Context window and polysemy interpretation: A case of Korean adverbial postposition –*(u)lo*

Université Paris Nanterre

Seongmin Mun[1] & Gyu-Ho Shin[2]
Université Paris Ouest Nanterre La Défense[1], University of Hawaii at Manoa[2]

Introduction

- **Korean postposition –*(u)lo*** has three major functions
  - INS (instrumental), FNS (final state), DIR (directional)

(1b) **–*(u)lo* as INS (instrument)**

na-nun	kamca-lul	khal-lo	ssel-ess-ta.
I-TOP	potato-ACC	knife-INS	cut-PST-DECL

'I cut a potato with a knife.'

(1a) **–*(u)lo* as FNS (final state)**

Yongho-lul	pancang-ulo	senchwul.ha-ass-ta.
Yongho-ACC	class.leader-FNS	election.do-PST-DECL

'(We) elected Yongho as the leader of the class.'

(1c) **–*(u)lo* as DIR (direction)**

Mia-ka	mikwuk-ulo	ttena-ss-ta.
Mia-NOM	America-DIR	leave-PST-DECL

'Milan left to America.'

- **Assumption:** construal of a polysemous word occurs in conjunction with a series of words, delivering various frame-semantic meanings (Goldberg, 2006) and yet purporting similar interpretations (Harris, 1954)
- **Context window:** a range of words surrounding a target word, affecting the determination of its characteristics
- **Question:** *how does context window address polysemy interpretation in Korean?*

Methods [1]

- **Input:** A portion of Sejong corpus (Shin, 2008), with semantic annotations of –*(u)lo* cross-verified by three native speakers of Korean (*κ* = 0.95)
  - Training set: 1,890 sentences
  - Test set: 210 sentences

밭/NNG 에서/JKB 채소/NNG **(으)로_INS/JKB** 가꾸/VV 다/EF
슬로/NNG 모션/NNG **(으)로_FNS/JKB** 보이__01/VV 다/EF
우리/NP 그만/MAG 포항/NNP **(으)로_DIR/JKB** 가/VV 자/EF

- **Model training:** Adapting a distributional semantic model (Harris,1954), an unsupervised learning algorithm was devised by combining Singular Value Decomposition with Positive Pointwise Mutual Information
- **Classification:** similarity-based estimate (Dagan et al., 1993) by calculating cosine similarity scores between *-(u)lo* and its co-occurring content words

Abbreviation: ACC = accusative case marker; DAT = dative marker; DECL = declarative; EF = final ending; JKB = adverbial case marker; MAG = general adverb; NNG = common noun; NNP = proper noun; NOM = nominative case marker; NP = pronoun; PST = past tense marker; TOP = topic; VV = verb;

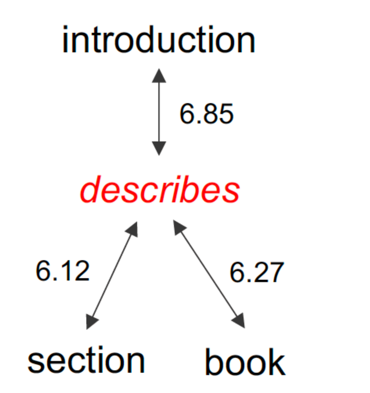
REFERENCES

- Bullinaria, John A & Levy, Joseph P. 2007. Extracting semantic representations from word co- occurrence statistics: A computational study. *Behavior Research Methods*, 39(3), 510–526.
- Dagan, Ido, Marcus, Shaul, & Markovitch, Shaul. 1993. Contextual word similarity and estimation from sparse data. *In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Columbus, 164-171.
- Goldberg, Adele. E. 2006. Constructions at work: The nature of generalization in language. *Oxford: Oxford University Press.*
- Harris, Zellig S. 1954. Distributional Structure. *WORD*, 10(2-3), 146-162.
- Patel, Malti, Bullinaria, John A. & Levy, Joseph P. 1997. Extracting semantic representations from large text corpora. *Proceedings of the 4th Neural Computation and Psychology Workshop*, London, 199–212.
- Shin, Hyo-Pil. 2008. The 21st sejong project : with a focus on selk(sejong electronic lexicon of korean) and the knc(korean national corpus). *In In The 3rd International Joint Conference on Natural Langauge Processing.*

Methods [2]

Similarity-based estimate (Dagan et al., 1993)

Network from the training set



Input as a test item

[introduction, *chapter* (unknown), book, section]

Q: How to calculate the similarity score between '*describes*' and '*chapter*'?

describes ←?→ *chapter* (unknown)

(w_1, w_2)	$\hat{I}(w_1, w_2)$	$f(w_1, w_2)$	$f(w_1)$	$f(w_2)$
(introduction, describes)	6.85	5	464	277
(book, describes)	6.27	13	1800	277
(section, describes)	6.12	6	923	277
Average:	6.41			

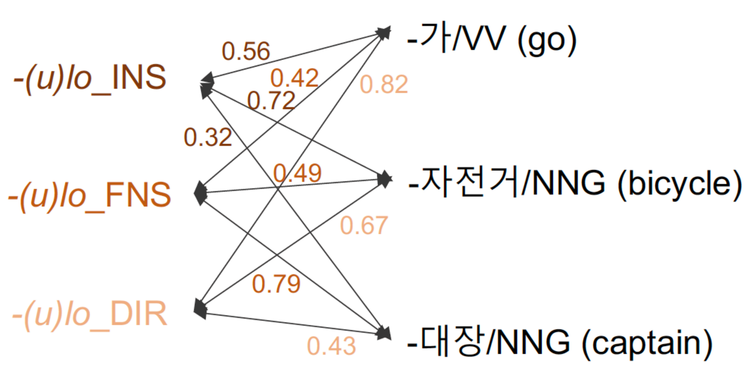
Table 1: The similarity based estimate as an average on similar pairs: $\check{I}(chapter, describes) = 6.41$

describes ←6.41→ *chapter* (unknown)

Our approach (adapted from Dagan et al., 1993)

Three functions of –*(u)lo* : INS (instrumental), FNS (final state), DIR (directional)

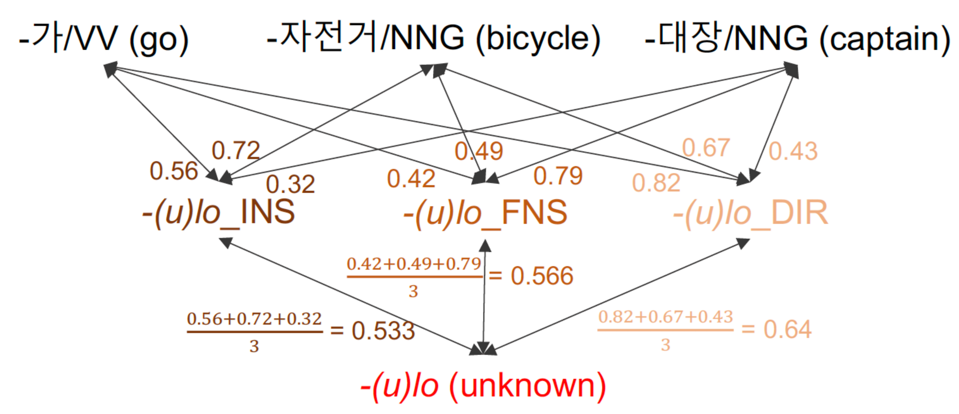
Network from the training set (window size: 1; normalized cosine)



Input as a test item

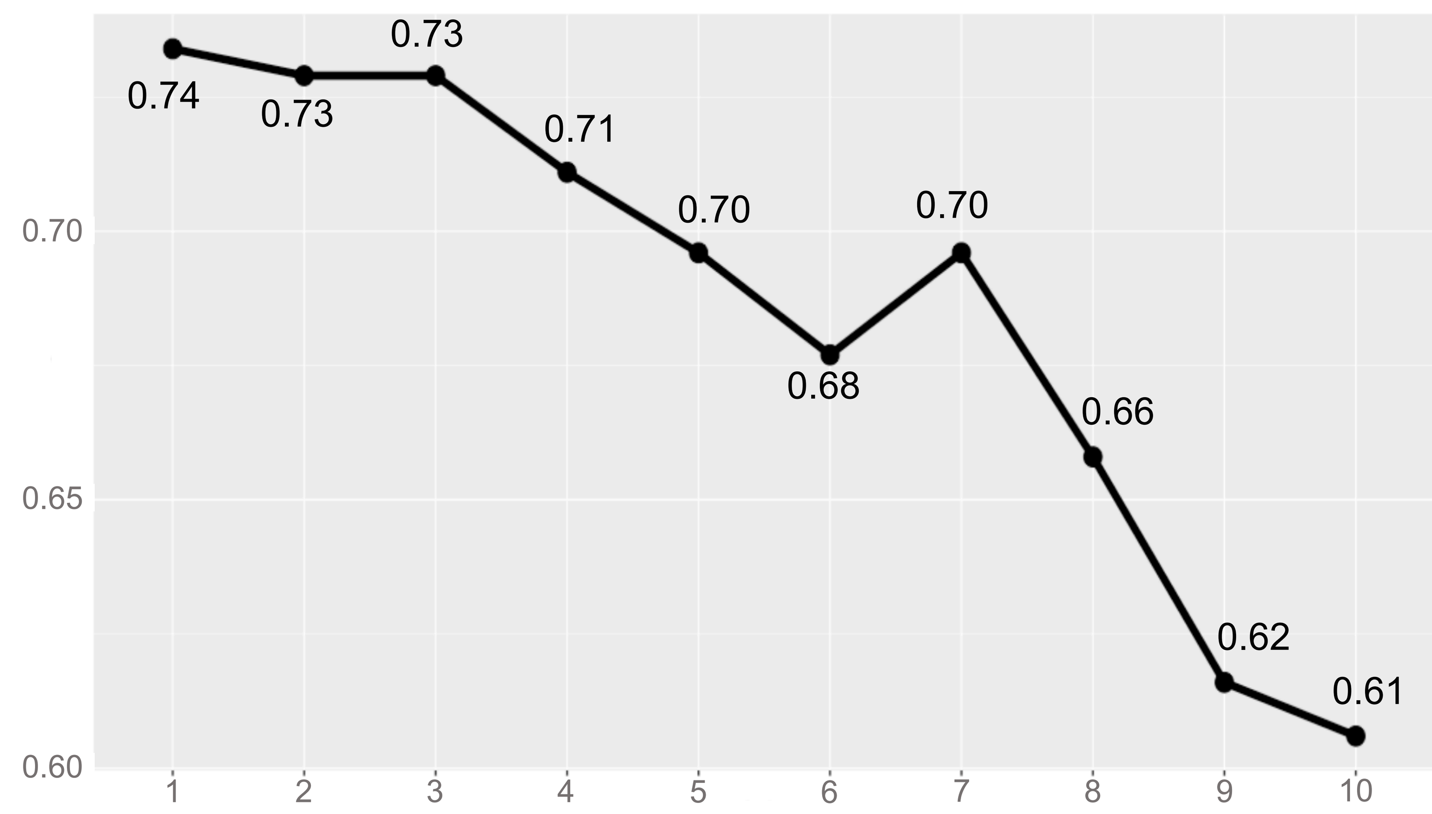
[-가/VV (go), -*(u)lo* (unknown), -자전거/NNG (bicycle), -대장/NNG (captain)]

Q: Which function is the intended function of -*(u)lo*?



-*(u)lo*_INS: 0.533
-*(u)lo*_FNS: 0.566 ⇨ -*(u)lo* (unknown) ⇨ **DIR**
-*(u)lo*_DIR: **0.64**

Results



X-axis: window size
Y-axis: accuracy (%)

- Our model achieved the highest classification accuracy rate in the window size of one, and the accuracy rates decreased as the window size increased
- **Interpretation**
  - ✓ This trend aligns with advantages of small window sizes (Bullinaria & Levy, 2007)
  - ✓ Considering that a narrower range of context window relates more to syntactic than to semantic information (Patel et al., 1997), our model may have employed structural, more than semantic, characteristics of tri-grams (word-target-word) for the best classification performance

Presented at the IMPRS Conference 2020 (03-05 JUN 2020)

Corresponding Author: Seongmin Mun(*simon.seongmin.mun@gmail.com*)