What Has LeBenchmark Learnt about French Syntax

Zdravko Dugonjić🤖, **Adrien Pupier**🏔️, Benjamin Lecouteux🏔️, Maximin Coavoux🏔️

🤖 LASR Lab, Technische Universität Dresden, School of Embedded and Composite AI (SECAI), Germany
🏔️Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

LREC 2024 – Torino

Context and Questions

- Large pretrained models are the de facto backbones of current natural language and speech processing systems
  - **textual models**: BERT (Devlin et al., 2019), and many others
  - **acoustic models**: wav2vec2 (Baevski et al., 2020), and many others
- BERT has been shown to acquire syntactic knowledge, although it was not trained explicitly with syntax supervision (Tenney et al., 2019; Lin et al., 2019; Rogers et al., 2020)
- Outstanding questions:
  - Do acoustic models also learn syntactic knowledge, despite being trained on very low level information (acoustic signal)?
  - How is syntactic information distributed across layers?

Contributions

- We present a study of a French acoustic model: LeBenchmark (Evain et al., 2021)
- Probing for 2 tasks: **POS tagging, unlabeled dependency parsing**
- Results:
  - Has LeBenchmark learnt syntax?

Contributions

- We present a study of a French acoustic model: LeBenchmark (Evain et al., 2021)
- Probing for 2 tasks: **POS tagging, unlabeled dependency parsing**
- Results:
  - Has LeBenchmark learnt syntax? Yes ... to some extent

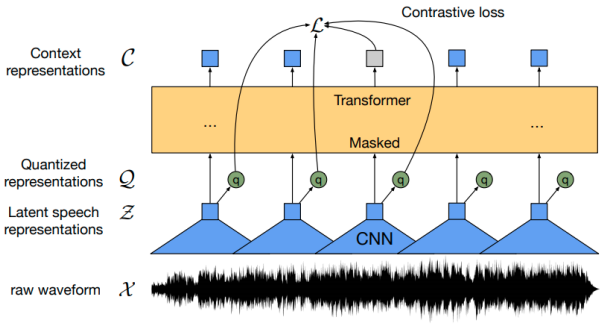
Contributions

- We present a study of a French acoustic model: LeBenchmark (Evain et al., 2021)
- Probing for 2 tasks: **POS tagging, unlabeled dependency parsing**
- Results:
  - Has LeBenchmark learnt syntax? Yes ... to some extent
  - How is syntactic knowledge distributed across the architecture layers?

Contributions

- We present a study of a French acoustic model: LeBenchmark (Evain et al., 2021)
- Probing for 2 tasks: **POS tagging, unlabeled dependency parsing**
- Results:
  - Has LeBenchmark learnt syntax? Yes ... to some extent
  - How is syntactic knowledge distributed across the architecture layers? It peaks in the middle layers and almost disappears in the final layers

wav2vec2 (Baevski et al., 2020)

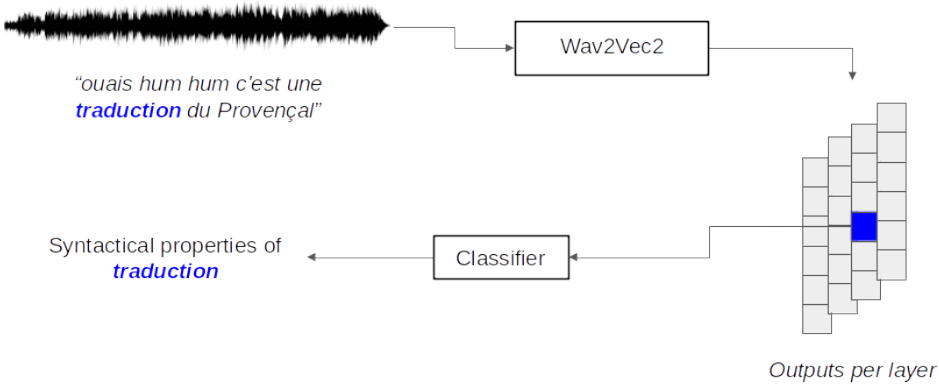


- Transformer-based (with 24 layers)
- Self supervised training using contrastive loss
- Input: raw audio signal
- Pretrained model produces vector representation of any sound file to use for downstream tasks (Automatic Speech recognition, Spoken Language Understanding, Speech Translation)

LeBenchmark (Evain et al., 2021)

- Wav2vec2 architecture, trained on **French** spoken corpora
- `Lebenchmark7K` train on 7k hours of spoken French (diverse types of speech situations)
- `https://huggingface.co/LeBenchmark/wav2vec2-FR-7K-large`

Probing tasks

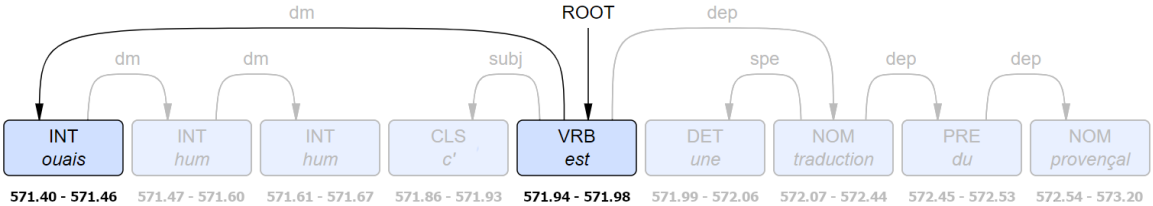


We focus on two properties:

- POS tagging: what is the part-of-speech of a token?
- Unlabelled dependency parsing: what is the syntactic head of the token?

Probing tasks

We address both tasks as token-tagging tasks:



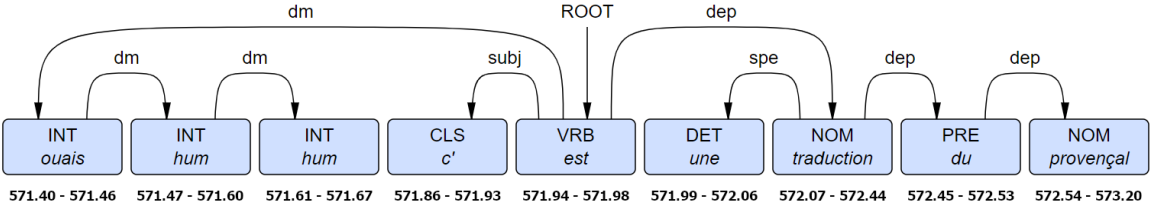
`Ouais` should be assigned

- INT (interjection) as POS tag
- +4 as head, i.e. its **relative position**

Dataset

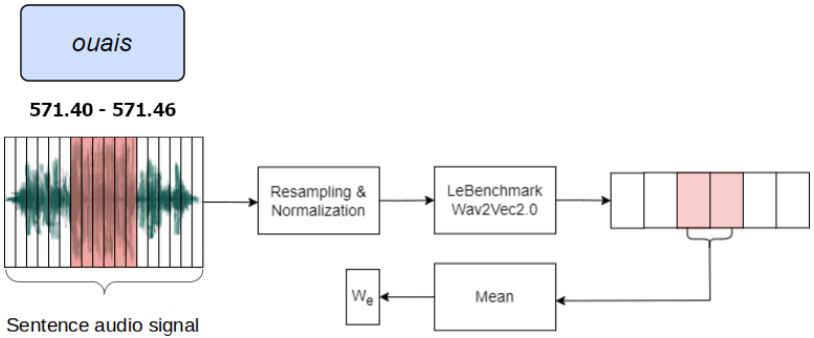
Orféo treebank (Benzitoun et al., 2016)

- Large treebank of spoken French (196h, 2.5M tokens)
  - 9h have gold syntactic annotations
  - the rest of the corpus have good quality silver syntactic annotations
- (Silver) token-level timestamp available!



Word Representations

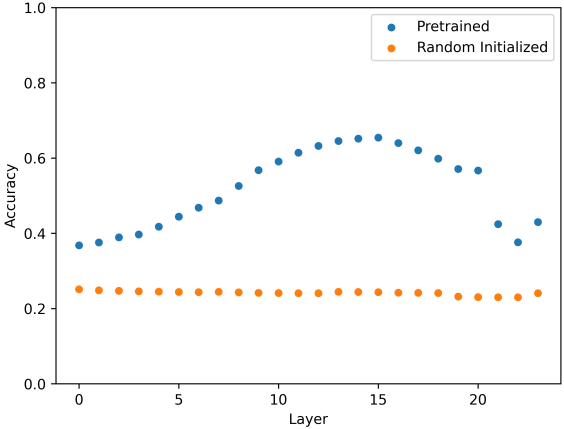
- Since probing tasks are word-tagging tasks, we need word representations
- Mean-pooling over all frames that are within the timestamps of a word

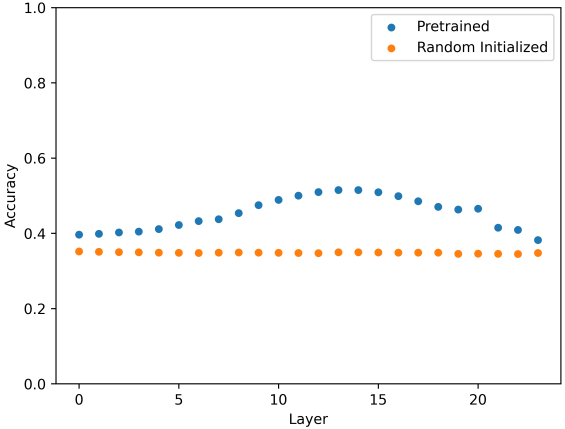


$$P(\text{pos}|w_e, \text{input signal}) = \text{Softmax}(\mathbf{W}^{\text{pos}} \cdot w_e)$$

$$P(\text{head}|w_e, \text{input signal}) = \text{Softmax}(\mathbf{W}^{\text{head}} \cdot w_e)$$

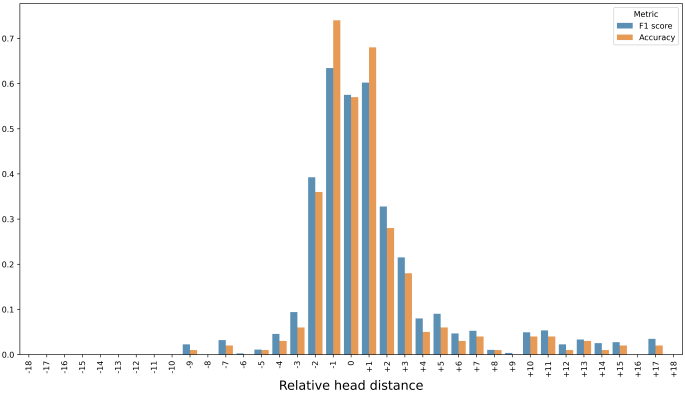
Experiments: Results





- Similar pattern with peak at layer 14
- Syntax almost disappears in last layers

Experiments: Results on layer 14



- Highest accuracy on close range dependency

Conclusion discussion

- Our results in line with Shen et al. (2023) who investigate probing
  - for English — whereas we experiment on French
  - read/prepared speech — whereas we use spontaneous speech
  - with another probing methodology

- Compared to BERT probing for syntax:
  - LeBenchmark also has a peak in the middle layers
  - but extremely sharp decrease in the final layers
  - whereas for BERT: the decrease is very mild

References I

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33:12449–12460, 2020.

Christophe Benzitoun, Jeanne-Marie Debaisieux, and Henri-José Deulofeu. Le projet orféo: un corpus d'étude pour le français contemporain. *Corpus*, (15), 2016.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.

Solène Evain, Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval, Didier Schwab, and Laurent Besacier. LeBenchmark: a reproducible framework for assessing self-supervised representation learning from speech. In *Proc. Interspeech 2021*, pages 1439–1443, 2021. doi: 10.21437/Interspeech.2021-556.

Yongjie Lin, Yi Chern Tan, and Robert Frank. Open sesame: Getting inside BERT's linguistic knowledge. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 241–253, Florence, Italy, August 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-4825. URL https://aclanthology.org/W19-4825.

Anna Rogers, Olga Kovaleva, and Anna Rumshisky. A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, 8:842–866, 2020. doi: 10.1162/tacl_a_00349. URL https://aclanthology.org/2020.tacl-1.54.

Gaofei Shen, Afra Alishahi, Arianna Bisazza, and Grzegorz Chrupała. Wave to syntax: Probing spoken language models for syntax. *arXiv preprint arXiv:2305.18957*, 2023.

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Sam Bowman, Dipanjan Das, and Ellie Pavlick. What do you learn from context? probing for sentence structure in contextualized word representations. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=SJzSgnRcKX.