Linguistic Structure and Language Models: An Annotated Bibliography

Suchir Salhan

Michaelmas 2025

Motivation

The *Introduction to Natural Language Syntax & Parsing* course explores the intersections between linguistic theory and computational modelling. It motivates why linguistics remains central to the study of language technology. [1]

LLMs are not a silver bullet. Linguistics provides theoretical structure, methodological discipline, and typological diversity that we can use to understand the capabilities and limitations of LLMs, and how we can improve the linguistic capabilities of models. Despite the sophistication of large language models (LLMs), linguistic theory continues to offer indispensable insights for interpretability, evaluation, and understanding of cross-linguistic structure.

[1] Handout prepared by Suchir Salhan for L95 Introduction to Natural Language Syntax and Parsing

Motivating Paper:
Ferrer-i-Cancho, R. & Solé, R. (2024). *Natural Language Processing relies on Linguistics.*
arXiv:2405.05966.

Areas of Study

- **Interpretability:** How do LLMs encode syntactic, morphological structure?
- **Resources:** Linguistically informed datasets and benchmarks to evaluate linguistic capabilities of models.
- **Evaluation:** Designing controlled experiments and minimal pairs in English and beyond.
- **Low-resource languages:** Evaluating cross-linguistic generalisation and typological gaps.
- **Study of language:** Grammar formalisms, morphological systems. Applications to Parsing Tasks in Computational Linguistics and Language Models tokenization.

Linguistics and Language Models – General Readings

Lots of "positional" works on LLMs and Linguistics [some reviewed in my CoPiL position paper: Section 3].

This sub-literature was sparked by a provocative position paper by Steve Piantadosi:

- **Piantadosi, S. T. (2023). Modern language models refute Chomskys approach to language. From fieldwork to linguistic theory: A tribute to Dan Everett, 15, 353414.** link
- Juri Optiz. Natural Language Processing RELIES on Linguistics
- **Millière, R. & C. Rathkopf. (2024). Anthropocentric bias and the possibility of artificial cognition. ICML 2024 Workshop on LLMs and Cognition**
- Fox, D. & Katzir, R. (2024). Large language models and theoretical linguistics. Theoretical Linguistics 50(1-2), 7176. link [Generativist/Chomskyan Response]

General Readings

- **Millière, R. (2024). Language models as models of language.** link
- Cognitive Plausibility in NLP (Lisa Beinborn & Nora Hollenstein, 2023) link

Technical Expositions – Language Models, Probing, Interpretability, Tokenizers

- Belinkov, Y. (2022). Probing Classifiers: Promises, Shortcomings, and Advances. Computational Linguistics 48(1), 207219. link
- Ferrando, J., Sarti, G., Bisazza, A., Costa-Jussà, M. R. (2024). A primer on the inner workings of transformer-based language models. link
- Doddapaneni, S., Ramesh, G., Khapra, M., Kunchukuttan, A., Kumar, P. (2025). A primer on pretrained multilingual language models. ACM Computing Surveys, 57(9), 139. link

Introductory Tokenization / Morphology Papers

- Gutierrez-Vasques, X., Bentz, C., Samardi, T. (2023). Languages Through the Looking Glass of BPE Compression. Computational Linguistics 49(4), 9431001. link
- Schmidt, C. W., et al. (2024). Tokenization Is More Than Compression. EMNLP 2024, 678702.
- Uzan, O., Schmidt, C. W., Tanner, C., Pinter, Y. (2024). Greed is All You Need: Evaluation of Tokenizer Inference Methods. ACL 2024, 813822.

Interpretability – Probing for Morphosyntactic Structure in Language Models

Key takeaway for L95: Syntax seems mostly solved in English LLMs, but probing shows persistent agreement errors and misalignment in multilingual / low-resource setups.

- Saphra, N., Wiegreffe, S. (2024). Mechanistic?. BlackboxNLP 2024, 480498.
- **Hewitt, J., Manning, C. D. (2019). A structural probe for finding syntax in word representations.** NAACL 2019, 41294138. link
- Pimentel, T., et al. (2020). Information-Theoretic Probing for Linguistic Structure. ACL 2020. link
- White, J. C., et al. (2021). A Non-Linear Structural Probe. NAACL 2021, 132138.
- Maudslay, R. H., Cotterell, R. (2021). Do Syntactic Probes Probe Syntax? NAACL 2021, 124131.

Agreement Probing

- **Finlayson, M., et al. (2021). Causal analysis of syntactic agreement mechanisms in neural language models.** ACL 2021, 18281843. link
- Marks, S., et al. (2025). Sparse Feature Circuits: Discovering and Editing Interpretable Causal Graphs in Language Models. ICLR 2025. link

Morphosyntactic Probing – Multilingual Language Models

- Chi, E. A., Hewitt, J., Manning, C. D. (2020). Finding universal grammatical relations in multilingual BERT. ACL 2020. link
- Arnett, C. (2025). On the Acquisition of Shared Grammatical Representations in Bilingual Language Models. ACL 2025. link
- Michaelov, J. A. (2023). Structural Priming Demonstrates Abstract Grammatical Representations. EMNLP 2023. link

Cross-Lingual Evaluation of Morphosyntax

- **Warstadt, A., et al. (2020). BLiMP: Benchmark of Linguistic Minimal Pairs for English.** TACL 2020, 377392.

- Xiang, B., et al. (2021). CLiMP: Chinese Language Model Evaluation. EACL 2021, 27842790.
- Song, Y., et al. (2022). SLING: Sino Linguistic Evaluation of Large Language Models. EMNLP 2022, 46064634.
- Taktasheva, E., et al. (2024). RuBLiMP: Russian Benchmark of Linguistic Minimal Pairs. EMNLP 2024, 92689299.

Tokenization and Low-Resource Languages

- Beinborn, L., Pinter, Y. (2023). Analyzing Cognitive Plausibility of Subword Tokenization. EMNLP 2023, 44784486.
- Arnett, C., Bergen, B. (2025). Why do language models perform worse for morphologically complex languages? COLING 2025, 66076623.
- Limisiewicz, T., et al. (2024). MYTE: Morphology-Driven Byte Encoding. ACL 2024, 1505915076.

Recent Parsing / Syntactic Language Models

- Dynamic Head Selection for Neural Lexicalized Constituency Parsing (ACL 2025). link
- Gomez-Rodriguez, C. (EMNLP 2024 / NoDaLiDa 2025). Revisit Supertagging / Dependency Parsing.
- Zhao, Y., et al. (ACL 2025). Compositional Syntactic Transformer Language Models. link

NLP for Very Low Resourced Languages / Treebanking

- Improving Low-Resource Morphological Inflection via Self-Supervised Objectives (ACL 2025). link
- Uzbek Treebank link, Luxembourgish link, Torwali link.

Other Applications of Morphosyntax

- Flickinger, D. (ACL 2025). Comparing LLM-generated and human-authored news text. link

Session Plan

Mon 12 Oct (Fermín)

Introduction: Natural Language Processing relies on Linguistics. arXiv:2405.05966.

Wed 15 Oct (Fermín)

Morphosyntax.

Mon 20 Oct (Paula)

Constituency, Heads, and Phrase Structure Grammar: Arguments, adjuncts, and long-distance dependencies.

Wed 22 Oct

Probing Syntax in LLMs.

- Hewitt & Manning (2019). *A Structural Probe for Finding Syntax in Word Representations.* NAACL 2019.
- Hall Maudslay & Cotterell (2021). *Do Syntactic Probes Probe Syntax? Experiments with Jabberwocky Probing.* NAACL 2021.
- Finlayson et al. (2021). *Causal Analysis of Syntactic Agreement Mechanisms in Neural Language Models.* ACL 2021.

Mon 27 Oct (Suchir/David)

Introducing Minimal Pair Datasets.

Wed 29 Oct (Suchir/David)

Minimal Pair Evaluation.

- Xiang et al. (2021). *CLiMP: A Benchmark for Chinese Language Model Evaluation.* EACL 2021.
- Song et al. (2022). *SLING: Sino Linguistic Evaluation of Large Language Models.* EMNLP 2022.

Mon 3 Nov (Paula)

Grammars I: PCFGs and Dependency Grammar.

Wed 5 Nov

Parsing and Benchmarking.

- *Dynamic Head Selection for Neural Lexicalized Constituency Parsing.* ACL 2025.
- *Better Benchmarking LLMs for Zero-Shot Dependency Parsing.* NoDaLiDa 2025.

Mon 10 Nov (Paula)

Grammars II: CCG and HPSG.

Wed 12 Nov

Cognitive Dependency and Supertagging.

- Gómez-Rodríguez (2024). *Revisiting Supertagging for Faster HPSG Parsing.* EMNLP 2024.

Mon 17 Nov (Fermín)

Morphology.

- Morphological Compositional Generalization in BabyLM 2025 Evaluation Pipeline.
- Ismayilzada et al. (2025). *Evaluating Morphological Compositional Generalization in Large Language Models.* NAACL 2025.

Mon 24 Nov (Suchir)

Tokenization.

Wed 26 Nov

Tokenization and Morphological Complexity.

- Beinborn & Pinter (2023). *Analyzing Cognitive Plausibility of Subword Tokenization.* EMNLP 2023.
- Arnett & Bergen (2025). *Why Do Language Models Perform Worse for Morphologically Complex Languages?* COLING 2025.

Mon 1 Dec

Introducing the Holiday Assignment.

Wed 3 Dec

Q & A Drop-in Session on the Holiday Task.

Course Summary

By the end of the course, you will understand how linguistic theory interfaces with modern computational models and will have experience critically evaluating LLMs through linguistically motivated experiments and datasets.