NLP and Linguistics

Juri Opitz
University of Zurich
www.juriopitz.com

It’s 2024

It's 2024

LLMs produce fluent text **without any specialised modules**

It’s 2024

Foreshadows:
1990s: “Statistical revolution”[1],
2011: “Neural revolution” [2]

LLMs produce fluent text **without any specialised modules**

It’s 2024

LLMs produce fluent text **without any specialised modules**

What is the role of linguistics in NLP?

Co-Thinkers

Shira Wein

Nathan Schneider

“CL”, “NLP”, and “Linguistics”

“CL”, “NLP”, and “Linguistics”

Linguistics

Study systematicity and variation in communication between humans

“CL”, “NLP”, and “Linguistics”

Linguistics

Study systematicity and variation in communication between humans

NLP

Developing technology for sophisticated computational processing of text

“CL”, “NLP”, and “Linguistics”

Linguistics

Study systematicity and variation in communication between humans

NLP

Developing technology for sophisticated computational processing of text

CL

As a broad field: includes NLP

A narrow focus (“cL”): answer RQs about language (rather than technology)

It’s 2024

LLMs produce fluent text **without any specialised modules**

What is the role of linguistics?

It’s 2024

LLMs produce fluent text **without any specialised modules**

What is the role of linguistics?

What NLP areas rely on it?

Thought experiment: Without Linguistics

Thought experiment: Without Linguistics

Resources: No carefully curated data sets

Thought experiment: Without Linguistics

Resources: No carefully curated data sets

Evaluation: Lack of expertise for designing human evaluation and cannot characterize many phenomena that challenge systems

Thought experiment: Without Linguistics

Resources: No carefully curated data sets.

Evaluation: Lack of expertise for designing human evaluation and cannot characterize many phenomena that challenge systems.

Low-Resource Settings: Struggle to understand why approaches that work for some language don’t work for another.

Thought experiment: Without Linguistics

Resources: No carefully curated data sets

Evaluation: Lack of expertise for designing human evaluation and cannot characterize many phenomena that challenge systems

Low-Resource Settings: Struggle to understand why approaches that work for some language don’t work for another.

Interpretability and Explanation: Can’t develop and test many hypothesis, harder to talk about observed system behavior.

Thought experiment: Without Linguistics

Resources: No carefully curated data sets

Evaluation: Lack of expertise for designing human evaluation and cannot characterize many phenomena that challenge systems

Low-Resource Settings: Struggle to understand why approaches that work for some language don’t work for another.

Interpretability and Explanation: Can’t develop and test many hypothesis, harder to talk about observed system behavior.

Study of Language: Purely commercial technology be indifferent to applications connected with scholarly or community-driven linguistic work.

Resources

Evaluation

Low-resource

RELIES

Study of Language

Explanation

Interpretability

NLP RELIES on Linguistics

Resources

The field of NLP is committed to an empirical methodology

Machine learning models are trained and evaluated on language data

The field of NLP is committed to an empirical methodology

Machine learning models are trained and evaluated on language data

Resources are supported by various degrees of linguistic knowledge

From proficiency in a language to formal training in linguistics

Examples

Examples

Linguistics increases data set diversity and quality

Examples

Linguistics increases data set diversity and quality

[3]: Linguistic diversity by controlling prepositions and complementizers

Examples

Linguistics increases data set diversity and quality

[3]: Linguistic diversity by controlling prepositions and complementizers

[4]: MT references need be paraphrased by linguists to ensure quality

Examples

Linguistics increases data set diversity and quality

[3]: Linguistic diversity by controlling prepositions and complementizers

[4]: MT references need be paraphrased by linguists to ensure quality

L-awareness in resource creation and development:

Examples

Linguistics increases data set diversity and quality

[3]: Linguistic diversity by controlling prepositions and complementizers

[4]: MT references need be paraphrased by linguists to ensure quality

L-awareness in resource creation and development:

Document characteristics and intended uses of datasets

Examples

Linguistics increases data set diversity and quality

[3]: Linguistic diversity by controlling prepositions and complementizers

[4]: MT references need be paraphrased by linguists to ensure quality

L-awareness in resource creation and development:

Document characteristics and intended uses of datasets

Resources with linguistic annotations: ED, UD, AMR, DRS, etc.

Evaluation

Evaluation

Evaluation

Contribute to big AI benchmarks

Evaluation

Contribute to big AI benchmarks

- A large portion of BigBench datasets are linguistic tasks [5]

Evaluation

Contribute to big AI benchmarks

A large portion of BigBench datasets are linguistic tasks [5]

System diagnostics and Fine-Grained Evaluation

Evaluation

Contribute to big AI benchmarks

A large portion of BigBench datasets are linguistic tasks [5]

System diagnostics and Fine-Grained Evaluation

[6, 8]: Challenge sets for linguistic and sociolinguistic performance

Evaluation

Contribute to big AI benchmarks

A large portion of BigBench datasets are linguistic tasks [5]

System diagnostics and Fine-Grained Evaluation

[6, 8]: Challenge sets for linguistic and sociolinguistic performance

[7]: Finer Evaluation through semantic structures

Evaluation

Contribute to big AI benchmarks

- A large portion of BigBench datasets are linguistic tasks [5]

System diagnostics and Fine-Grained Evaluation

- [6, 8]: Challenge sets for linguistic and sociolinguistic performance
- [7]: Finer Evaluation through semantic structures

Meta-Evaluation and Metric design

Evaluation

Contribute to big AI benchmarks

A large portion of BigBench datasets are linguistic tasks [5]

System diagnostics and Fine-Grained Evaluation

[6, 8]: Challenge sets for linguistic and sociolinguistic performance

[7]: Finer Evaluation through semantic structures

Meta-Evaluation and Metric design

E.g., What Linguistic features does “BERTscore” actually measure? [18]

Evaluation

Contribute to big AI benchmarks

A large portion of BigBench datasets are linguistic tasks [5]

System diagnostics and Fine-Grained Evaluation

[6, 8]: Challenge sets for linguistic and sociolinguistic performance

[7]: Finer Evaluation through semantic structures

Meta-Evaluation and Metric design

E.g., What Linguistic features does “BERTscore” actually measure? [18]

New measures: E.g., Incorporate sociolinguistic lexica to measure social bias [19]

Low-Resource

“Low-resource” - Goal based definition

“Low-resource” - Goal based definition

NLP goal:

Process low-resource languages

“Low-resource” - Goal based definition

NLP goal:

Process low-resource languages

Science goal:

Learning efficiency: Scientific insights for/on efficient learning

“Low-resource” - Goal based definition

NLP goal:

Process low-resource languages

Science goal:

Learning efficiency: Scientific insights for/on efficient learning

Practitioner’s goal:

Low-budget, Low computational resources

Low-budget

Low-budget

In our toolbox: Taggers, Parsers, etc.

Low-budget

In our toolbox: Taggers, Parsers, etc.

Explore, filter, structure large data sets

Low-budget

In our toolbox: Taggers, Parsers, etc.

Explore, filter, structure large data sets

Can be much cheaper than using LLMs; low computational cost

Low resource languages

Low resource languages

Linguistics in LLMs for processing low-resource languages:

Low resource languages

Linguistics in LLMs for processing low-resource languages:

[10] Use language-balanced tokenization with morphemes;

Low resource languages

Linguistics in LLMs for processing low-resource languages:

[10] Use language-balanced tokenization with morphemes;

[11] Add linguistic features and call for “Hiring a linguist”;

Low resource languages

Linguistics in LLMs for processing low-resource languages:

[10] Use language-balanced tokenization with morphemes;

[11] Add linguistic features and call for “Hiring a linguist”;

Linguistically sensitive supervision

Low resource languages

Linguistics in LLMs for processing low-resource languages:

[10] Use language-balanced tokenization with morphemes;

[11] Add linguistic features and call for “Hiring a linguist”;

Linguistically sensitive supervision

[12]: LLMs can be harmful to local language communities, if applied in a top-down approach, linguists can help understand communication situations

Interpretability and Explanation

Linguistics takes center stage

Linguistics takes center stage

Meta-language: NLP is pervaded by Linguistic meta-language

Linguistics takes center stage

Meta-language: NLP is pervaded by Linguistic meta-language

Interpretability method goal: binding observations to this language

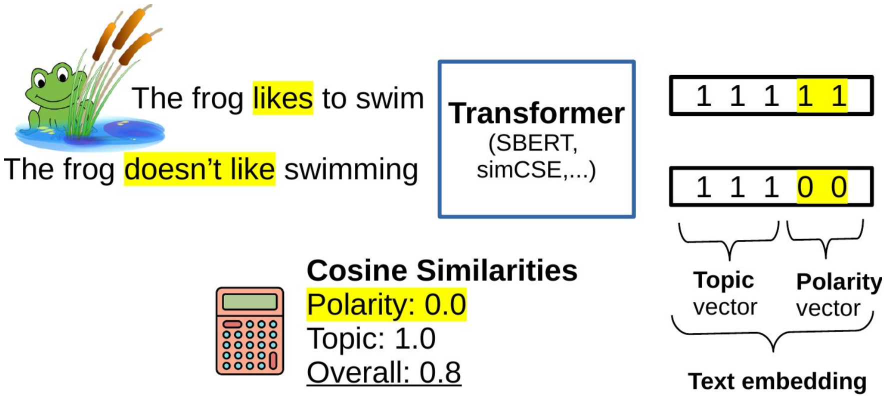
Example of presenter’s work [13]

Semantic similarity is a crucial NLP task

But how is similarity assessed?

Neural models are large black boxes

Idea: Bind embedding parts to concepts



Ideas from linguistics and adjacent fields in debates

How to **interpret what NLP models can** represent?

How to define machine ‘understanding’? [14, 15]

Is grounding required for a model to capture meaning? [e.g., 16, 17]

Study of Language

Study of Language

Linguistics as the application domain

Study of Language

Linguistics as the application domain

Make content available vs. **Study language systems**

Study language systems

Study language systems

Classic cL tools and tasks

Study language systems

Classic cL tools and tasks

- Corpora, Pattern search

Study language systems

Classic cL tools and tasks

- Corpora, Pattern search

Documentary and historical linguistics

Study language systems

Classic cL tools and tasks

- Corpora, Pattern search

Documentary and historical linguistics

- Noisy data, image or audio form (without transcriptions), orthography?; basic grammar?

Study language systems

Classic cL tools and tasks

Corpora, Pattern search

Documentary and historical linguistics

Noisy data, image or audio form (without transcriptions), orthography?; basic grammar?

Need not: Run-off-the-mill LLMs; **Need**: noise tolerant interdisciplinary research

Study language systems

Classic cL tools and tasks

- Corpora, Pattern search

Documentary and historical linguistics

- Noisy data, image or audio form (without transcriptions), orthography?; basic grammar?
- **Need not**: Run-off-the-mill LLMs; **Need**: noise tolerant interdisciplinary research

Language-focused study in fields beyond linguistics proper

Study language systems

Classic cL tools and tasks

Corpora, Pattern search

Documentary and historical linguistics

Noisy data, image or audio form (without transcriptions), orthography?; basic grammar?

Need not: Run-off-the-mill LLMs; **Need**: noise tolerant interdisciplinary research

Language-focused study in fields beyond linguistics proper

Law, Literature, Humanities, etc.

Summary

Summary

The acronym RELIES is non-exhaustive

Summary

The acronym RELIES is non-exhaustive

RELIES indicates ways in which NLP did, does, will continue to, and is going to rely on linguistics

Summary

The acronym RELIES is non-exhaustive

RELIES indicates ways in which NLP did, does, will continue to, and is going to rely on linguistics

Read more: https://arxiv.org/abs/2405.05966

Summary

Your thoughts?

The acronym RELIES is non-exhaustive

RELIES indicates ways in which NLP did, does, will continue to, and is going to rely on linguistics

Read more: https://arxiv.org/abs/2405.05966

Thank you for listening

References (for an exhaustive list, please refer to https://arxiv.org/abs/2405.05966)

[1] Cf. Brown, Peter F. (1993). "The mathematics of statistical machine translation: Parameter estimation". Computational Linguistics (19): 263–311.

[2] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12:2493–2537.

[3] Abhilasha Ravichander, Matt Gardner, and Ana Marasovic. 2022. CONDAQA: A contrastive reading comprehension dataset for reasoning about negation. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing.

[4] Markus Freitag, David Grangier, and Isaac Caswell. 2020. BLEU might be guilty but references are not innocent. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing.

[5] Aarohi Srivastava, et al. 2023. Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. Transactions on Machine Learning Research.

[6] Rachel Rudinger, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme. 2018. Gender bias in coreference resolution. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

[7] Juri Opitz and Anette Frank. 2021. Towards a decomposable metric for explainable evaluation of text generation from AMR. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics

[8] Marco Tulio Ribeiro, Tongshuang Wu, Carlos Guestrin, and Sameer Singh. 2020. Beyond accuracy: Behavioral testing of NLP models with CheckList. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics

[10] Mathias Creutz and Krista Lagus. 2002. Unsupervised discovery of morphemes. In Proceedings of the ACL-02 Workshop on Morphological and Phonological Learning

[11] Kexun Zhang, Yee Choi, Zhenqiao Song, Taiqi He, William Yang Wang, and Lei Li. 2024. Hire a linguist! In Findings of ACL 2024.

[12] Steven Bird. 2022. Local languages, third spaces, and other high-resource scenarios. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)]

[13] Juri Opitz and Anette Frank. 2022. SBERT studies meaning representations: Decomposing sentence embeddings into explainable semantic features. In AACL 2022.

[14] Jesse Dunietz, Greg Burnham, Akash Bharadwaj, Owen Rambow, Jennifer Chu-Carroll, and Dave Ferrucci. 2020. To test machine comprehension, start by defining comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics

[15] Sagnik Ray Choudhury, Anna Rogers, and Isabelle Augenstein. 2022. Machine reading, fast and slow: When do models “understand” language? In Proceedings of the 29th International Conference on Computational Linguistics

[16] Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: on meaning, form, and understanding in the age of data. In Proc. of ACL.

[17] Ellie Pavlick. 2023. Symbols and grounding in large language models. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences

[18] Hanna, Michael, and Ondřej Bojar. "A fine-grained analysis of BERTScore." Proceedings of the Sixth Conference on Machine Translation. 2021.

[19] Julius Steen and Katja Markert. 2024. Bias in news summarization: Measures, pitfalls and corpora. In Findings of the Association for Computational Linguistics ACL 2024

Addendum

Study on LLM translation from a grammar book

LLMs can translate from one grammar book [12]

- But cannot reach the level of a linguist
- A linguist is also needed to establish the upper-bound