

On the Interplay between Positional Encodings, Morphological Complexity, and Word Order Flexibility

Kushal Tatariya*, Wessel Poelman*, Miryam de Lhoneux

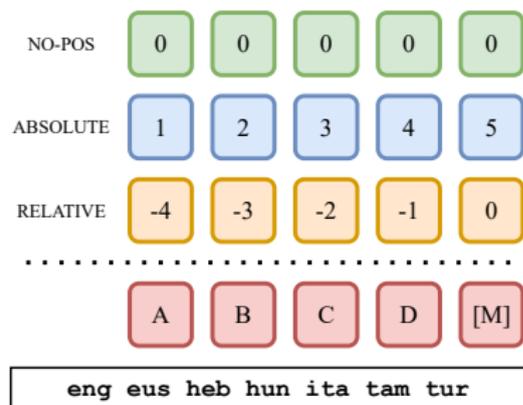
KU Leuven



The Trade-off Hypothesis

Languages with high word order flexibility tend to be more morphologically complex. What is the interplay between position encodings, word order flexibility and morphological complexity?

Intuition: The higher the word order flexibility, the less useful positional encodings might be, and vice-versa. Conversely, a more morphologically complex language could depend more on positional encodings (if longer words are segmented into more tokens), and vice-versa.



The Proxies

| Language | Dominant Order | Word Order | | | Morphological Complexity | | |
|-----------|----------------|------------|------|--------|--------------------------|--------|-------|
| | | HDE | ROE | SO-ROE | AV | η | MATTR |
| Basque | SOV | 0.50 | 0.21 | 0.72 | 30.22 | 0.38 | 0.68 |
| English | SVO | 0.16 | 0.03 | 0.20 | 25.20 | 0.34 | 0.53 |
| Hebrew | SVO | 0.38 | 0.13 | 0.25 | 27.48 | 0.32 | 0.69 |
| Hungarian | NDO | 0.40 | 0.12 | 0.83 | 42.57 | 0.37 | 0.63 |
| Italian | SVO | 0.26 | 0.06 | 0.27 | 24.63 | 0.33 | 0.58 |
| Tamil | SOV | 0.07 | 0.09 | 0.94 | 41.50 | 0.45 | 0.69 |
| Turkish | SOV | 0.22 | 0.13 | 0.31 | 33.32 | 0.36 | 0.65 |

Word Order Flexibility

UD-based metrics by Futrell et al., (2015)¹:

1. **Head Direction entropy (HDE)**: conditional entropy of whether a head is to the right or left of a dependent.
2. **Relation Order Entropy (ROE)**: conditional entropy of the order of words in a local sub-tree.
3. **Subject Object ROE (SO-ROE)**: conditional entropy of the order of the subject and object in the main clause.

¹Richard Futrell, Kyle Mahowald, and Edward Gibson. “Quantifying Word Order Freedom in Dependency Corpora”. In: *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*. Ed. by Joakim Nivre and Eva Hajičová. Uppsala, Sweden: Uppsala University, Uppsala, Sweden, Aug. 2015, pp. 91–100. URL: <https://aclanthology.org/W15-2112/>.

Morphological Complexity

Measured on a sample of 250k lines of the training data, using the same tokenizer as the model:

1. **Accessor Variety (AV)**: how often types from the vocabulary co-occur, either on the left or right in a corpus.
2. **Accessor Efficiency (η)**: the Shannon efficiency of AV.
3. **Moving Average TTR (MATTR)**: the type-token ratio of the corpus as calculated in a sliding window of tokens.

Main Contributions

1. We systematically select languages with diverse typological profiles in terms of morphological complexity and word order strategies.
2. We train monolingual models from scratch with three positional encoding methods (**ABSOLUTE**, **RELATIVE**, **NO-POS**), keeping all other variables constant.
3. We analyse the impact of position encodings in the context of the trade-off hypothesis, using fine-grained proxies of morphological complexity and word order flexibility.

Results: Pretraining

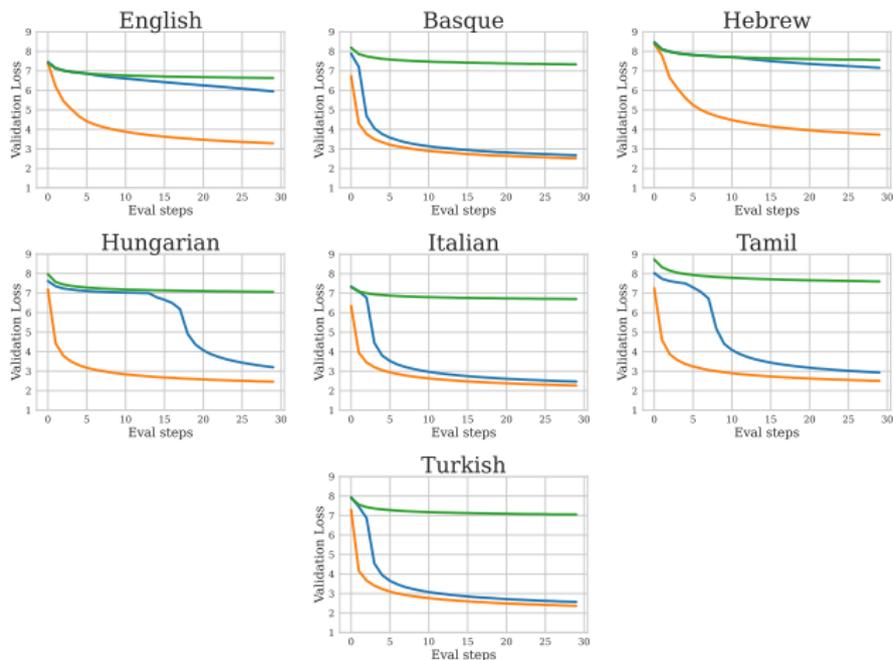


Figure 1: Loss curves on validation set for pretraining for **ABSOLUTE**, **RELATIVE**, and **NO-POS**.

Results: Finetuning

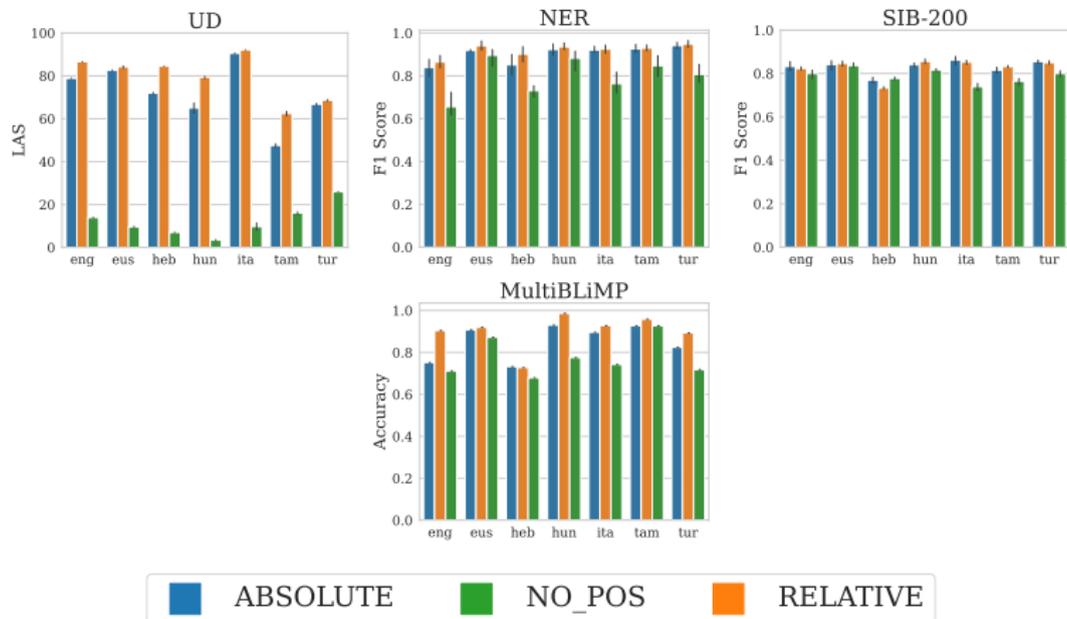


Figure 2: Results per task, language, and positional encoding type. Scores are averaged over 5 runs.

Position Encodings and Morphological Complexity

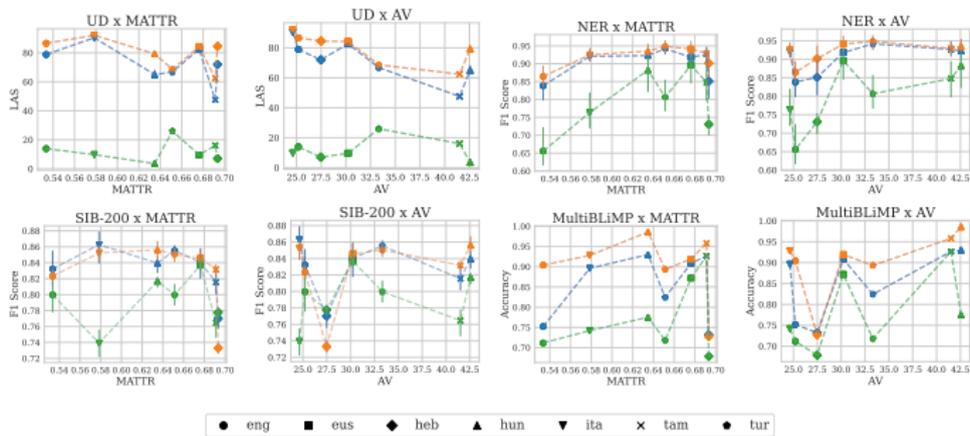


Figure 3: Relation between proxies for morphological complexity and downstream performance. The line shows the groupings of positional encoding type.

Position Encodings and Word Order Flexibility

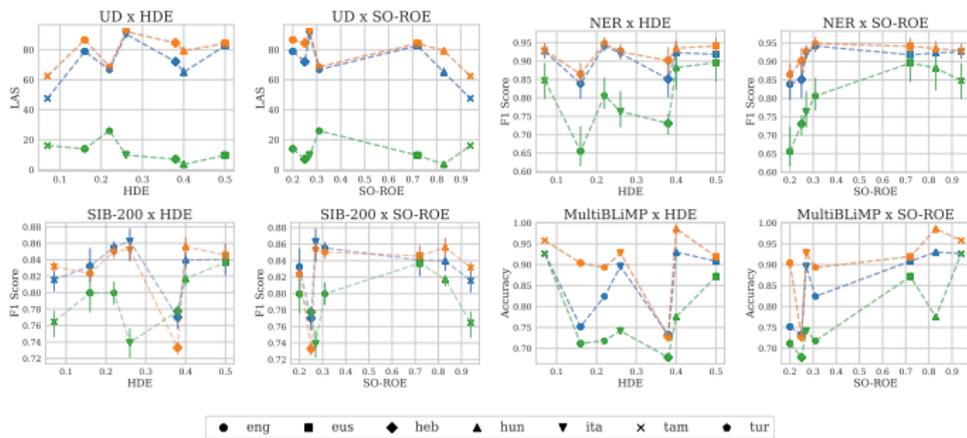


Figure 4: Relation between proxies for word order flexibility and downstream performance. The connecting line is just to show groupings of positional encoding type, not a regression line.

Findings

1. The impact of position encodings is task specific.
2. Relative position encoding gives the most consistent performance across languages and tasks.
3. Position encodings are important to learn syntax, regardless of word order flexibility or morphological complexity.
4. Morphological complexity and word order flexibility do not have a direct interaction with position encodings, contradicting previous research.