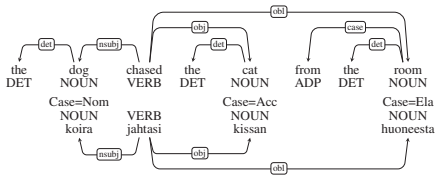
Nucleus Composition in Transition-Based Dependency Parsing

Joakim Nivre[1,2] Ali Basirat[3] Luise Dürlich[1,2] Adam Moss[1]
[1]Uppsala University [2]RISE Research Institutes of Sweden [3]Linköping University

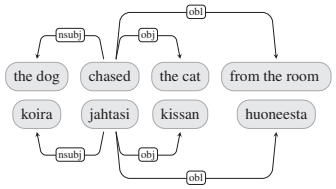
Computational Linguistics 48(4), 849–886, 2022.

Syntactic Nuclei

Are the elementary units of syntax **words** . . .

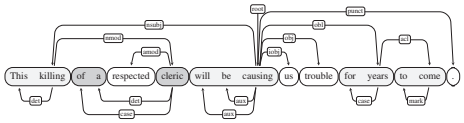


. . . or **syntactic nuclei** [Tesnière, 1959]?



Syntactic Nuclei in UD

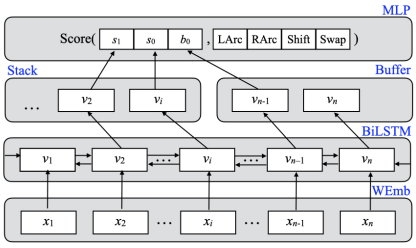
Nucleus = Subtree with only functional relations: **aux**, **case**, **cc**, **clf**, **cop**, **det**, **mark**.



Experiments use UD v2.8.1 [Zeman et al., 2021].

Nucleus Composition

Transition-based parser with BiLSTM encoder [Kiperwasser and Goldberg, 2016]



How do we represent a nucleus $f(h, d, l)$?

Baseline:

$$f(h, d, l) = \vec{h}$$

Nucleus composition (NC):

$$f(h, d, l) = \begin{cases} \vec{h} + g(\vec{h}, \vec{d}, \vec{l}) & \text{if } l \in F \\ \vec{h} & \text{otherwise} \end{cases}$$

where

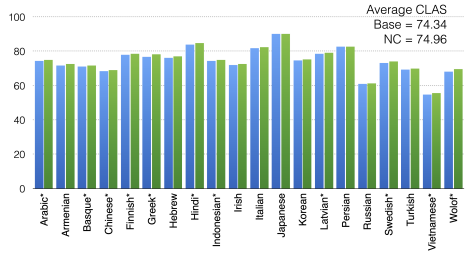
$$g(\vec{h}, \vec{d}, \vec{l}) = \sigma(W(\vec{h} \odot \vec{d} \odot \vec{l}) + b)$$

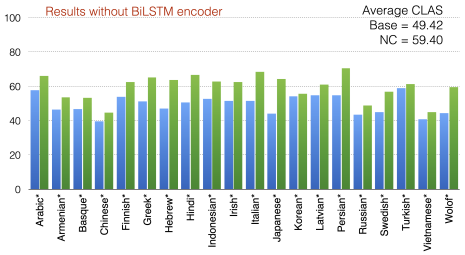
Research Questions

1. To what extent does nucleus composition improve parsing accuracy?
2. What factors determine the rate of improvement for different languages?
3. Which linguistic constructions benefit most from nucleus composition?
4. What information is captured by the learned composition function?

Does parsing accuracy improve?

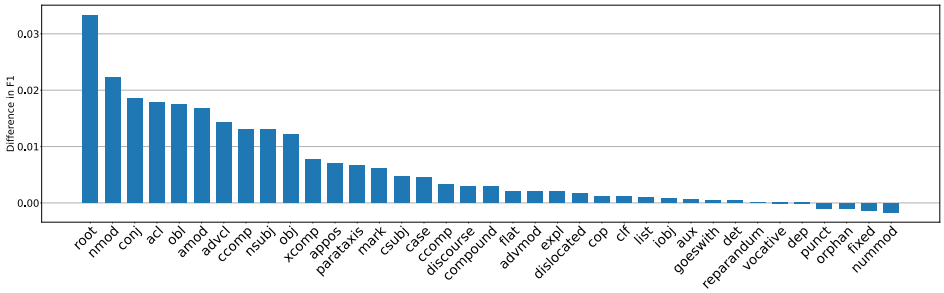
Small but consistent improvement across languages; much larger improvements when BiLSTM is ablated.





Which linguistic constructions benefit?

Improvements on main predicates, nominal dependents, clausal dependents, and coordination.



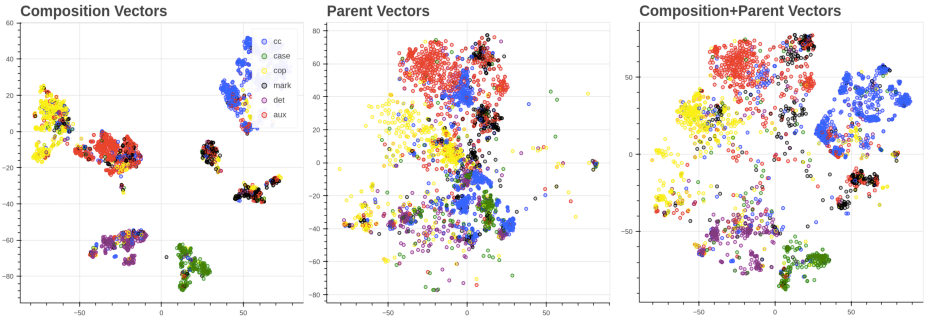
Can we explain the improvements?

Significant effects of a linear-mixed effects model for predicting improvement in CLAS score:

With BiLSTM encoder	Without BiLSTM encoder
Frequency of **det** relation	Frequency of **case** relation
Relational entropy of **cc** head	Frequency of **det** relation
Categorial entropy of **cc** head	Frequency of **cop** relation
	Frequency of **aux** relation

What is captured by composition?

Composition increases similarity of vectors representing nuclei of the same syntactic type (Finnish).



References

Eliyahu Kiperwasser and Yoav Goldberg. Simple and accurate dependency parsing using bidirectional LSTM feature representations. *Transactions of the Association for Computational Linguistics*, 4:313–327, 2016.

Lucien Tesnière. *Éléments de syntaxe structurale*. Editions Klincksieck, 1959.

Daniel Zeman et al. Universal dependencies 2.8.1. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, http://hdl.handle.net/11234/1-3687, 2021.