Site Update: Korea University

Sanghoun Song and others
(sanghoun@korea.ac.kr)

Korea University

2021 July 19

Research Fellows

- Eunkyung Yi
  - SUNY at Buffalo PhD (Jean-Pierre Koenig)
  - Deep learning's ability of **discourse expectation**
  - taught HPSG (2020 Fall semester)
  - Now, a tenure-track at Ewha Womans University
- Kyeong-min Kim
  - Simon Fraser University PhD (Chung-hye Han)
  - Experimental syntax & corpus analysis
    - Relative clauses in Korean
    - Unbound reflexives in English

Students

- Guehyun Wang (MA, completed)
  - Paraphrasing Training Data for Neural Machine Translation (`wgh218@korea.ac.kr`)
  - With respect to the data augmentation, English sentences were augmented by paraphrasing using English Resource Grammar (ERG), a resource grammar-based Head-Driven Phrase Structure Grammar (HPSG), and Korean sentences were augmented by paraphrasing using a pretrained language model. This study trained a total of 40 machine translation models with differ data sizes, paraphrasing methods, model architectures, and sources to target language pairs (EnKo and KoEn).

Students (cont'd)

- Gyu-min Lee (MA, on-going)
  - HPSG/MRS-Based Natural Language Generation Using Transformer
  - a short presentation on **Wed.** / a poster presentation at **HPSG 2021**
  - With the massive success of stochastic language models, some claim that linguistics is not necessary for natural language processing. However, recent studies made use of linguistically motivated meaning representation of abstract meaning representation and minimal recursion semantics in natural language generation task and achieved impressive result while hinting that neural networks do make extensive use of linguistic information. This study replicates previous studies by applying Transformer.

Students (cont'd)

- Kwonsik Park (PhD, on-going): a computational approach to nativelikeness
    - A Deep Learning-based Understanding of Nativelikeness: A Linguistic Perspective
    - Deep learning can contrast the minimal pairs of syntactic data
- Unsub Shin (MA, on-going): computational psycholinguistics
    - BERT seldom suffers NPI illusion.
- Seongtae Kim (MA, on-going): a probing test
    - Embedding Calculus with Nonword Properties Improves Word Sense Disambiguation.
- Kang San Noh (MA, on-going): corpus-based L2 studies
    - How Korean EFL Learners Use the Progressives: A Collostructional Analys

Students (cont'd)

- Junghyun Lee (MA, on-going): A Coronavirus corpus
- Hayul Park (MA, on-going): HateSpeech detection
- Daeun Kang (MA, on-going): Accusative-oblique alternations in Korean
- Jimin Lee (MA, on-going): Ditranstive verbs in Korean

Projects

- K-GLUE benchmark (Korean General Language Understanding Evaluation)
  - KoLA (Korean Corpus of Linguistic Acceptability)
  - STS (Semantic Textual Similarity)
  - NLI (Natural Language Inference)
- Medical NLP
  - Korea University Research Institute for **Medical Bigdata Science**
  - Hyung Joon Joo, M.D., Ph.D / Division of Cardiology
  - A Word Pair Dataset for Semantic Similarity and Relatedness in Korean Medical Vocabulary: Reference Development and Validation (JMIR Medical Informatics, 2021 Jun.)
  - KM-BERT: A Pre-trained BERT for Korean Medical Natural Language Processing (decision in progress)

Grants

- Korean Commitmentbank (PI)
  - Building a testsuite of natural language inference
  - National Institute of Korean Language (about 111,000 USD)
- Computational linguistic analysis and implementation toward the Coronavirus crisis (PI)
  - Building a Coronavirus corpus in Korean (2020-2022)
  - Korean Research Foundation (about 22,000 USD)
- An interdisplinary study of North Korean text data
  - A joint work with the Institute for Far Eastern Studies (about 20,000 USD)