

Learning Language Structures through Grounding

Freda Shi

University of Waterloo
Vector Institute, Canada CIFAR AI Chair
fhs@uwaterloo.ca

Abstract

Language is highly structured, with syntactic and semantic structures, to some extent, agreed upon by speakers. With implicit or explicit awareness of such structures, humans can learn and use language efficiently and generalize to sentences that contain unseen words. Motivated by human language learning, in this presentation, I will introduce a family of machine learning tasks that learns language structures through grounding, where distant supervision from other data sources (i.e., grounds), including but not limited to different modalities (e.g., vision), execution results of programs, and other languages, are used to guide the learning of language structures. I will demonstrate the potential of this task formulation, advocate for its adoption through three schemes, and discuss the possibility of the general language learning problem through grounding.

Humans can learn language naturally and efficiently, as well as using natural language to interact with the world. Language structures, such as syntactic and semantic parses of sentences, play an important role in such processes: with awareness of structure, humans can judge whether a sentence is grammatical, compose sentence meaning, and produce grammatical sentences describing new objects and events. Even more impressively, though humans implicitly develop and use structure for language processing in their daily communication, the explicit structure of sentences is almost never given. Drawing inspiration from human language learning, in this presentation, I will introduce my work on learning language structures through grounding, from the following three perspectives:

Part I: Learning syntax from visual grounding. Language is rarely text in isolation: humans learn and use language through interaction with others and the world. Inspired by this, I introduce the task of visually grounded grammar induction and develop the first promising system on the task (Shi et al. 2019), as well as follow-up work extending the model to induce grammar from visually grounded speech without text annotations (Lai et al. 2023).

Part II: Learning semantics through execution. Meanings of natural language can be represented by translating them into executable programs, and can usually be grounded

into the real world through execution. In this part, I will introduce our work on joint induction of syntax and semantics (Mao et al. 2021), as well as the execution-based minimum Bayes risk decoding algorithm for natural language to code translation using large language models (Shi et al. 2022).

Part III: Learning syntax from cross-lingual grounding. There are more than 7,000 languages all over the world. While models have reached high performance on high-resource languages, a universal NLP system should be able to process and understand other languages as well. I will introduce our work on learning cross-lingual word alignment (Shi, Zettlemoyer, and Wang 2021) and syntax (Shi, Gimpel, and Livescu 2022) from cross-lingual grounding. Going beyond, I will also cover our recent work on cross-lingual reasoning with large language models (Shi et al. 2023).

The presentation will conclude with a discussion on the potential of the general problem of learning language, including but not limited to structures, through grounding.

References

- Lai, C.-I. J.; Shi, F.; Peng, P.; Kim, Y.; Gimpel, K.; Chang, S.; Chuang, Y.-S.; Bhati, S.; Cox, D.; Harwath, D.; Zhang, Y.; Livescu, K.; and Glass, J. 2023. Audio-Visual Neural Syntax Acquisition. In *ASRU*.
- Mao, J.; Shi, F.; Wu, J.; Levy, R. P.; and Tenenbaum, J. B. 2021. Grammar-Based Grounded Lexicon Learning. In *NeurIPS*.
- Shi, F.; Fried, D.; Ghazvininejad, M.; Zettlemoyer, L.; and Wang, S. I. 2022. Natural Language to Code Translation with Execution. In *EMNLP*.
- Shi, F.; Gimpel, K.; and Livescu, K. 2022. Substructure Distribution Projection for Zero-Shot Cross-Lingual Dependency Parsing. In *ACL*.
- Shi, F.; Suzgun, M.; Freitag, M.; Wang, X.; Srivats, S.; Vosoughi, S.; Chung, H. W.; Tay, Y.; Ruder, S.; Zhou, D.; Das, D.; and Wei, J. 2023. Language models are multilingual chain-of-thought reasoners. In *ICLR*.
- Shi, H.; Mao, J.; Gimpel, K.; and Livescu, K. 2019. Visually Grounded Neural Syntax Acquisition. In *ACL*.
- Shi, H.; Zettlemoyer, L.; and Wang, S. I. 2021. Bilingual Lexicon Induction via Unsupervised Bitext Construction and Word Alignment. In *ACL*.