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

Kushal Tatariya*, Wessel Poelman*, Miryam de Lhoneux

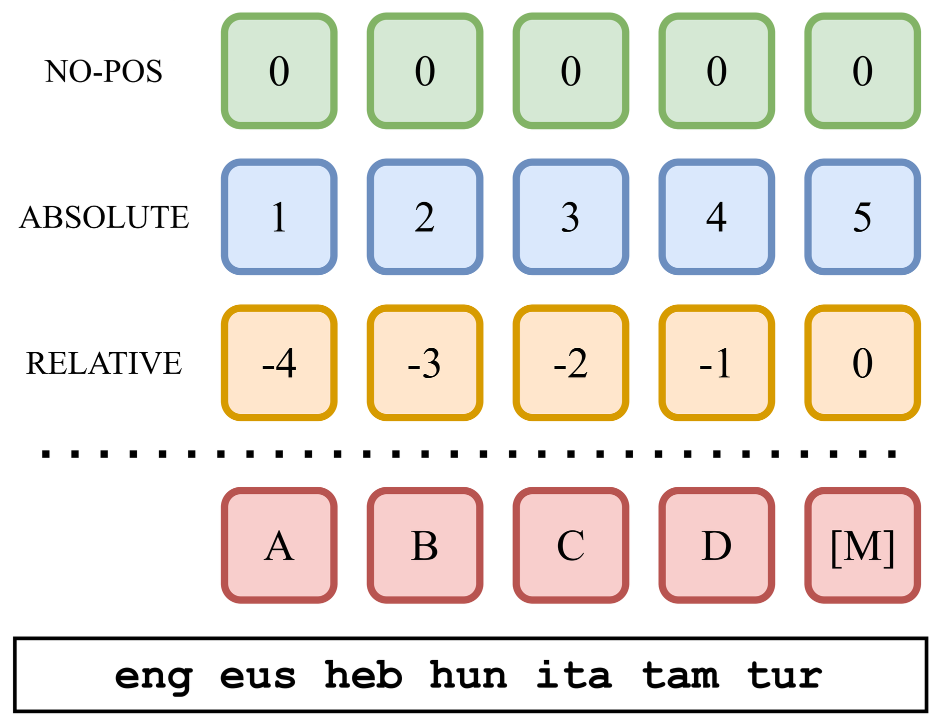
Department of Computer Science, KU Leuven

kushaljayesh.tatariya@kuleuven.be

1. 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.



1.1 The Proxies

Language	Dominant Order	Word Order HDE	ROE	SO-ROE	Morphological Complexity 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

1. **Word Order Flexibility:** UD-based metrics by Futrell et al., (2025):
   - (a) **Head Direction entropy (HDE):** conditional entropy of whether a head is to the right or left of a dependent.
   - (b) **Relation Order Entropy (ROE):** conditional entropy of the order of words in a local sub-tree.
   - (c) **Subject Object ROE (SO-ROE):** conditional entropy of the order of the subject and object in the main clause.
2. **Morphological Complexity:** Measured on of 250k lines of the training data, using the same tokenizer as the model:
   - (a) **Accessor Variety (AV):** how often types from the vocabulary co-occur, either on the left or right in a corpus.
   - (b) **Accessor Efficiency (η):** the Shannon efficiency of AV.
   - (c) **Moving Average TTR (MATTR):** the type-token ratio of the corpus as calculated in a sliding window of tokens.

2. 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.

3. Results

3.1 Pretraining

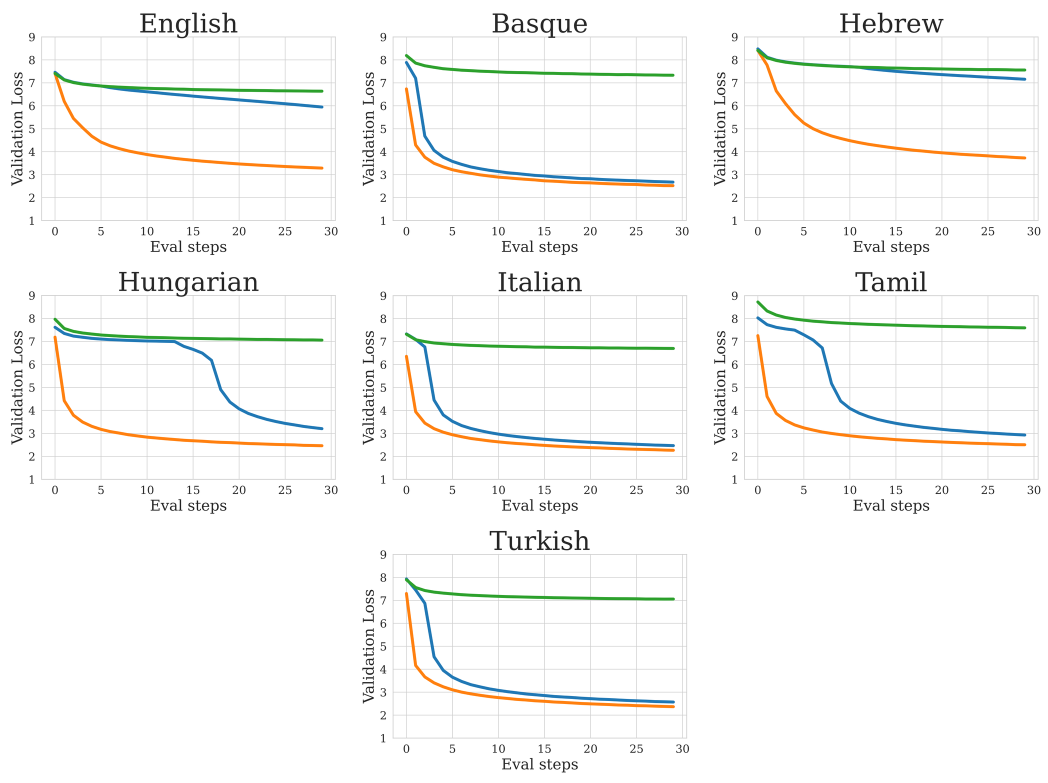


Figure 1: Loss curves on validation set for pretraining.

3.2 Finetuning

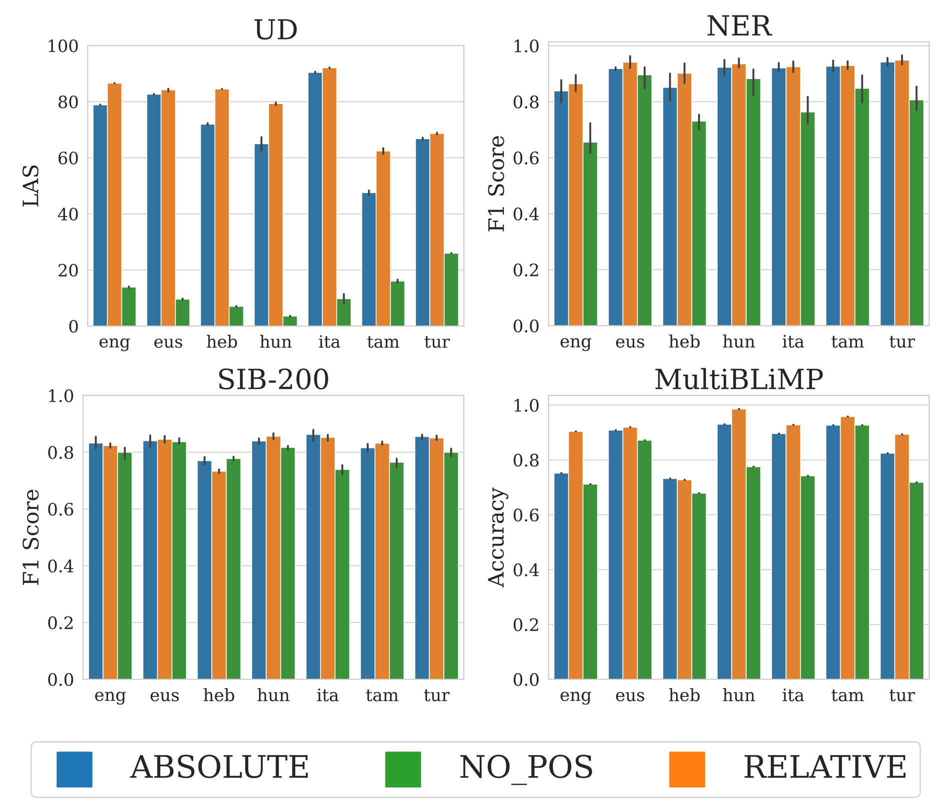
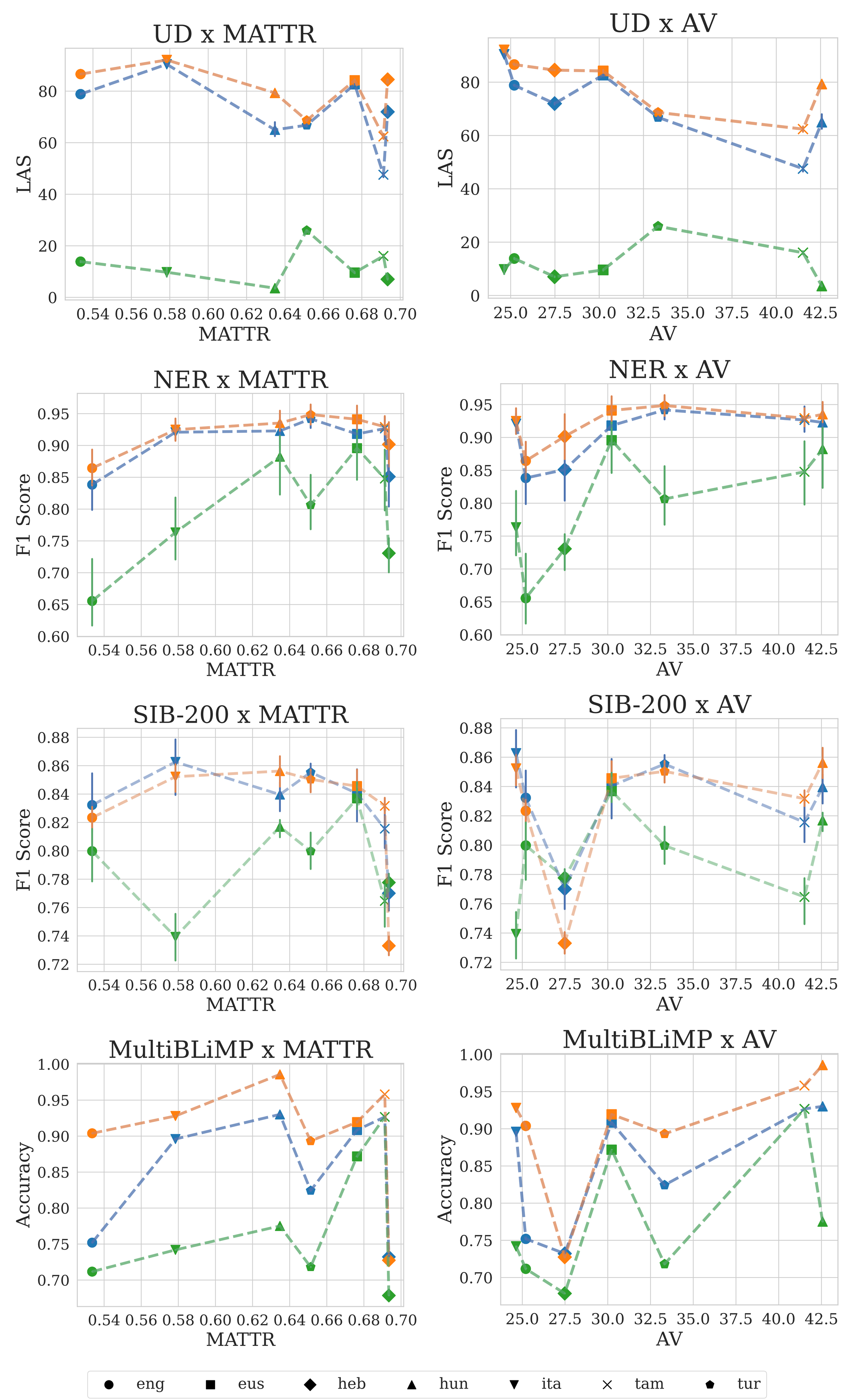


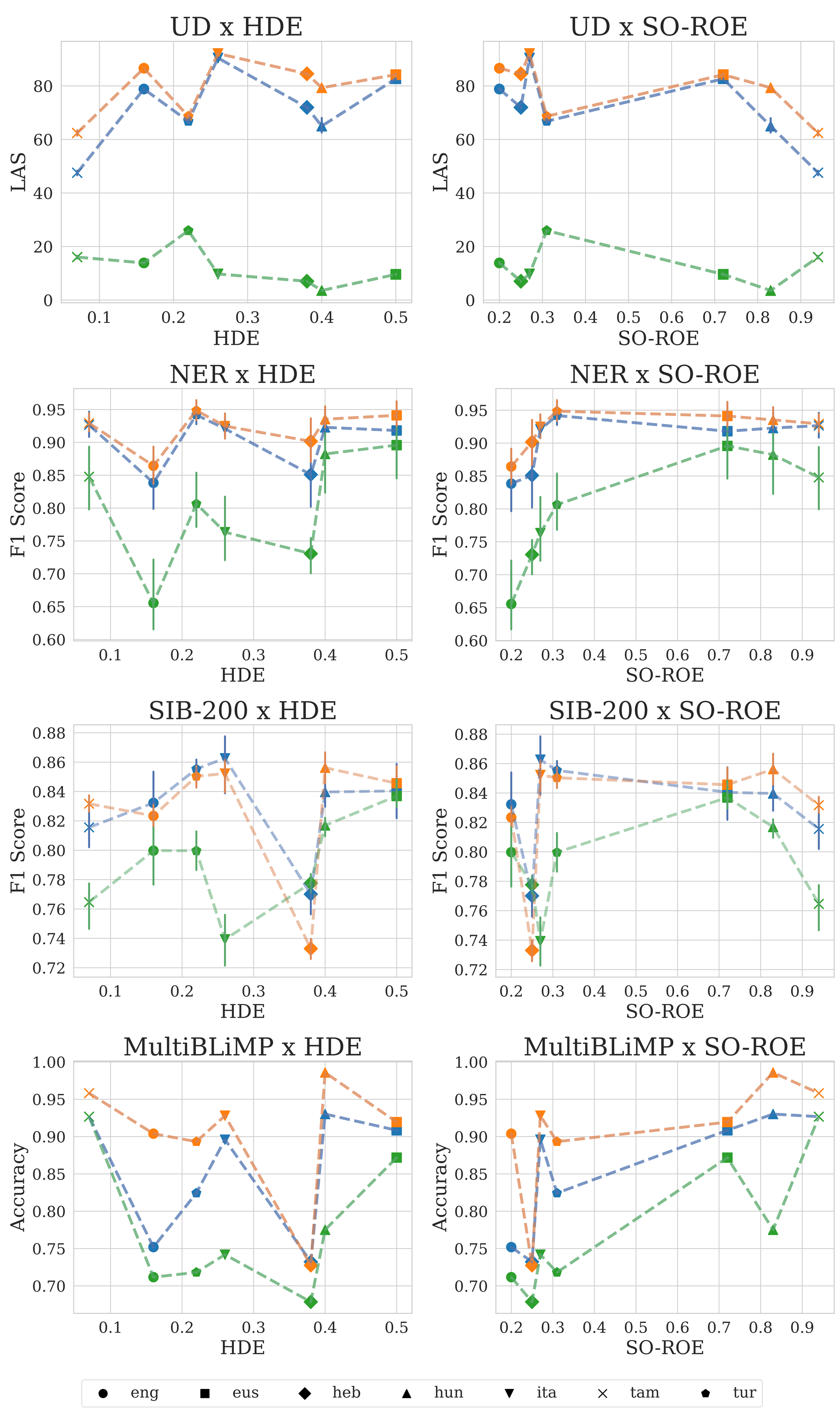
Figure 2: Results per task, language, and positional encoding type.

4. Analysis

4.1 Position Encodings and Morphological Complexity



4.2 Position Encodings and Word Order Flexibility



5. Findings

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

Paper

KU LEUVEN

LinkedIn