

# Going Beyond Zero-Shot MT: Combining Phonological, Morphological and Semantic Factors

Cristina Espana-Bonet & Josef van Genabith

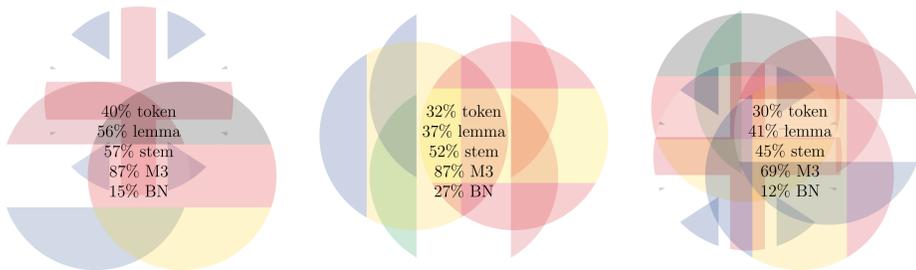
Our systems based on factored NMT are designed to fully exploit multilinguality by including factors that increase the number of common elements among languages such as phonetic coarse encodings and synsets, besides shallow part-of-speech tags, stems and lemmas. Document level information is also considered by including the topic of every document. This approach improves a baseline without any additional factor for all the language pairs and even allows beyond-zero-shot translation. That is, the translation from unseen languages is possible thanks to the common elements —especially synsets in our models— among languages.

## Factors

- Topic:** Biterm Topic Model learned for 20 topics, so that each topic is the main topic of at least 1% of the training documents
- M3 encoding:** Approximate phonetic encoding with Metaphone 3
- BN Synset:** For nouns (incl. named entities, foreign words and numerals), adjectives, adverbs and verbs. Negation particles are tagged with NEG
- PoS:** {NOUN, VERB, PREPOSITION, PRONOUN, DETERMINER, ADVERB, ADJECTIVE, CONJUNCTION, ARTICLE, INTERJECTION}  
Tool: IXA pipeline & TreeTagger + Mappings
- Lemma:** Tool: IXA pipeline & TreeTagger
- Stem:** Tool: Snowball API

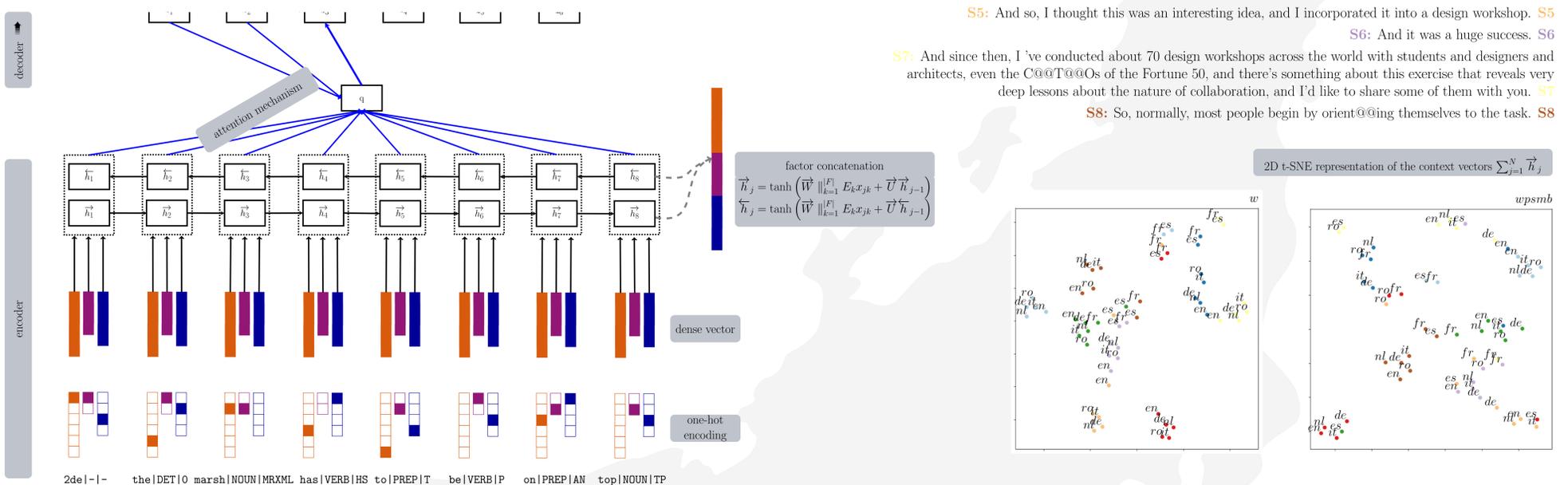
orient@@|VERB|orient|ARNTNK|bn:00091197v|t2 ing|VERB|orient|ARNTNK|bn:00091197v|t2  
(https://github.com/cristinae/BabelWE)

## Multilingual TED

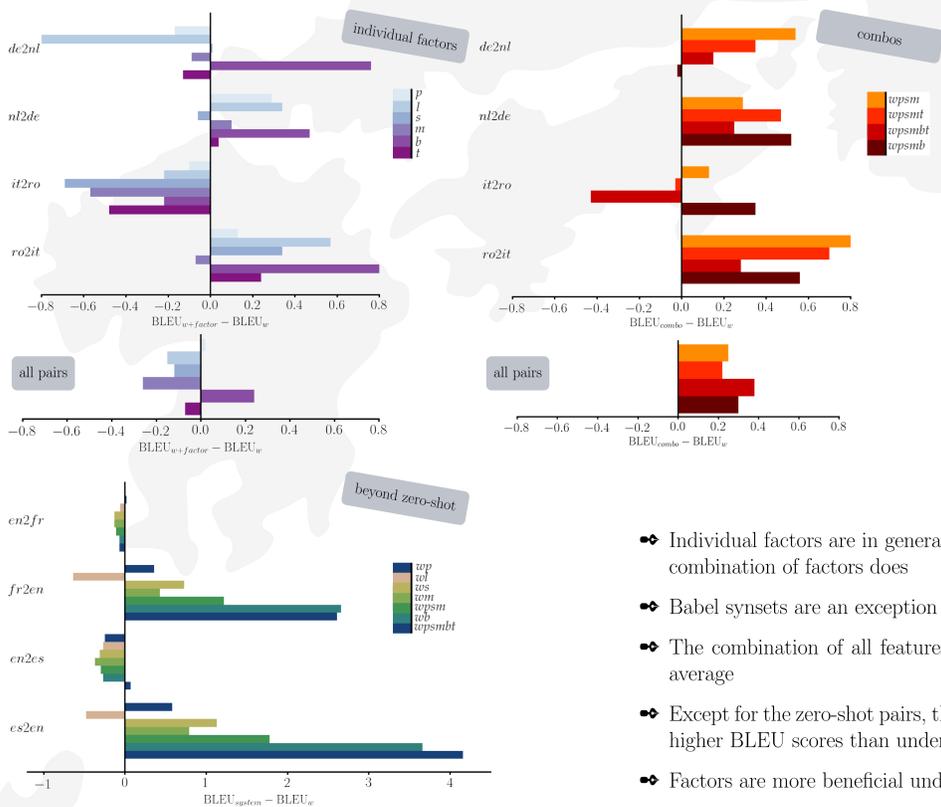


- High intersection of factors among all languages
- Tokens and lemmas are more similar in West Germanic languages, Babel synsets in Latin languages
- Location in the ML-space is not improved by factors
- Further improvements: ML-WSD for synsets, PoS selection for M3 encodings, etc

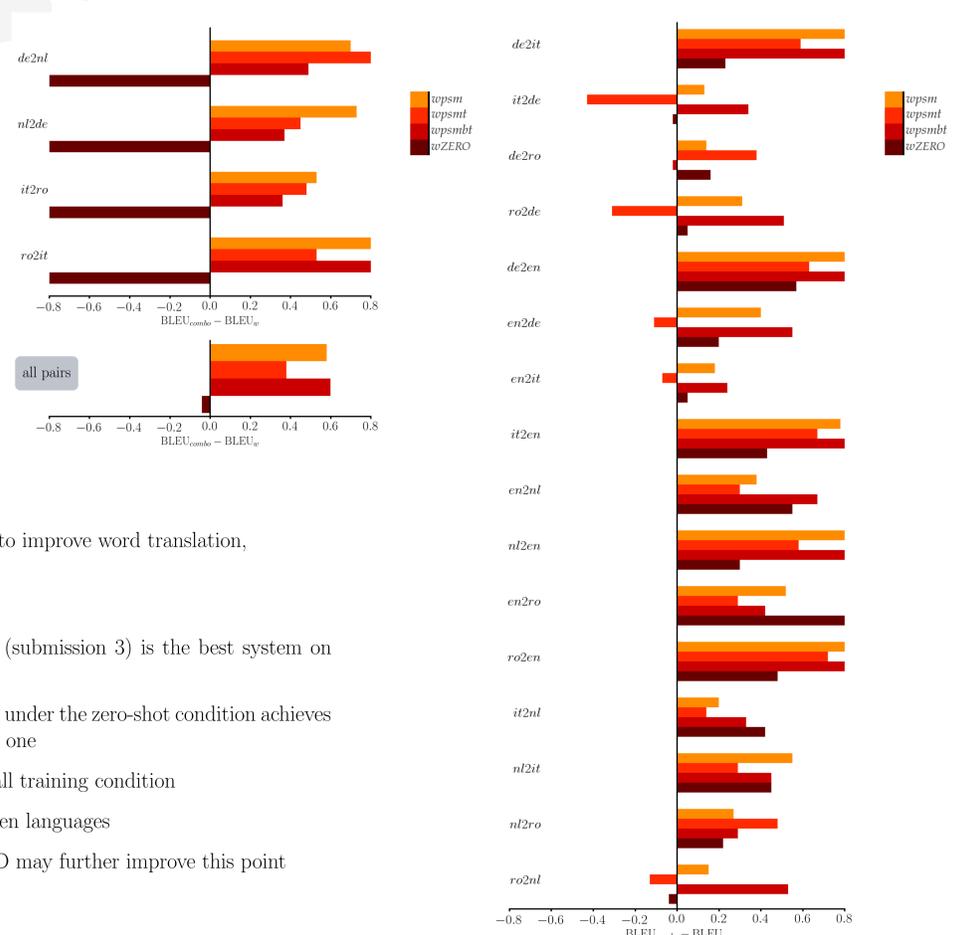
## Architecture



## Zero-Shot Training Condition



## Small Training Condition



- Individual factors are in general not able to improve word translation, combination of factors does
- Babel synsets are an exception
- The combination of all features *wpsmbt* (submission 3) is the best system on average
- Except for the zero-shot pairs, the training under the zero-shot condition achieves higher BLEU scores than under the small one
- Factors are more beneficial under the small training condition
- Babel synsets allow translation from unseen languages
- Joint learning for ML-NMT and ML-WSD may further improve this point