promised you* were found to contribute to the addition of the construction's property. Also, while examining the number of instances in each case, the corpus also gave quite a lesser number of instances with object modifiers than expected such as *I sing you a **British** lullaby (COCA 2014 TV)*, the actual usage of the previously discussed meaning differences in speech and writing, etc, from various genres of the corpus have also been discussed briefly in this study. To summarize, the study is based upon real-life usage and how often even the simplest forms of the said construction are used by the people according to COCA. This gives a clear view of how a single ditransitive construction acts accordingly to the type of verbs and other syntactic properties, and extends towards various branches grammatically, paving way to new concepts which would be of a great aid for analysing in the future.

References

Dryer, Matthew S. 1986. Primary objects, secondary objects, and antidative. *Language* 62: 808-845.

Lakoff, George, and Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Malchukov, Andrej, Martin Haspelmath, and Bernard Comrie. 2007. Ditransitive constructions: A typological overview. *Studies in ditransitive constructions: A comparative handbook*. Berlin: De Gruyter Mouton.

Rappaport Hovav, Malka and Beth Levin. 2008. The English dative alternation: The case for verb sensitivity. *Journal of Linguistics* 48(1): 129-167.

Korean EFL learners' underuse of the English perfect forms

Kang San Noh and Sanghoun Song
(Korea University)

The main purpose of this study is to apply the collostructional analysis into the underuse behavior of Korean EFL learners. Underuse behavior refers to L2 learners' showing the symptom of avoidance of certain expressions or structures (Brown 2006). For instance, Hebrew EFL learners tend to avoid phrasal verbs and prefer one-word verbs because the phrasal verb system does not exist in Hebrew (Dagut and Laufer 1985). Likewise, Korean EFL learners might have difficulty learning the English perfect system and end up avoiding it. Some point out that Korean EFL learners do have difficulty acquiring the English perfect system as the perfect is concerned with both the past and the present (Lee 2011). With this in mind, this study intends to show how Korean EFL learners actually avoid the English perfect forms.

English perfect system is a complex morpho-syntactic construction made of an auxiliary ("have", "be") followed by a past participle (Ritz 2012). There have been a series of arguments on whether the perfect is a tense, an aspect, or both. Those who argue that the perfect is closer to a tense claim that the interpretation of "completion" or "result" are not intrinsic to the perfect itself (McCoard 1978). In contrast, Huddleston (1988) argues that the perfect falls under the category of an aspect. This study follows the traditional approach which considers the perfect as an aspectual system.

As for the methodology for analysis, this study uses the collostructional analysis. Collostructional analysis is a type of collocational analysis which investigates the association strength between the target structure and particular lexemes (Gries and Stefanowitsch 2003). The lexemes that are either attracted or repelled by the target structure are called *collexemes* (i.e. occur more frequently or less frequently). Collostructional analysis is applied to find out the collexemes of a collostruct which one intends to investigate. By doing so, the collostruction strength of the collexemes can be produced. Collostructional analysis is distinguished from the previous collocational analysis which focuses on raw frequencies of occurrence. Raw frequency itself can be ineffective to measure the association strength of collocates due to the following reasons. First, the most frequent collocates of any given constructions are generally function words which are not that significant semantically. Second, considering Ziph's law, the majority of collocates generally have a very low frequency. This is the reason why collostructional analysis intends to produce the collostruction strength, which is calculated via the Fisher-Yates exact test.

This study uses the two Korean EFL learners' corpora. The first one is the Gachon Learner Corpus (GLC, Carlstrom 2015, http://thegachonlearnercorpus.blogspot.kr) and the latter is the Yonsei English Learner Corpus (YELC, Rhee and Jung 2014). The GLC consists of 2,507,899 words and the YELC consists of 1,082,295 words, respectively. In order to compare the distribution between English and Korean corpus data, this study uses the COCA (Corpus of Contemporary American-English, Davies 2008-) as the source of a reference corpus.

Using the collostructional analysis, the list of the top 30 most repelled collexemes is given as in Table 1.

<Table 1> The List of the Most Repelled Collexemes

Rank	Lexeme	Collostruction Strength	Rank	Lexeme	Collostruction Strength
1	make	0	16	choose	3.60748E-50

2	have	4.40973E-188	17	find	3.93376E-49
3	get	2.96176E-174	18	study	7.26058E-49
4	give	2.56292E-132	19	cause	6.06613E-48
5	change	7.64912E-120	20	develop	8.11066E-48
6	use	8.59037E-120	21	ban	6.38881E-47
7	become	6.97179E-106	22	spend	2.83524E-45
8	do	8.53125E-91	23	avoid	5.87085E-44
9	go	7.67485E-86	24	grow	1.07351E-43
10	learn	5.3948E-67	25	come	3.29356E-43
11	lose	8.69772E-67	26	improve	2.10475E-39
12	think	1.98629E-66	27	affect	2.48103E-39
13	take	4.78807E-61	28	talk	4.28934E-39
14	help	3.59307E-55	29	agree	8.60098E-39
15	allow	3.79404E-55	30	struggle	3.51922E-38

Among the top 30 collexemes repelled by Korean EFL learners, it turns out that the 3 state verbs (e.g. *agree, have, think*) are found. The other 27 collexemes except for those aforementioned state verbs are dynamic verbs (e.g. *make, get, give, change, use, become, do, go, learn, lose, take, help, allow, choose, find, study, cause, develop, ban, spend, avoid, grow, come, improve, affect, talk, struggle*). In this way, the list of the repelled collexemes makes it possible to analyze how Korean EFL learners actually underuse the English perfect forms.

By applying the collostructional analysis into the corpus data, the list of the most repelled collexemes was produced. It is hoped that this study can provide the empirical and pedagogical contribution to the understanding of Korean EFL learners' underuse of the English perfect forms.

References

Brown, Douglas. 2007. *Principles of language learning and teaching*. New York: Pearson Education.

Carlstrom, Brian. 2015. The Gachon Learner Corpus. Available online at http://korean learnercorpusblog.kr/p/corpus.html

Davies, Mark. 2008-. The Corpus of Contemporary American English (COCA): 560 Million Words, 1990-Present. Available online at https://corpus. byu.edu/coca/.

Dagut, Menachem and Batia, Laufer. 1985. Avoidance of phrasal verbs: a case for contrastive analysis. *Studies in Second Language Acquisition* 7: 73-79.

Gries, Th. Stefan and Anatol, Stefanowitsch. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209-243.

Huddleston, Rodney. 1988. *English grammar: an outline*. Cambridge university press.

Lee, Ji-Hye. 2011. An effective grammar acquisition of English tense and aspect: focused on present perfect aspect and past tense. *Secondary English Education* 4(1): 33-60.

McCoard, Robert. William. 1978. *The English perfect: tense choice and pragmatic inferences*. Amsterdam: North Holland.

Rhee, Seok-Chae and Jung, Chae-Kwan. 2012. Yonsei English Learner Corpus (YELC). Proceedings of the First Yonsei English Corpus Symposium, 26-36.

Ritz, Marie-Eve. 2012. Perfect tense and aspect. In Robert I. Binnick, ed., *The Oxford Handbook of Tense and Aspect*, 881-907. Oxford university press.

How can 'what'-questions be interpreted as 'why'-questions?

Okgi Kim
(University of Wisconsin-Milwaukee)

1. Introduction This paper aims at addressing the research question of how 'what'-questions like that in (1) can be construed as 'why (the hell)'-questions asking causes/reasons in Korean:

(1) {mwe-l/way} kulehkey manhun nonmwun-ul ilk-ess-ni?
what-ACC/why so many paper-ACC read-PST-QUE
'Why (the hell) did you read so many papers?'

As indicated above, the replacement of *mwe-l* 'what-ACC' with *way* 'why', with the intended meaning intact, would lead one to assume that *mwe-l* itself may mean 'why'. However, this simple view is challenged by the fact that unlike *way*, *mwe-l* cannot stand alone to ask causes/reasons:

(2) A: I read 10 papers on *wh*-questions yesterday. B: Way 'why'/*Mwe-l 'what-ACC'?

Here I propose an alternative view, according to which the 'why'-question meaning is derived via the tight interaction between *mwe-l* and a CAUSE operator situated at the split-Force domain.

2. External merge of *mwe-l* in Spec-NP I analyze the nominal adjunct *mwe-l* in (1) as directly merged in Spec-NP, as in [$_{NP}$ mwe-l [$_{N'}$ [$_{AP}$ kulehkey manhun] nonmwun-ul]].[1] A compelling piece of evidence for the external merge of *mwe-l* in Spec-NP comes from the following ungrammatical examples regarding overt scrambling:

(3) a. *[kulehkey manhun nonmwun-ul] mwe-l ilk-ess-ni?
so many paper-ACC what-ACC read-PST-QUE
'(int.) Why did you read so many papers?'
b. *[mwe-l] ecey kulehkey manhun nonmwun-ul ilk-ess-ni?
what-ACC yesterday so many paper-ACC read-PST-QUE
'(int.) Why did you read so many papers?'

(3a) is ruled out because only the N′ constituent has scrambled, leaving *mwe-l* in its specifier position. (3b) is bad because *mwe-l* occupying Spec-NP has scrambled out of the NP, violating the Left Branch Constraint.

3. Double-CP structure McCloskey (2006) investigates why embedded inversion is available for question verbs like *wonder*, but not for resolutive verbs like *know*, as illustrated in (4).

(4) a. I wonder what should we do. b. *I usually know who might they hire.

As noted by McCloskey, T-to-C raising in embedded contexts is problematic under the view that T-to-C raising is licensed if and only if the target C-position is not selected by a lexical head: if we assume the embedded *wh*-clause in (4a) to be selected by *wonder*, T-to-C fronting should not happen, contrary to fact. In order to explain this and the contrast in (4), McCloskey proposes that whilst resolutive verbs take a single-CP complement, question verbs take a double-CP complement in which the higher C may correspond to Force in the sense of Rizzi (1997). On this view, the sentence in (4a) is assumed to involve the syntactic structure presented in (5).

(5) [$_{VP}$ wonder [$_{CP/ForceP}$ [$_{C/Force}$ Ø [$_{CP}$ [$_{C}$ should$_1$ [$_{TP}$ we t$_1$ do]]]]]]

[1] Following Ochi (1999) and Nakao & Obata (2009), I take *mwe-l* to have an 'inherent' Case, not a 'structural' Case.

Since the target C-position (i.e., lower C) is selected by the higher functional head C/Force, not by the lexical head *wonder*, the T-to-C fronting of the auxiliary *should* is allowed. Note that like embedded inversion, and unlike *way* 'why', the nominal adjunct *mwe-l* is compatible not with resolutive predicates but with question predicates:

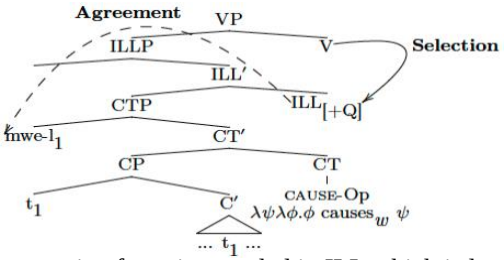
(6) a. na-nun [John-i mwe-l kulehkey manhun nonmwun-ul ilk-ess-nunci] kwungkumha/*an-ta.
I-TOP John-NOM what-ACC so many paper-ACC read-PST-QUE wonder/*know-DECL
'I wonder/know why John read so many papers.'

b. na-nun [John-i way kulehkey manhun nonmwun-ul ilk-ess-nunci] kwungkumha/an-ta.
I-TOP John-NOM why so many paper-ACC read-PST-QUE wonder/know-DECL
'I wonder/know why John read so many papers.'

To explain the contrast above, following McCloskey's insights, I take *mwe-l*-interrogatives to involve double-CP structures, with an additional assumption that Force, selecting a CP, is split up into C(lause)T(ype)P and ILL(ocutionary Force)P (following Coniglio & Zegrean (2012)).

4. Analysis I suggest that the well-formed *mwe-l*-interrogative in (6a) has the LF structure given in (7).

(7)



(here I adopt Shlonsky's (2021) *wh*-licensing system)

The interrogative force is encoded in ILL which is locally selected by (question) V. The ILL selects CTP headed by CAUSE-Op which denotes a function of the form $\lambda\psi\lambda\phi.\phi$ causes_w ψ, where ϕ is a proposition/event expressing the cause, and ψ is a proposition/event expressing the result that is presupposed to be true (cf. Dowty 1979). The nominal adjunct *mwe-l* bears two features [+wh] and [+Q]. Probe of [+wh] by C (question particle *nunci*/*ni*) leads *mwe-l* to move (from Spec-NP) to Spec-CP. Subsequently, *mwe-l* moves up to Spec-CTP, where it gets licensed by Agree with $\text{ILL}_{[+Q]}$, in order to *wh*-question the second argument of CAUSE-Op (i.e., the cause), characterizing its clause type as a *wh*-interrogative; the CP selected by CT serves as the first argument of the operator (i.e., the result).[2] On this analysis, the meaning of *mwe-l*-interrogatives can be paraphrased as "What is the proposition ϕ such that ϕ causes ψ?" For example, (1) is construed as "What is the proposition ϕ such that ϕ causes the past event of the addressee's reading so many papers?": the answer can be "I have a talk at LSK next week."

5. Conclusion The proposed analysis treats the nominal adjunct *mwe-l* as a canonical 'what' ranging over propositions, not as a special 'what' meaning 'why'. The tight interaction of *mwe-l* with CAUSE-Op situated in Force captures why, unlike *way*-interrogatives, *mwe-l*-interrogatives always involve the double-CP structure whose higher functional projection is (split-)ForceP.

Selected References

McCloskey, James. 2006. Questions and Questioning in a Local English. In Zanuttini, R., Campos, H., Herburger, E., Portner, P. (eds.), *Negation, Tense and Clausal Architecture: Cross-linguistic Investigations*, 86-126.

Nakao, Chizuru & Miki Obata. 2009. When 'What' Means 'Why': On Accusative *wh*-adjuncts in Japanese. *University of Pennsylvania Working Papers in Linguistics* 15(1): 18.

[2] I follow Tsai (2008) in arguing that both reason and causal questions involve a 'causal' relation between two events: reason Qs have a causal relation called Enable (one event is a necessary condition for the other: e.g., *Pasuya entered the pool, and then he drowned*.) while causal Qs have a causal relation called Cause (one event is a sufficient condition for the other: e.g., *It just snowed outside, so the snow is white*).

Processing the dynamicity of events in language

Sarah Hye-yeon Lee and Elsi Kaiser
(University of Southern California)

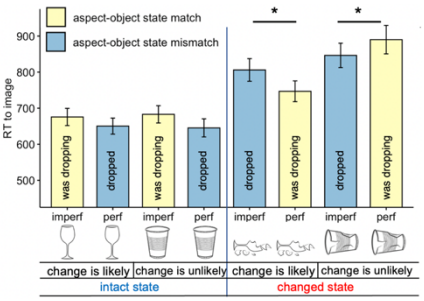
The world is a dynamically changing place, filled with different kinds of events. Humans' perception and conceptualization of events – 'things that happen' – is central for theories of human perception, action, language, and the human mind. Tracking of object states as they undergo change (e.g. intact onion → chopped onion, in an event of chopping an onion) is a fundamental mechanism in event comprehension (e.g. Altmann & Ekves 2019). Here we investigate how *temporal information encoded in grammar (Exp 1: grammatical aspect, Exp 2: tense)* and *real-world event knowledge* interact to modulate comprehenders' mental representations of object states. Grammatical markers about *time* provide information about whether an event is ongoing or completed (e.g. grammatical aspect: *is chopping*, *(has) chopped*) and whether it occurred in the past or future (e.g. tense: *chopped*, *will chop*) and are used during online processing to guide comprehenders' mental representations of object states (e.g. Altmann & Kamide 2007, Madden-Lombardi et al. 2017, Misersky et al. 2021). Event understanding is also tied to *real-world knowledge*. We possess knowledge of how events involving different objects normally take place in the real world (e.g. a wine glass is likely to break when dropped, but a plastic cup is less likely to break) and use this knowledge to guide event comprehension (e.g. McRae et al. 2005). What is less well-understood is how these two types of information interact during online comprehension. How do (i) grammatical aspect (Exp. 1) and (ii) tense (Exp. 2) affect the processing of events wherein the object's change-of-state (i) is likely (e.g. dropping of a wine glass) vs. (ii) unlikely (e.g. dropping of a plastic cup), according to event knowledge?

In Experiment 1, 114 participants read event descriptions in imperfective or perfective aspect (e.g. Kevin was dropping vs. Kevin dropped …). In Experiment 2, 107 participants read event descriptions in future or past tense (e.g. Kevin will drop vs. Kevin dropped …) (24 targets, 36 fillers in both experiments; 16 different verbs were used in targets; different objects in each item). In both studies, the object is either likely or unlikely to undergo a change-of-state as a result of the action (e.g. likely to break if dropped: wine glass vs. unlikely to break: plastic cup). Both studies used word-by-word self-paced reading with a rebus sentence paradigm, where some sentence constituents are presented as images (e.g. Madden & Therriault 2009, Madden-Lombardi et al. 2017). Crucially, this paradigm allows us to manipulate whether the image depicts the object in its intact (unchanged) state vs. its changed (resultant) state (e.g. intact wine glass vs. broken wine glass). The depicted change(s) were not entailed by verb semantics but were (highly or minimally) inferable from event knowledge.

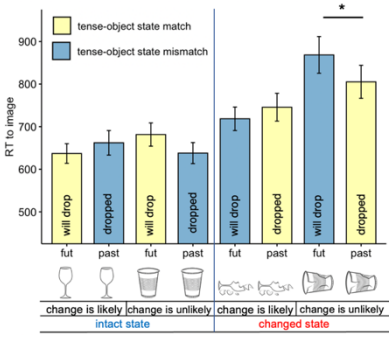
Exp. 1 ***Predictions*** Perfective aspect (e.g. *Kevin dropped …*) evokes a completed event representation with focus on the event's endpoint; imperfective aspect (e.g. *Kevin was dropping …*) evokes an ongoing event representation. We predict that perfective aspect will focus comprehenders' attention on state change. Focusing on state change will have different consequences for events in which the depicted change is likely vs. unlikely. With *likely-changes*, processing of changed state images will be faster in perfective aspect (when there is a match between the nature of the grammatical aspect and the object state) than in imperfective aspect. With *unlikely-changes*, attentional focus on state change will magnify the penalty (RT slowdown) associated with seeing an unlikely changed state, so RTs will be slower in perfective aspect (vs. imperfective aspect.)

Results We report analyses that compare the same image presented in different linguistic contexts. When processing intact object images, aspect and change-likeliness did not modulate RTs. When processing *changed*

object images: in *likely-change* conditions, the changed-state image was processed *faster* in *perfective* sentences than in *imperfective* sentences (t=2.59, p=0.02). However, in *unlikely-change* conditions, the changed-state image was processed *slower* in *perfective* sentences than in *imperfective* sentences (t=2.30, p=0.03), due to attention on state change magnifying the penalty of seeing an unlikely changed state.



Exp. 2 *Predictions* Tense temporally locates an event on a timeline with respect to the utterance time (Reichenbach 1947), without putting specific focus on any part of the event's internal structure. One might hypothesize that comprehenders may struggle with integrating a change-of-state image given a future tensed sentence, as future tense signals that the event is yet to occur. Does this type of 'tense-object state mismatch' penalty (RT slowdown) hold across both event types (likely-change & unlikely-change)? We predict that this penalty will be modulated by event knowledge. If the change is *likely*, change-of-state representations may be available to some degree even in future tense. However, if the change is *unlikely*, comprehenders will have difficulty integrating changed state images in future tense sentences.



Results When processing intact object images, tense did not modulate RTs (t=0.48, p=0.63). When processing changed object images in events where change was *likely*, image RTs in future tense and past tense sentences did not differ in a meaningful way (t=0.97, p=0.35), suggesting that change-of-state representations can be rapidly integrated in future tense sentences, as long as the depicted change is highly inferable. When change was *unlikely*, however, the changed object image was processed *slower* in *future tense* sentences than in *past tense* sentences. (t=2.95, p=0.008), suggesting an RT penalty associated in the case of 'tense-object state mismatch.'

Discussion Both experiments show that grammatical markers that encode temporal properties of events play a fundamental role in constraining and guiding event representations and activations of object states. We also show that the effects of grammatical information about an event's temporal properties are modulated by real-world event knowledge. We conclude that comprehenders rapidly integrate linguistic information with information rooted in the real-world to integrate visual information into the mental event representation.

References

Altmann, G., and Ekves, Z. 2019. Events as intersecting object histories: A new theory of event representation. *Psychological Review 126*(6).

Madden, C. J., and Therriault, D. J. 2009. Short article: Verb aspect and perceptual simulations. *Quarterly Journal of Experimental Psychology 62*(7): 1294-1303.

Madden-Lombardi, C., Dominey, P. F., and Ventre-Dominey, J. 2017. Grammatical verb aspect and event roles in sentence processing. *PLoS One 12*(12).

McRae, K., Hare, M., Elman, J. L., and Ferretti, T. 2005. A basis for generating expectancies for verbs from nouns. *Memory & Cognition 33*(7): 1174-1184.

Misersky, J., Slivac, K., Hagoort, P., and Flecken, M. 2021. The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition*.

The dative puzzle in Chinese revisited

Hyunjun Park
(Dongguk University)

Since Huang's (1982) work in Chinese, it has been widely assumed that scope in languages such as Chinese and Korean is basically determined by the surface word order. To be specific, only surface scope reading for quantifiers is allowed in the canonical word order, as shown in the examples of Chinese and Korean, respectively, in (1).

(1) a. You yi-ge xuesheng mai-le mei-ben shu.
have one-CL student buy-ASP every-CL book
'One student bought every book.' (∃>∀, *∀>∃) (Gan and Tsai 2020)
b. etten haksayng-i motun chayk-ul ilk-ess-ta.
some student-NOM every book-ACC read-PST-DECL
'Some student read every book.' (∃>∀, *∀>∃) (Kim 2015)

By looking closely, however, it can be observed that scope ambiguity occurs in dative constructions of Chinese, as illustrated in (2).

(2) a. Zhangsan mail-le liang-ben shu gei mei-ge ren. [Theme-Goal]
Zhangsan buy-ASP two-CL book GEI every-CL man
'Zhangsan bought two books to everyone.' (∃>∀, ∀>∃) (Huang 1982)
b. Wo song san-ben shu gei mei-ge ren. [Theme-Goal]
I give three-CL book GEI every-CL man
'I gave three books to everyone.' (∃>∀, ∀>∃) (Aoun and Li 1989)

In dative constructions, as can be seen from the above examples, the [Theme-Goal] orders reveal scope ambiguity whereas the [Goal-Theme] orders do not display it, indicating a scope asymmetry, as seen in (3).

(3) a. Zhangsan mai-gei-le mei-ge ren liang-ben shu le. [Goal-Theme]
Zhangsan buy-GEI-ASP every-CL man two-CL book ASP
'Zhangsan bought two books to everyone.' (*∃>∀, ∀>∃) (Huang 1982)
b. Wo song-gei mei-ge ren san-ben shu. [Goal-Theme]
I give-GEI every-CL person three-CL book
'I gave everyone every book.' (*∃>∀, ∀>∃)

In addition, as pointed out by Gan and Tsai (2020), when the Goal consists of the universal quantifier 'every' and the Theme includes the existential quantifiers such as 'two' and 'three', as in (4), the only available interpretation is the inverse scope, in which the Goal takes scope over the Theme.

(4) a. Wo song-le mei-yi-ben shu gei yi-wei laoshi.
I give-ASP every-one-CL book GEI one-CL teacher
'I gave every book to one teacher.' (∃>∀, *∀>∃)
b. Lisi juan-le mei-yi-fu hua gei yi-jia meishuguan.
Lisi donate-ASP every-one-CL painting GEI one-CL art.museum

'Lisi donated every painting to one art museum.' (∃>∀, *∀>∃) (Gan and Tsai 2020)

The patterns of quantifier scope interpretation in Chinese dative constructions are shown in (5).

(5) Quantifier scope interpretation in Chinese dative constructions

Quantifier	Word order	Goal > Theme	Theme > Goal
Goal: ∃, Theme: ∀	Goal-Theme	✔	✘
	Theme-Goal	✔	✘
Goal: ∀, Theme: ∃	Goal-Theme	✔	✔
	Theme-Goal	✔	✘

Based on this fact, we can see that only the Goal takes scope higher that the Theme or the wide scope interpretation of the Goal is strongly preferred in Chinese dative constructions. As noted by Aoun and Li (1989), this pattern is very similar to English. Hence, in this paper, adopting Janke and Neeleman's (2012) ascending/descending VP structures, we would like to suggest that Chinese dative constructions can be divided into right-to-left scope in ascending structure like (6a) and left-to-right scope in descending structure like (6b).

(6) a. [[V Q_1] Q_2] (rightward-ascending: Q_2>Q_1 >> Q_1>Q_2)
b. [V [Q_1 [t_v Q_2]]] (rightward-descending: Q_1>Q_2; *Q_2>Q_1) (Janke and Neeleman's 2012)

Under the analysis, a scope asymmetry in Chinese dative constructions becomes available in the following examples.

(7) a. Zhangsan [$_{VP}$ [$_{V'}$ mai-le liang-ben shu] gei mei-ge ren]. (Ascending VP)
'Zhangsan bought two books to everyone.' (∀>∃ >> ∃>∀)
b. Wo [$_{VP}$ [$_{V'}$ song-le mei-yi-ben shu] gei yi-wei laoshi].
'I gave every book to one teacher.' (∃>∀, *∀>∃)

(8) a. Zhangsan mai-gei-le [$_{VP}$ mei-ge ren [$_{V'}$ t_v liang-ben shu le]]. (Descending VP)
'Zhangsan bought everyone two books.' (*∃>∀, ∀>∃)
b. Wo song-le [$_{VP}$ mei-ge ren [$_{V'}$ t_v san-ben shu]].
'I gave everyone every book.' (*∃>∀, ∀>∃)

References

Bruening, Benjamin. 2010. Double object constructions disguised as prepositional datives. *Linguistic Inquiry* 41: 287-305.

Bruening, Benjamin. 2020. Implicit arguments in English double object constructions. *Natural Language and Linguistic Theory*

Gan, Tian and Cheng-Yu Edwin Tsai. 2020. The syntax of Mandarin dative alternation: An argument from quantifier scope interpretation. *International Journal of Chinese Linguistics* 7: 187-222.

Janke, Vikki and Ad Neeleman. 2012. Ascending and descending VPs in English. *Linguistic Inquiry* 43: 151-190.

Lee, Hanjung. 2020. Ditransitivity hierarchy, semantic compatibility and the realization of recipients in Korean dative construction. *Journal of Linguistics*: 1-36.

Kim, Lan. 2015. Asymmetric ditransitive constructions: Evidence from Korean. *Lingua* 165: 28-69.

Two types of semi-modal V-*to* in English

Lehan Yang
(Pusan National University)

The aim of this study is to provide a structural analysis of certain semi-modals that are based on a V-*to* construction, such as *have to*, *ought to*, and *need to*, as in the following sentences:

(1) a. Anna *ought to* eat vegetables.
b. Simon does not *have to* write a paper.
c. You *need to* study Chinese.

Semi-modals are transformed from main verbs as their meaning as a main verb gradually weakens and their modal meaning strengthens. However, as they retain characteristics of main verbs, they are called semi-modals. Banks (2019) adduces two types of relationships with speakers that the interpersonal metafunction encodes: mood and modality. He also proposes that modalization and modulation are two types of modalities: epistemic modality and deontic modality. Under his analysis, semi-modals, which are assimilating to the modal auxiliaries and can convey verbal meanings, express kinds of modality.
Previous studies have shown that semi-modals can be divided into diverse categories based on their particular semantic meanings and structural properties. Coates (1983) claims that semi-modals can be considered quasi-modals, while Quirk et al. (1985) divide semi-modals into semi-modals, modal idioms, semi-auxiliaries, marginal modals, and other categories. On the other hand, a few studies have proposed various syntactic analyses that divide those semi-modals into mono-clauses and bi-clauses. That is, some researchers deem the semi-modals as usually referring to a flat structure, as in the following data from Radford's analysis (2004):

(2) a. [CP [C Ø] [TP John [T ought] not [AUXP [AUX to] say anything]]]].
b. [CP [C Ø] [TP RPO [T to] not co-operate with the police]].

Contrariwise, Hong (2007) refers to a configuration with multiple layers rather than a flat structure, as in the following example:

(3) TP1__ [T' Ø [VP have [TP2 John$_k$ [T'2 to [VP study syntax]]]]]].

However, the existing analyses face certain difficulties. For example, in Radford (2004)'s mono-clausal analysis, the grammatical category of *to* is ambiguous. He proposes two possible positions (T and Aux) in the tree diagram of *ought to*. Similarly, Hong's (2007, 2014) view of bi-clausal analysis posits that V exists as a lexical verb for all semi-modal verbs taking a V-*to* construction. On the other hand, Radford (2004) considers that V can be treated as an auxiliary verb. In Huddleston and Pullum's (2002) opinion, auxiliaries and lexical verbs can be distinguished by the NICE properties: Negation, Inversion, Code, and Emphasis. For example, they regard the verb *need* as a lexical verb because lexical *need* often occurs in affirmative sentences, as in the following sentence with affirmative *need*:

(4) We *need to* consider both options.

Therefore, the present study anatomizes the inadequacies of existing syntactic analyses of semi-modals and proposes a structurally sophisticated analysis of semi-modals within a minimalist framework, adopting the erstwhile Split-TP Hypothesis and NICE Properties.

References

Banks, D. 2019. *A Systemic Functional Grammar of English: a simple introduction*. Routledge.

Coate, J. 1983. *The semantics of the Modal Auxiliaries [M]*. London & Canberra: Croom Helm.

Hong, S. 2007. On the Syntax of "Have To". *Studies in Generative Grammar.* 17.3: 397-413.

Hong, S. 2014. A Bi-clausal Account of English 'to'-Modal Auxiliary Verbs. *Language and Information.* 18.1: 33-52.

Huddleston, R., & Pullum, G. K. 2002. *The Cambridge Grammar of English Language*. Cambridge University Press.

Quirk, R. 1985. *A Comprehensive Grammar of the English Language.* Pearson Education India.

Radford, A. 2004. *Minimalist Syntax: Exploring the Structure of English.* Cambridge University Press.

A comparative analysis of certain double object constructions in Spanish and Korean

Nakyung Yoon
(Korea University)

In the literature there are two opposing approaches to dative alternation: a projection approach and a transform approach. Both approaches are related to the status of double object constructions (DOCs) and pre-/postpositional dative constructions (PDCs) in English. In this study, I discuss how in Spanish and Korean, certain DOCs can be categorized into two types – basic DOCs and derived DOCs (see (1)-(2) for Korean, (3)-(4) for Spanish). Specifically, I argue that based on nominalization and passivization data, both types of DOCs are syntactically derived from their variant of PDCs (see (5)-(8)). In the derived DOCs, benefactive dative arguments are introduced via a low applicative head, establishing the relationship between the two objects that implicates possession interpretation, as in Spanish, or via a high applicative head, in this case, not necessarily entailing successful transference of possession, as in Korean (see (9)-(10)). This analysis extends not only to Spanish (inflected languages), but also to Korean (agglutinative languages).

KOREAN

(1) a. [John-un$_{TOP}$] [Yenghi-eykey$_{DAT}$] [senmul-ul$_{ACC}$] **ponay**-ss-ta. PDC
John-$_{TOP}$ Yenghi-$_{ACC/-DAT}$ present-$_{ACC}$ send-$_{PST-DECL}$
'John sent a present to Yenghi'
b. [John-un$_{TOP}$] [Yenghi-lul$_{ACC}$] [senmul-ul$_{ACC}$] **ponay**-ss-ta. BASIC DOC
John-$_{TOP}$ Yenghi-$_{ACC/-DAT}$ present-$_{ACC}$ send-$_{PST-DECL}$
'John sent Yenghi a present'

(2) a. [John-un$_{TOP}$] [ai-eykey$_{DAT}$] [chayk-ul$_{ACC}$] **ilke-cwu**-ess-ta. PDC
John-$_{TOP}$ child-$_{ACC/-DAT}$ book-$_{ACC}$ read-$_{BEN-PST-DECL}$
'John read a book for a child'
b. [John-un$_{TOP}$] [ai-lul$_{ACC}$] [chayk-ul$_{ACC}$] **ilke-cwu**-ess-ta. DERIVED DOC
John-$_{TOP}$ child-$_{ACC}$ book-$_{ACC}$ read-$_{BEN-PST-DECL}$
'John read a child a book'

SPANISH

(3) a. Pablo (le) envió un diccionario a Gabi. PDC
Pablo $_{CL.DAT}$ sent a dictionary $_{A}$ Gabi
'Pablo sent a dictionary to Gabi.'
b. Pablo (le) envió un diccionario a Gabi. BASIC DOC
Pablo $_{CL.DAT}$ sent a dictionary $_{A}$ Gabi
'Pablo sent Gabi a dictionary.'

(4) a. Valeria (*le) diseñó una pollera a Andreína. PDC
Valeria $_{CL.DAT}$ designed a skirt $_{A}$ Andreína.$_{DAT}$
'Valeria designed a skirt for Andreína'
b. Valeria (*le) diseñó una pollera a Andreína. DERIVED DOC
Valeria $_{CL.DAT}$ designed a skirt $_{A}$ Andreína.$_{DAT}$
'Valeria designed Andreína a skirt'

KOREAN

(5) Thim-uy Chelswu-uy mapep-uy **kaluchi**-m.
Tim- GEN Chelswu- GEN magic- GEN teach- NML
'The teaching of magic Chelswu by Tim'

(6) *John-uy ai-uy chayk-uy **ilke-cwu**-m.
John-GEN child- GEN book- GEN read- BENE-NML
'The reading of the book the child by John'

(7) Chelswu-ka wuywenhoy-ey-uyhay sang-i cwue-ci-essta.
Chelswu-NOM committee-by prize-NOM give-PASS-PAST-DEC
'The award was given to Chelswu by the committee.'

(8) *ai-ka John- ey-uyhay chayk-i **ilke-cwu**-ci-ess-ta.
child-NOM. John-by book NOM read-BEN- PASS-PAST-DEC
'The book was read by the child.'

(9) a. ?Chulswu-ka ppalli Yonghi-lul steak-lul cwu-ess-ta.
Chulswu-NOM rápido Yonghi-ACC steak-ACC give-PST-DEC
'Chulswu quickly gave Yonghi the steak.'
b. Chulswu-ka ppalli Yonghi-ekye steak-lul cwu-ess-ta.
Chulswu-NOM rápido Yonghi-DAC steak-ACC give-PST-DEC
'Chulswu quickly gave the steak to Yonghi.'

(10) a. ??Chulswu-ka ppalli Yonghi-lul kabang-lul tul-e cwu-ess-ta.
Chulswu-NOM quickly Yonghi-ACC bag-ACC hold BEN-PST-DEC
'Chulswu quickly hold the bag for Yonghi.'
b. *Chulswu-ka ppalli Yonghi-ekye kabang-lul tul-e cwu-ess-ta.
Chulswu-NOM rápido Yonghi-DAC bag-ACC hold BEN-PST-DEC
'Chulswu quickly hold the bag for Yonghi.'

Selected References:

Cuervo, C. 2003a. *Datives at large*, doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Demonte, V. 1995. "Dative alternation in Spanish", *Probus*, 7, 5-30.

Harley, H. 2002. "Possession and the double object construction", *Linguistic variation yearbook*, 2(1): 31-70.

Kim, L. 2015. "Asymmetric ditransitive constructions: Evidence from Korean"*, Lingua,* 165, 28-69.

MacDonald, J. E. (2015). A movement analysis of some double object constructions. In *Proceedings of the 32nd West Coast Conference on Formal Linguistics* (pp. 276-285). Cascadilla Proceedings Project.

Oh, E. & M. L. Zubizarreta. 2010. "A comparison of the English and Korean double object", *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, CSLI Publications, Stanford, CA, 185-199.

Ormazabal, J. & J. Romero. 2010. "The derivation of dative alternations"*, Argument structure and syntactic relations: A cross-linguistic perspective*, 203-232.

Park, S-D. & J Whitman. 2003. "Direct movement passives in Korean and Japanese", *Japanese/Korean Lingustics12, 307-321*.

Pineda, A. 2020. "Double-Object Constructions in Romance", *Syntax,* 23(3), 203-240.

Yoon, J. 2015. "Double nominative and double accusative constructions", *The Handbook of Korean Linguistics*, 79-97.

Emergent constructional knowledge of the masked language model: A corpus-based approach to do-be construction

Un-sub Shin and Sanghoun Song
(Korea University)

Recent advances in artificial neural networks showed that probabilistic vector representations of corpora are increasingly sensitive to the acceptability of rarely occurring language data (Linzen & Baroni, 2021). The present study builds a corpus of statistically marginal do-be construction (DBC) (e.g., roughly one example out of ten thousand instances), such as *what you have to do is get ready* (Flickinger & Wasow, 2013). In addition, we evaluate a masked language model called BERT (Devlin et al., 2019) by examining how BERT scores the contrasts between the well-formed DBC and unacceptable one.

DBC is referred to as an instance of conventionalized knowledge about the language usage, characterized by the obligatory use of the verb *do* and copula *be*. In addition, only the highly restricted set of words, such as *thing*, *what* and *all*, can occupy the subject position of DBCs as shown in (1).

(1) a. All he$_i$ could do was spread *himself*$_i$ /**themselves*$_i$ out over a flat place.
b. First thing I try to *do*/**make* is suppress everything with sedatives.
c. What the fragment does *not*/**always* do is shed **any** light on the marital status of the historical Jesus.
d. *All*/**Everything* he does is make the other Democratic candidates look sensible.

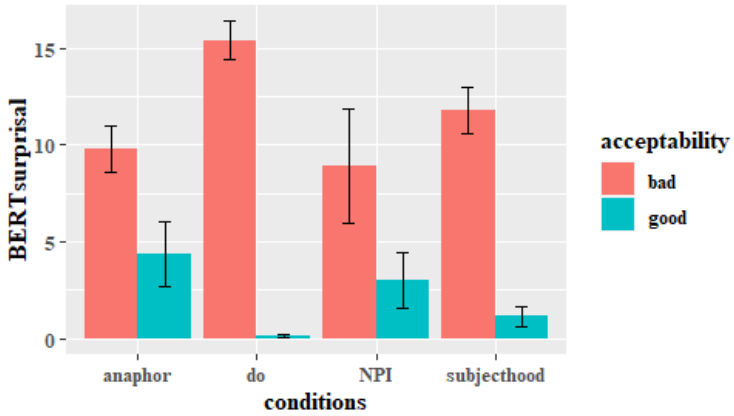
Notably, DBCs systematically constrain the distribution of anaphora and negative polarity items (NPIs), which counter the traditional conception of such linguistic phenomena. In (1a), the interpretation of the reflexive anaphor *himself* is dependent on the pronoun *he* in the subordinate clause, where the antecedent does not c-command *himself* (Reinhart, 1981). Similarly, NPI *any* in (1c) is licensed by a structurally lower word *not*, which constitutes a negative context but cannot scope over *any* in the matrix clause (Linebarger, 1987).

Importantly, DBC may test the ability of natural language processing systems that are potentially generalizable to peripheral but linguistically expressive data (Flickinger & Wasow, 2013). In these regards, we first aggregated arguably 17,737 DBC examples from COCA (Davies, 2009), GloWbE (Davies, 2013) and TV and Movie corpora (Davies, 2021) as shown in Table 1. We then extracted 20 instances of DBCs containing reflexive anaphora and 14 instances bearing NPIs for testing the performance of BERT. Furthermore, we also sampled 68 instances of DBCs, analogous to (1b) and (1d), to examine whether BERT captures the distributional property of DBCs. In our experiment, BERT is fed upon the whole sentences and calculated the output probability of target items *himself*, *do*, *not*, *all* and their counterparts. We then transformed the probability using surprisal estimation, i.e., higher surprisal may indicate an unacceptability of tested items due to an increased processing difficulty (Levy, 2008).

The results of our experiment using ANOVA shows that BERT clearly captures the difference between the well-formed and unacceptable DBCs in terms of surprisal rates ($p < .001$) (Figure 1). The result of our work suggests that BERT is sensitive to the syntactic licensing of reflexives and NPIs, and to the distributional restriction on the word occurrences in DBCs. Thus, we conclude that BERT encodes generalizable constructional knowledge required for processing marginal but linguistically essential language data like DBCs.

Table 1. Data descriptions of DBC corpus. Search pattern denotes the any form of verb *do*, *be* and any verb.

Search pattern: [vd*] [vb*] [v*]						
		base: [vvi*]	infinite: [vv0*]	present: [vvg*]	past: [vvn*]	Total
COCA	acad	76 (0.81)	10 (0.68)	25 (0.66)	31 (1.02)	**142**
	fic	121 (1.28)	10 (0.68)	16 (0.42)	37 (1.22)	**184**
	mag	180 (1.91)	19 (1.29)	34 (0.89)	58 (1.91)	**291**
	news	435 (1.46)	61 (4.15)	134 (3.51)	151 (4.99)	**781**
	spok	838 (8.89)	93 (6.33)	385 (10.10)	308 (10.17)	**1,624**
	Total	**1,650 (17.50)**	**193 (13.13)**	**594 (15.59)**	**585 (19.32)**	
GloWbE	ca	771 (8.18)	108 (7.35)	215 (5.64)	202 (6.67)	**1,296**
	gb	2,721 (28.86)	337 (22.93)	853 (22.38)	757 (25.00)	**4,668**
	ie	617 (6.54)	60 (4.08)	175 (4.59)	158 (5.22)	**1,010**
	us	2,823 (29.94)	402 (27.35)	913 (23.96)	901 (29.76)	**5,039**
	Total	**6,932 (73.52)**	**907 (61.70)**	**2,156 (56.57)**	**2,018 (66.64)**	
TV/Movie	0-19	**845 (8.96)**	**368 (25.03)**	**1,061 (27.84)**	**476 (15.72)**	
Total		**9,428 (%)**	**1,470 (%)**	**3,811 (%)**	**3,028 (%)**	**17,737**



<Figure 1> BERT's surprisal to grammatical conditions. Error bars indicate the 95% confidence intervals.

References

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2), 159-190.

Davies, M., & Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide*, 36(1), 1-28.

Davies, M. (2021). The TV and Movies corpora: Design, construction, and use. *International Journal of Corpus Linguistics*, 26(1), 10-37.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL*, 4171–4186.

Flickinger, D., & Wasow, T. (2013). A corpus-driven analysis of the do-be construction. *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 35-63.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

Linebarger, M. C. (1987). Negative polarity and grammatical representation. *Linguistics and philosophy*, 325-387.

Linzen, T., & Baroni. M. (2021). Syntactic structure from deep learning. *Annual Review of Linguistics*, 7, 195-212.

Reinhart, T. (1981). Definite NP anaphora and c-command domains. *Linguistic Inquiry*, 12(4), 605-635.

Progressives with dense and discrete interval verbs in English

Minji Kang
(Kyungpook National University)

It is well documented that stative verbs and achievement verbs do not come in the progressive form as in (a). This study aims to account for under what semantic condition(s) the progressive form is allowed to occur with stative and achievement verbs as in (2).

(1) a. *We are owning a house in a country. b. *I am recognizing a mistake.
c. *John is resembling his father.

(2) a. John is being polite. b. John is sitting on the chair.
c. John is winning the game. d. John is eating an apple.
e. John is resembling his father more and more as each day goes by.

To this end, the current study assumes that the event structure is mereological as in Krifka (1989), and intervals of eventualities of verbs are either dense or discrete. And the discrete intervals are again divided into two subcategories, *simplex atomic subinterval* and *complex atomic subinterval* as follows:

(3) a. $\forall P$ [simplex(P) $\leftrightarrow \forall t, t'[t, t' \in \text{Atom}(T_P) \rightarrow t=t']$ b. $\forall P$ [complex(P) $\leftrightarrow \forall t, t'[t, t' \in \text{Atom}(T_P) \rightarrow t \neq t']$

c. $\text{Atom}(T_P) = \lambda t\ [t \in T_P \wedge \neg \exists t'\ [t' \sqsubset t]]$, where T_P stands for the set of intervals which the entire eventuality of P takes over.

The present study contends that only *complex atomic subinterval* verbs can be used in progressive form, and stative and achievement verbs that can appear in progressive are characteristic of "complex atomic subinterval." As Dowty (1979) explains, the progressive *be waltzing* denotes subevents each of which consists of taking three steps. Likewise, (2d) can be said to denote three atomic subevents such as biting off a part of an apple, chewing it inside the mouth, and swallowing it. Since its interval structures are isomorphic to each other, verbs such as *waltz* and *eat* are complex atomic subinterval verbs (Krifka 1989).

Let us try to explain how stative verbs such as *be polite* and *sit(a stance*-predicate in terms of Dowty 1979) shift to atomic subinterval verbs. When stative verbs are used in progressive form, they denote temporary situations in which the subject is in. For example, (2a) means that John is temporarily pretending to be polite. Similarly, (2b) means that John is temporary on the chair, and might leave the chair soon. So John's being on the chair is more or less temporary (cf. similar explanation in Dowty 1979). Taking into account this temporariness, the current study calls such stative verbs "Temporary Discrete Interval" (hereafter TDI) verbs, and they can be defined as follows:

(4) TDI Verb P: $\forall P$ [TDI (P) $\Leftrightarrow \forall t$ [P(t) $\rightarrow \Diamond \exists t'$ [~P(t') $\wedge$ t $\leq$ t']], where '~P' means the state opposite to that of P.

On the contrary, pure stative verbs denote purely dense interval eventualities, defined as in (5)

(5) Pure Dense Interval Verb P: $\forall P$ [PDI (P) $\Leftrightarrow \forall t$ [P(t) $\rightarrow \exists t', t'', t'''$ [P(t') $\wedge$ t' $\sqsubset$ t $\wedge$ P(t'') $\wedge$ t$\sqsubset$t'' $\wedge$ P(t''') $\wedge$ t''' $\sqsubset$ t'' $\wedge \forall t''''[t'''' \sqsubset t''' \rightarrow t<t'''']]$

The possibility that the opposite interval might ensue and contrast with the interval of the eventuality of P can be considered a change or event. Relying on this sense of change, native speakers of English coerce them into complex atomic subinterval verbs and in turn, use them in progressive form (cf. Zucchi 1998). If a stative verb

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A2A01073540).

is a Pure Dense Interval Verb (hereafter PDI), they do not shift to complex atomic subinterval verbs so they cannot be used in progressive form.

For the coercion, the present study proposes an operator MEO, which coerces a stative predicate into an eventive one, very similar to the operator *PO* of Herweg (1991). MEO stands for *Makeshift Event Operator,* whose formal definition can be given as follows:

(6) MEO: λS, e[MEO(S)(e)]

a. ∀t[S-Phase(t) → τ(e)=t] b. ∀e[MEO(S)(e) → S-Phase (τ(e))] c. Discrete(τ(e))=1

Once a verb is or becomes an atomic subinterval one, the progressive operator can take it and yield a progressive form as in (2). For the semantics of the progressive operator "Prog", this study proposes (7a) (cf. Herweg 1979). And the progressive of "be being polite" can be formally represented as (7b).

(7) a. ⟦Prog⟧: λEλt ∃e[E(e) ∧ PROG(e,t)], where PROG(e,t) is defined as [t⊂τ(e) ∧ t⊂Atomic(τ(e))]

b. ⟦be being polite⟧=λt ∃e[MEO([polite(t) ∧ ◇∃t′[~polite(t′) ∧ t≤t′]])(e)] ∧ PROG(e,t)]

Lastly, as (1b) shows, achievement verbs are incompatible with progressive form. But verbs like (2b), *win, realize, die, etc.* can appear in progressive form. We can assume that a kind of preparatory phase may be attached right before the event so that it is not a moment anymore. *Start* is known to capture the preparatory stage of the main eventuality of its complement predicate, while *begin* is not. *He started to sneeze, but he didn't* is acceptable whereas *began* is not. (Tobin 1993).

Based on this, this study calls these Preparatory Stage Extendable (hereafter PSE) verbs and presents as follows:

(8) Preparatory Stage Extendable(hereafter PSE) Verbs P:

∀P[PSE (P) ⇔ ∀e, t [P(e)(t) → ◇∃t′, t″ e′ [P-act(e′)(t′) ∧ τ(e′) ∞ t ∧ Result-in (e′, e)(t)

∧ ¬∃t″ [t″⊂ τ(e′) ∧ t < t″], where ∞ is the "touch" relation between two intervals (cf. Krifka 2021).

As in (2e), the stative verb *resemble* can be shifted into PSE achievements with adjuncts such as *more and more.* The adjunct puts a preparatory stage, and the degree of resemblance increases as time goes by.

From the discussion above, we can conclude that any predicate can be used in the progressive form if they denote eventualities whose intervals have atomic subintervals.

References

Dowty, D. (1979) *Word Meaning and Montague Grammar.* Dordrecht: Kluwer Academic Publishers.

Dowty, D. (1986). The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics? *Linguistics and Philosophy*, 9(1), 37-61. Retrieved August 30, 2021, http://www.jstor.org/stable/25001231.

Herweg, Michael. (1991) Perfective and imperfective aspect and the theory of events and states. *Linguistics* 29: 969–1010.

Krifka, M. (2021). Individuating Matter over Time. In H. Filip (Ed.), Countability in Natural Language (pp. 121-144). Cambridge: Cambridge University Press. doi:10.1017/9781316823774.006.

Krifka, M. (1989). Nominal Reference and Temporal Constitution: Towards a Semantics of Quantity", in Jeroen J. Groenendijk, Martin Stokhof & Frank Veltman (eds.), *Proceedings of the 6th Amsterdam Colloquium*, Institute of Linguistic, Logic and Information, University of Amsterdam, 1987. 153-173.

Tobin, Y. (1993). *Aspect in the English Verb: Process and Result in Language*. London: Longman.

Zucchi, S. (1998) Aspect Shift. In: Rothstein S. (eds) *Events and Grammar*, 349-370. Studies in Linguistics and Philosophy, vol 70. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-3969-4_15.

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A2A01073540).

Projecting presupposition “not again” in Korean

Kihyo Park and Yunju Nam
(Konkuk University)

This research investigated how the native Korean speakers process the presupposed meaning triggered by “*tto*” (again) when *tto* is either embedded or unembedded in the negation scope of the two different negation forms in Korean. It has been argued that, while short-form negation “*an-*” only allows negative presupposition, the long-form negation “*-ci anh-*” allows both affirmative and negative presuppositions (Park 2001). However, little empirical research on whether and/or how this asymmetrical generation of the presupposed meaning affects language processing. Previous research (Park and Nam, 2020), using offline tasks, confirmed that Korean speakers tend to allow affirmative presupposition for the long-form negation. Despite this, the speakers preferred negative presupposed meaning as default. Notably, the main effects of context type (affirmative or negative) and negation form (long or short) and the interaction effect between them were reported in the research. But it has remained unclear whether the context supports immediate processing, or a default meaning is generated by the presupposition trigger and revised later to fit the context in real-time discourse processing. To explore this, we experimented with the online presupposition and projection problem using the eye-tracking method.

A 2 x 2 x 2 factorial design was constructed, where independent variables are the context type (affirmative or negative), the embedding of presupposition trigger *tto* (embedded or unembedded), and the negation form (long or short) (see Table 1). Total 48 items and 60 fillers were constructed. We recorded the eye movement of 34 Korean speakers. *First-Pass Reading times* (the sum of all fixation durations from first entering a region from the left before leaving it in any direction), *Second-Pass Reading times* (the sum of all fixation durations of the second run of fixations in a region), and *Total reading times* (the sum of all fixation durations in a region over the entire trial) were analyzed in VP as the region of interest (ROI).

From the linear mixed effects regression that we specified, a significant context effect (β = 39.96, S.E. = 15.72, t = 2.54, $p < .05$) in first-pass reading times, embedding effect (β = 48.82, S.E. = 18.59, t = 2.63, $p < .05$) in second-pass reading times and both of the effects in total reading times (context: β = 50.03, S.E. = 24.00, t = 2.09, $p < 0.05$; embedding: β = 68.90, SE = 24.00, t = 2.87, $p < 0.005$) were found. Critically, we found no significant negation form or the interaction effects, unlike the previous research with offline tasks.

Our results showed that the effect of context influences overall real-time processing stages as expected. The negation form effect was detected neither in the early-stage processing nor in the late-stage one, which is not in line with the previous offline study. On the other hand, the results suggested that the effect of presupposition trigger processing could be reflected in the late-stage processing. These findings partially support the Projection-Takes-Time (PTT) hypothesis (Schwarz and Tiemann, 2016). Table 2 provides that reading times were faster in the negative context than the ones in the affirmative context in general. Furthermore, the overall reading times became faster in the embedded condition than when the presupposition trigger is unembedded in the negation scope. Interestingly, the first-pass reading times in the negative context showed the opposite pattern. This might indicate that the speakers try to retrieve the negative context anaphor aligning with the presupposition content in the negated verbs. To sum, the current research provided partial evidence on the presupposition processing in Korean that a default meaning (negative presupposed meaning) is generated by the presupposition trigger *tto* + negation (not again) and then considering the given context meaning later. In or after this late stage, the possibility

of the affirmative presupposition of the *tto* + long-form negation might be calculated to establish the relevance between the context and the upcoming sentence with the *tto* + negation. Thus, the effect of the negation form would be reflected in the offline processing but not in the online one.

References

Park, Jong-gap. 2001. On the ambiguity of Korean negative sentences (2). *The Korean Language and Literature* 74, 1-20.

Park, Kihyo and Nam, Yunju. 2020. The psychological reality of presupposition "tto" in Korean negation types. *Korean Journal of Linguistics* 45(4): 843-862.

Schwarz, Florian and Tiemann, Sonja. 2016. Presupposition projection in online processing. *Journal of Semantics* 34(1): 61-106

Table 1. Conditions, stimuli with the region of interest (underlined)

(a) Affirmative context	
그저께 윤서는 친구들을 만나 탁구를 치러 갔다. 몸 상태도 좋았고 그녀는 탁구를 실컷 쳤다.	Two days ago, Yoonseo went to play Ping-pong with her friends. She felt well and played it a lot.
(b) Negative context	
그저께 윤서는 친구들을 만나 탁구를 치고 싶었다. 그러나 몸 상태가 좋지 않아 그녀는 탁구 치기를 포기했다.	Two days ago, Yoonseo wanted to play Ping-pong with her friends. But she did not feel well and gave up playing it.
(1) Short-Unembedded	어제 윤서는 **또 |탁구를 안 쳤다|** Yesterday Yoonseo again Ping-Pong not played "Yesterday Yoonseo did not play Ping-Pong again."
(2) Short-Embedded	어제 윤서는 **|탁구를 또 안 쳤다|** Yesterday Yoonseo Ping-Pong again not played "Yesterday Yoonseo did not play Ping-Pong again."
(3) Long-Unembedded	어제 윤서는 **또 |탁구를 치지 않았다|** Yesterday Yoonseo again Ping-Pong play not-did "Yesterday Yoonseo did not play Ping-Pong again."
(4) Long-Embedded	어제 윤서는 **|탁구를 또 치지 않았다|** Yesterday Yoonseo Ping-Pong again play not-did "Yesterday Yoonseo did not play Ping-Pong again."

Table 2. Summary statistics of eye-tracking reading times by condition

(a) Affirmative context		**Reading times (ms)**	(b) Negative context		**Reading times (ms)**
Measure	**Condition**	**Mean (SD)**	**Measure**	**Condition**	**Mean (SD)**
First-pass	Short-Unembedded	503 (326)	First-pass	Short-Unembedded	442 (279)
	Short-Embedded	494 (338)		Short-Embedded	474 (322)
	Long-Unembedded	474 (322)		Long-Unembedded	432 (274)
	Long-Embedded	442 (279)		Long-Embedded	458 (303)
Second-pass	Short-Unembedded	203 (376)	Second-pass	Short-Unembedded	235 (443)
	Short-Embedded	199 (342)		Short-Embedded	164 (329)
	Long-Unembedded	231 (447)		Long-Unembedded	212 (351)
	Long-Embedded	169 (300)		Long-Embedded	149 (298)
Total	Short-Unembedded	706 (502)	Total	Short-Unembedded	752 (545)
	Short-Embedded	693 (483)		Short-Embedded	613 (387)
	Long-Unembedded	705 (520)		Long-Unembedded	644 (417)
	Long-Embedded	611 (374)		Long-Embedded	607 (410)

Forward-looking effects in pronoun interpretation

Jina Song and Elsi Kaiser
(University of Southern California)

It is well-known that pronoun interpretation is closely related to building discourse coherence (e.g., [4], [6], [7]). One of the discourse-based models, Centering Theory (CT) (e.g.,[4]), provides a key intuition that multiple-pronoun resolution (ex.1a, 2a) is different from resolution of a single pronoun in a single clause (ex.1b, 2b) due to their differences in the referential structure – whether one or both referents from the preceding clause are re-mentioned. Although CT predicts that the referential structure influences pronoun interpretation by contributing to transition coherence in discourse, few studies have investigated the referential structure effects in pronoun resolution. Moreover, it has been argued that the degree of topicality/salience is different depending on thematic roles such that Agent and Experiencer are more topical than Stimulus and Patient (e.g., [8]). However, it has yet to be answered whether difference in topicality/salience levels of the thematic roles plays a role in establishing coherence transition in discourse and, thus, affects pronoun resolution. To this end, we report two studies that investigate i) whether referential structure guides pronoun resolution (e.g., 1-pronoun structure vs. 2-pronoun structure), and ii) whether the referential strcture effects are modulated by thematic roles.

Method. Exp1 (n=40) and Exp2 (n=60) manipulated verb type in the preceding clause and referential structure type in the pronoun-containing clause (ex.1, 2). In the pronoun-containing clause, we manipulated the **referential structure** by the object position being occupied by either another pronoun (2-pronoun: *he daxed him*) or a new name (1-pronoun: *he daxed Tom*). Nonce verbs were used in the pronoun-containing clause to minimize variability from verb semantics. Regarding the **verb type**, Exp1 tested Stimulus/Experiencer (SE/ ES) verbs while Exp2 tested Agent-Patient (AP) verbs in the preceding clause. Note that all these verbs are known to elicit Implicit causality (IC) bias such that a following subject-position pronoun tends to prefer either the subject antecedent (IC1 verbs, subject-biased) or the object antecedent (IC2 verbs, object-biased) depending on the IC verb type. (e.g., [1,2,3,5]). In Exp1, changes in IC bias are involved with a change in thematic roles. Stimulus-Experiencer (SE) verbs elicit a subject bias (IC1); Experiencer-Stimulus (ES) verbs elicit an object bias (IC2). On the contrary, in Exp2, changes in IC bias are not associated with any changes in thematic roles. Some AP verbs elicit a subject bias (IC1); other AP verbs elicit an object bias (IC2). Both experiments had 24 targets and 36 fillers. We conducted a picture-writing task (Fig.2), where people typed the names of each stick figure in the boxes so that the picture described the event of the underlined part of the sentence.

Predictions. The referential structure effect will be confirmed, if the subject-position pronouns are more likely to be interpreted as referring to the preceding object in the 1-pronoun condition compared to the 2-pronoun condition. According to a core idea of the CT, demoting a higher-salient referent (preceding subject) to the less privileged position (following object position), while promoting lower-salient referent (preceding object) to the privileged position (following subject position), results in a less coherent transition in discourse. Moreover, if the thematic role modulates the referential effects, in Exp1, SE verbs will be less susceptible to the differences in referential structure than those with ES verbs. Experiencer subjects (ES verbs) may disprefer to be demoted to the lower-privileged position more than Stimulus subjects (SE verbs) do, because topicality could boost the salience of the Experiencer subject. On the contrary, in Exp 2, there will be no differences between the AP_IC1 and the AP_IC2 verb condition because both types of AP verbs assign the same thematic roles to their arguments.

Results. We found main effects of verb type (p's<.05, glmer) and referential structure (p's<.001) both in Exp1 and Exp2 (Fig.1), indicating that pronoun interpretation is guided both by verb cues (more object interpretations with IC2 verbs than with IC1) and by referential structure (less object interpretations in 2-pronoun condition than in 1-pronoun). However, a significant interaction was found only in Exp 1 (p's<.01), but not in Exp 2 (p's>.3). As expected, while SE conditions showed weaker effects of referential structure than the ES conditions in Exp1, the referential structure effects were not different depending on verb type in Exp 2. It shows that the referential

effect is modulated by thematic roles. **In sum**, our results show that (i) the referential structure guide pronoun resolution by contributing to transition coherence in discourse and (ii) the referential effects are modulated by the thematic roles and their relative topicality.

(1) Examples of Exp1 (all-male name items (50%), all-female name items (50%))

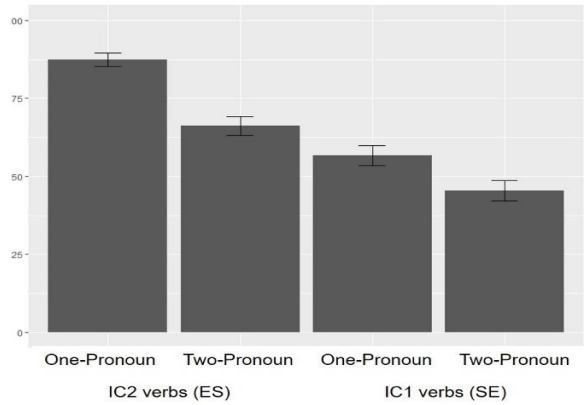
a. Bill {**surprised (SE $_{IC1}$) / hated (ES $_{IC2}$)**} Jack because **he daxed him**. [2-Pronoun]

b. Bill {**surprised (SE $_{IC1}$) / hated (ES $_{IC2}$)**} Jack because **he daxed Tom**. [1-Pronoun]

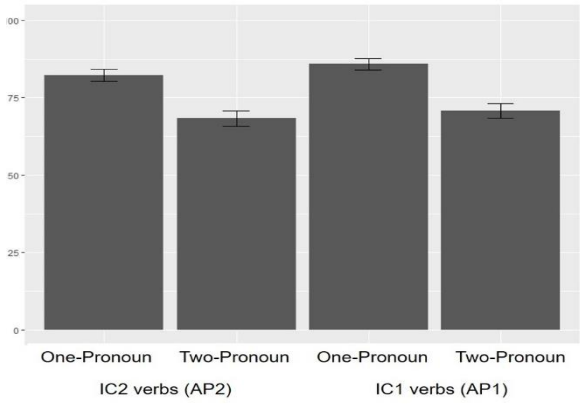
(2) Examples of Exp2 (all-male name items (50%), all-female name items (50%))

a. Bill {**cheated (AP$_{_IC1}$) / criticized (AP$_{_IC2}$)**} Jack because **he daxed him**. [2-Pronoun]

b. Bill {**cheated (AP$_{_IC1}$) / criticized (AP$_{_IC2}$)**} Jack because **he daxed Tom**. [1-Pronoun]



Exp1. ES vs. SE verb condition



Exp2. AP2 vs. AP1 verb condition

Figure 1. *Proportion of trials where the subject-position pronoun refers to preceding object.*

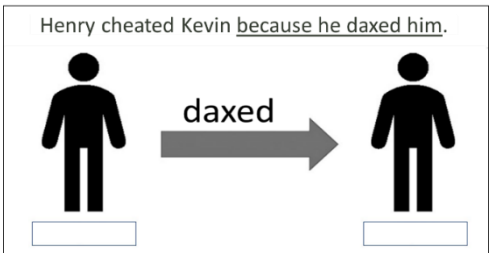


Figure 2a. *Sample Two-Pronoun stimulus.*

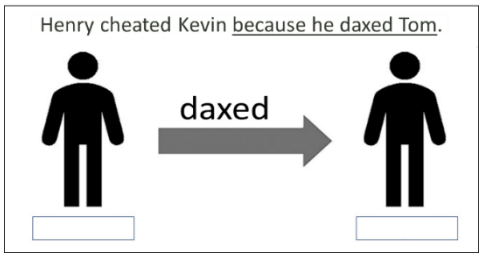


Figure 1. *Sample One-Pronoun stimulus.*

References

[1] Au, T. K. F. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of memory and language*, *25*(1), 104-122.

[2] Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of verbal learning and verbal behavior*, *16*(5), 601-609.

[3] Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods*, *43*(1), 124-135.

[4] Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). *Centering: A framework for modelling the local coherence of discourse.*

[5] Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*(10), 1474-1508.

[6] Kaiser, E. (2009). Investigating effects of structural and information-structural factors on pronoun resolution. *Information structure from different perspectives*, 332-353.

[7] Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI publications.

[8] Verhoeven, E. (2014). Thematic prominence and animacy asymmetries. Evidence from a cross-linguistic production study. *Lingua*, *143*, 129-161.

3. Conference Information

Biography of Plenary and Invited Speakers

Adele Goldberg (Princeton University)

Talk Title: Explain me this: Choosing good-enough constructions

Adele Goldberg has been a professor at Princeton University since 2004, initially in the linguistics program and currently in the psychology department. Her work focuses on the role of semantic, social, statistical, and memory-based factors in how languages are learned, represented, and used. She aims to gain a better understanding of human's creative but constrained language ability in adult sand children, L1 and L2, neurotypical and atypical populations. She is fascinated by many aspects of language including: word meaning, language change, island constraints, metaphor processing and emotion, good-enough language production, and the various functions of grammatical constructions. Goldberg is a fellow of the Linguistic Society of America, the Association of Psychological Science, and the Humboldt Foundation.

Manfred Krifka (Leibniz-Zentrum Allgemeine Sprachwissenschaft & Humboldt-Universität zu Berlin)

Talk Title: The layers of assertion: Propositions, judgements, commitments, acts

Manfred Krifka received his PhD in Munich, Germany, with a dissertation on aspectual classes and the mass/count distinction. He was professor at the University of Texas at Austin and is currently at Humboldt-University Berlin and director of the Leibniz-Centre General Linguistics. He has contributed to the topics of information structure, questions, vagueness, negation, and anaphora. Currently, he is working on languages in Vanuatu and on speech acts as PI of an ERC Advanced Grant, Speech Acts in Grammar and Discourse.

Sungbom Lee (Sogang University)

Talk Title: A macropragmatic approach to societal communication

Sungbom Lee is a professor of English at Sogang University in Seoul, Korea. He majored in English linguistics at Seoul National University and earned his MA in 1985. He received his PhD from Yale University in 1990 and taught at Hanyang University until 1997, when he moved to Sogang University. He was a visiting professor at UCSB in 2004 and at Nanjing University in 2019. Dr. Sungbom Lee has delivered keynote speeches and lectures on semantics and pragmatics in various places. His teaching passions include crosslinguistic pragmatics and intercultural communication, and his academic specializations are emotional expressions, aggressive speech acts, macropragmatics of language use and perception of defamatory utterances. He is the author of numerous scholarly books, articles and chapters and the recipient of the Korean National Academy of Sciences Award and the Ministry for Culture, Sports and Tourism for excellent books.

Seungshik Kang (Kookmin University)

Talk Title: New trends and major issues in deep-learning based approaches to language processing

Seungshik Kang is a professor of the Department of Computer Science at Kookmin University. He received his B.S, M.S and PhD from the Department of Computer Science and Engineering, Seoul National University. His research interests lie in Korean language processing, word embedding, text mining, and machine learning. He constructed large-scale Korean corpora and built a Korean word embedding model. He also developed a practical Korean morphological analyzer which is widely used by Korean NLP researchers. He served as the vice-president of the Korean Society of Information Sciences (2013-2016), and is currently the vice-president of the Linguistic Society of Korea. He is the author of 『Korean Morphological Analysis and Information Retrieval』 and 『Compiler and Automata』 (both books are written in Korean).

Eon-Suk Ko (Chosun University)

Talk Title: How we talk to infants and why it matters: Insights from Korean infants and their caregivers

Eon-Suk Ko is an Associate Professor at Chosun University. She holds a PhD in linguistics from University of Pennsylvania. Her doctoral research was in Korean phonology/phonetics, and she extended her research to child language acquisition at Brown University. Her current research programs include infants' perceptual and cognitive development, linguistic and developmental aspects of child-directed speech, and phonology/phonetics.

Abstracts of Plenary and Invited Talks

Explain me this: Choosing good-enough constructions

Adele E. Goldberg
(Princeton University)

What are children trying to do when they learn language? They aim to understand others' messages on the basis of witnessed formal cues, and aim to access patterns that will convey their intended messages. This requires them to learn a rich network of *constructions*, which are simply pairings of form and function, at varying levels of abstractness and complexity. What happens when an ideal combination of constructions is inaccessible? In this case, we all use a "good-enough" combination. I will explain how choosing good-enough constructions underlies a wide range of phenomena including: polysemy, language change, overextensions (labeling a cow, *dog*), grammatical overgeneralizations (e.g., *She laughed me*), L2 errors, and linguistic creativity. On this perspective, what makes an error an error is the existence of a different, more optimal combination of constructions.

The layers of assertion: Propositions, judgements, commitments, acts

Manfred Krifka
(Leibniz-Zentrum Allgemeine Sprachwissenschaft & Humboldt-Universität zu Berlin)

In this talk I will propose a fine-grained syntactic, semantic and pragmatic analysis of assertions. I will argue that what has been analyzed as the assertion operator applied to a proposition actually has three layers: a Judgement Phrase that can contain subjective epistemic, evidential, and evaluative operators that are outside of the proposition put forward; a Commitment Phrase that expresses the commitment of the speaker to the proposition and can contain operators such as *seriously*; and an Act Phrase that triggers a performative update of the common ground and can contain operators. I will argue that clause-embedding predicates are sensitive to these layers, e.g. *depend on* subcategorises for propositions, *think* for judgements, and *say* for commitments.

A macropragmatic approach to societal communication

Sungbom Lee
(Sogang University)

The objective of this presentation is to provide an overall view of 'societal communication' from a macropragmatic perspective. To begin with, the term 'societal communication' is used to include 'intergroup communication' and 'intercultural communication in Gudykunst (1993) and Ting-Toomey & Dorjee (2015). Societal communication refers broadly to a form of communication not so much between individuals in a micro context of speech but as between interlocutors that are conscious of or representative of a group or class each of them belongs to or identifies oneself with. Using some terminologies and principles of macropragmatics, we will explore several types of societal communication that are at issue in Korea nowadays for various reasons. They include communication between aged and young, communication between different genders, and communication between 'Kab', those who have power or privilege, and 'Ul', the underprivileged. The focus of our discussion is on the role played by 'noise' in societal communication that can interfere with a smooth and efficient flow of information or emotion in a verbal exchange. The important origins of those impediments are often 'the idols' in the sense of Francis Bacon that participants of societal communication have in their mind. We will discuss some real examples of such idols that can hinder and distort the meaning or interpretations of the message in various types of societal communication in Korea.

New trends and major issues in deep-learning based approaches to language processing

Seungshik Kang
(Kookmin University)

Natural language processing started from a rule-based method and developed into a probabilistic approach based on the construction of large-scale corpora. Recently, deep learning-based language processing is widely used. Much research which makes use of rule-based methods explores generative grammar focusing on the description of human knowledge as rules whereas statistical stochastic methods try to improve the performance of NLP systems by using the probability to which the rules are applied. Deep learning-based language processing has been developed on the basis of the construction of a large-scale text corpus. Accordingly, NLP technology is divided into the traditional symbolic NLP and the deep learning NLP. Traditional symbolic NLP methods develop rule-based systems based on human-readable symbols that express real-world objects from the perspective of symbolic AI, whereas deep learning methods are based on “sentence tokenization” and “word embedding”. Deep learning-based methods develop NLP systems by expressing sentences as vectors and composing them into document vectors. In this lecture, we discuss the sentence tokenization technique, which is a basic element of deep learning NLP, and the word embedding and sentence embedding technique as the basic method of constructing a document vector.

How we talk to infants and why it matters: Insights from Korean infants and their caregivers

Eon-Suk Ko
(Chosun University)

Across many cultures, speakers adapt their speech and language in some ways when talking to an infant. The extent of this modification, or the nature of such adaptation, however, is not fully understood. The research on the so-called infant-directed speech (IDS) and infants' processing of IDS has so far been mainly conducted on the WEIRD (Western, Educated, Industrialized, Rich and Democratic) population. In this talk, I present some findings from Korean infants and their caregivers that might help advance our understanding of the nature of IDS and the mechanism of how infants might process the input for learning.

I first present findings on Korean infants' preference for IDS over adult-directed speech (ADS). Studies on infants' word learning have shown that infants learn words better when they are presented in IDS and in ADS. This is likely because IDS holds infants' attention better than ADS, an effect which has been replicated in many previous studies. In the absence of a report on this issue from Korean infants, we tested Korean infants' preference for English and Korean presented in IDS and ADS for 6-9 month and 12-15 month old infants by tracking the duration of their head-turn towards the stimuli. Our results confirmed Korean infants' preference for IDS, with increased preference for the IDS in their native language in the older group.

What role then does IDS play in infants' language acquisition? Some argue that IDS carries a didactic function and directly facilitates infants' learning of language by providing enhanced features relevant to language learning. Others suggest that the benefit of IDS in language acquisition is mainly in its function to modulate infants' level of arousal and convey affection, which might help infants to maintain attention to the language input thereby facilitating language learning. We have been investigating aspects of Korean IDS at various levels of language to understand the extent to which IDS provides an enhanced language model directly relevant to language learning. I will present some of the findings from the analyses of Korean IDS, which would broaden our understanding of the nature of IDS.

The relevance of IDS to language development is not only in its quality, but also in the quantity to which an infant is exposed to it. We are investigating various factors that might be relevant to a child's experience of everyday language environments, and how they might affect children's vocabulary development. I present some of these findings with a focus on the benefits of introducing books during infancy. In a longitudinal study, we conducted a survey on how often the caregiver reads a book to their 7 month-old infant and made a day-long recording of all sounds that the infant is exposed to. When they were 14 months old, we tested their word recognition by tracking their eye gazes in the Preferential Looking Paradigm. We found a correlation between the frequency of book reading and the accuracy of word recognition. Crucially, the effect of book reading on word recognition was mediated by the amount of IDS, but not ADS, which highlights the importance of the amount of IDS as input to the language-learning children.

I close the talk with a brief discussion on the implications of these findings for the evidence-based intervention strategies in parenting, and the development of AI that reverse-engineers the human brain to mimic infants' learning.