Neural Supertag-based Parsing for Tree Wrapping Grammars

Tatiana Bladier, Jakub Waszczuk, Laura Kallmeyer

TreeGraSP meeting #5
Düsseldorf, 11 November 2020

hhu
Heinrich Heine
Universität Düsseldorf

erc
European
Research
Council

Tree Wrapping Grammar (TWG)

- Finite set of elementary trees, combined via:
  - → (simple) substitution,
  - → sister adjunction,
  - → wrapping substitution (Kallmeyer et al., 2013; Osswald and Kallmeyer, 2018).

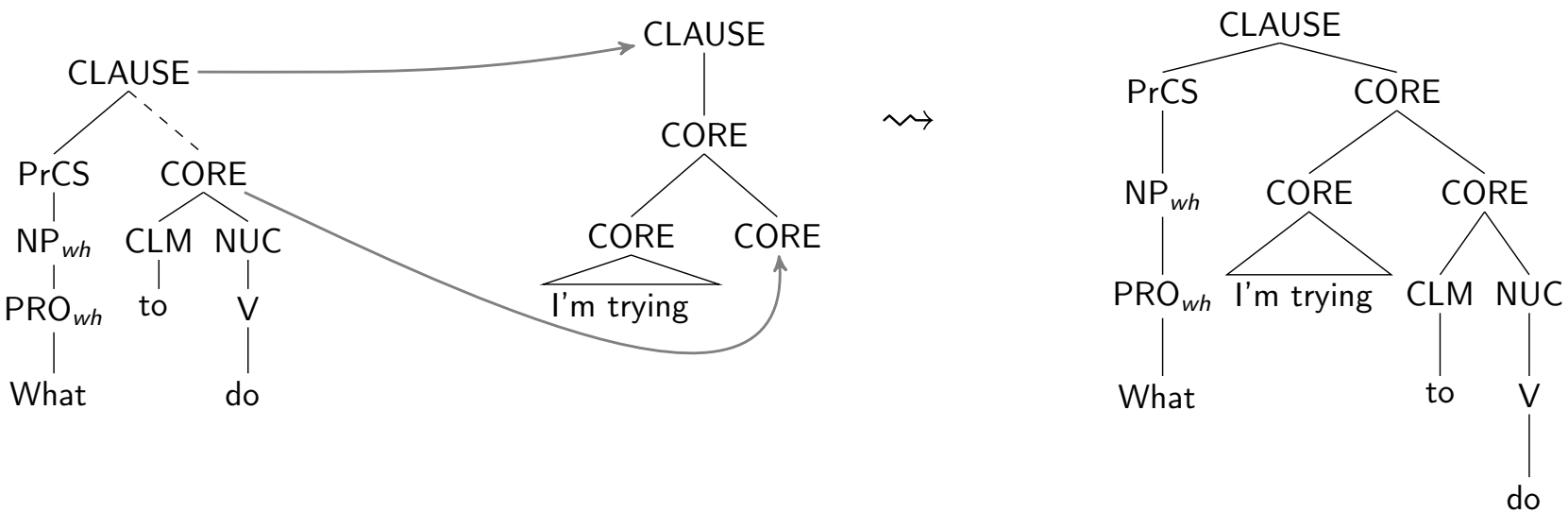


Figure 1: Wrapping substitution and a long distance dependency (LDD).

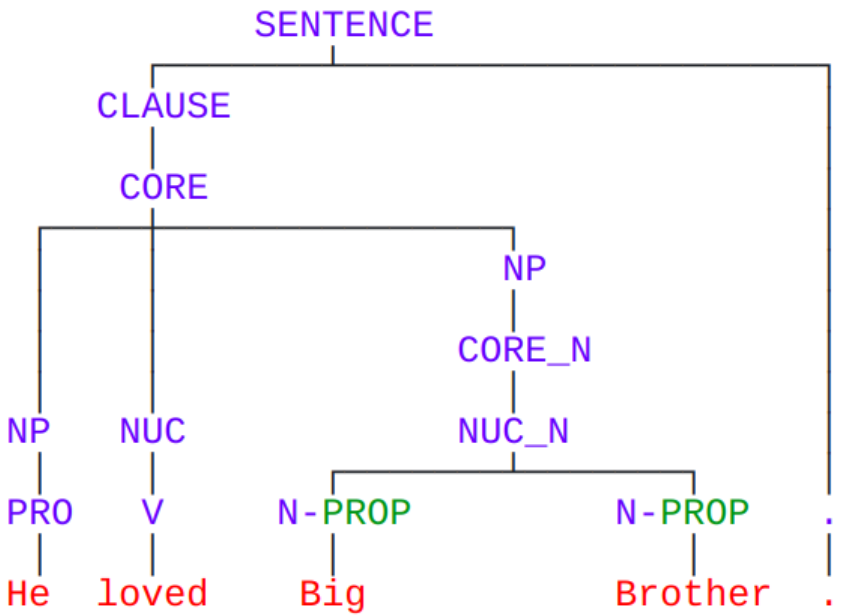
RRGbank and RRGparbank

- Corpora of RRG annotated sentences

  → RRG = Role and Reference Grammar (Van Valin and LaPolla, 1997; Van Valin Jr, 2005)

- PTB and George Orwell's novel '1984' with translations

  → annotated with RRG structures.



TWG Extraction: initial and auxiliary trees

- TWG extraction algorithm based on Xia (1999) for TAG.
- Percolation tables for head and modifier distinction.

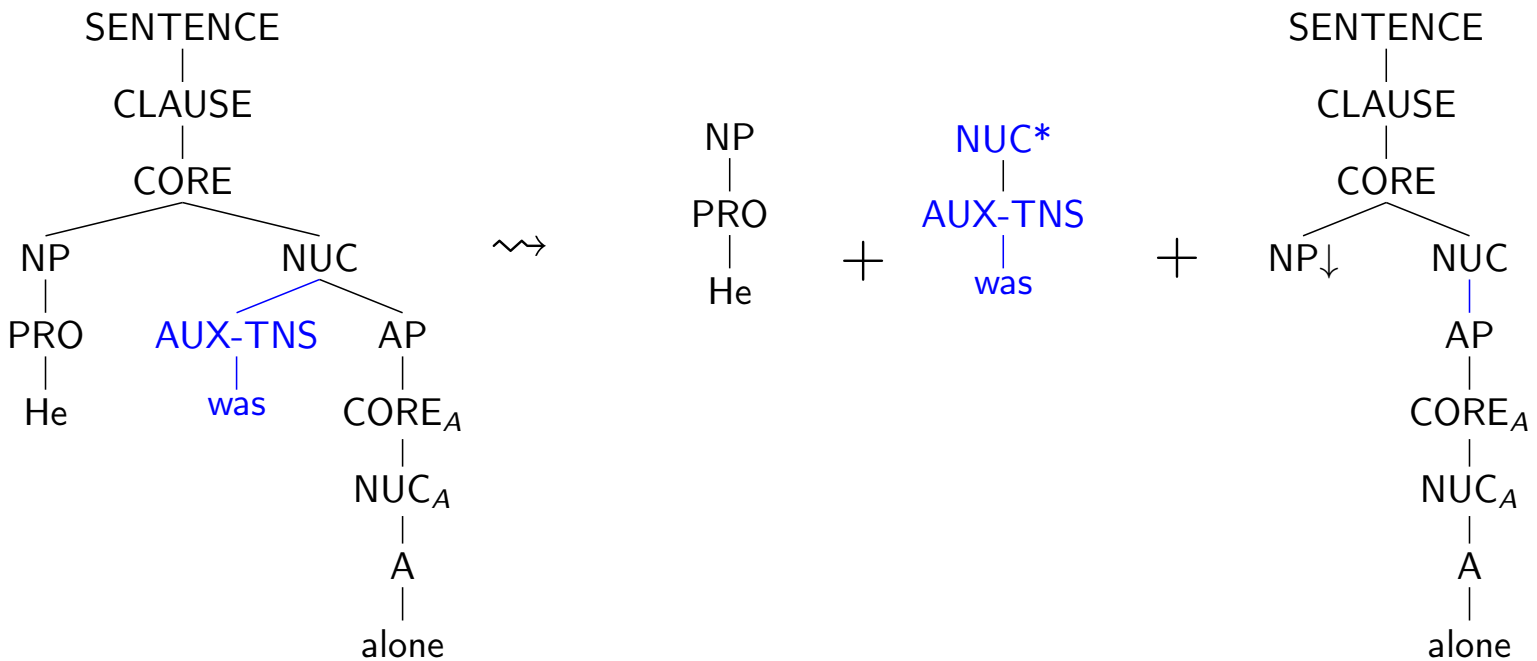


Figure 2: Extraction of initial and sister-adjoining trees.

TWG Extraction: d-edge trees for LDDs

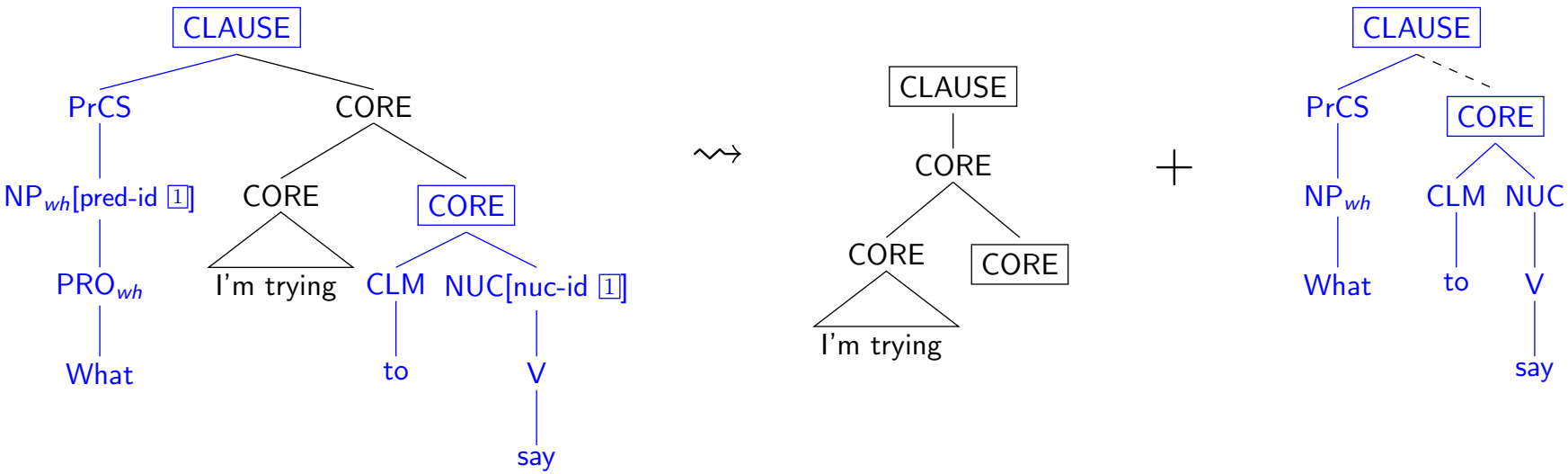


Figure 3: Extraction of a target tree and an elementary tree with a long distance dependency (LDD).

Extracted TWG Grammar

Parameters	TWG
Supertags	**3125**
Supertags occuring once	**1858**
Avg. sentence length	10.97
Sentences	7656
# initial trees	1527
# sister-adjoining trees	1549
# d-edge trees	49
# Long-dist. dependencies (LDDs)	**78**

Table 1: Statistics on subcorpora and extracted grammars.

Parsing architecture

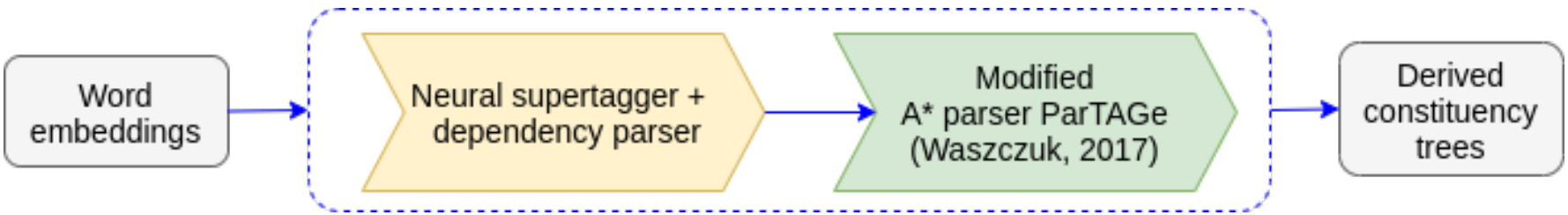


Figure 4: Pipeline of our neural statistical TWG parsing architecture.

Evaluation

- Train, dev, and test split: 4960, 551, and 2145 sentences

	DiscoDOP		Discoparset		ParTAGe	
	dev	test	dev	test	dev	test
Unlabeled Attachment Score	–	–	–	–	87.74	87.67
Supertagging accuracy	–	–	–	–	74.25	75.81
POS-tagging accuracy	92.02	93.25	$94.24_{(+3.04)}$	$94.92_{(+2.69)}$	**94.63**	**95.07**
Exactly matching parses	29.04	32.87	$28.68_{(+8.89)}$	$28.30_{(+13.19)}$	**36.12**	**38.32**
Labeled F1	79.26	80.96	$83.57_{(+6.83)}$	$84.56_{(+6.39)}$	**85.26**	**85.26**

Table 2: Parsing results compared with DiscoDOP[1] and Discoparset[2]. The numbers in subscript represent the relative gain provided by BERT[3].

[1] DiscoDOP (van Cranenburgh et al., 2016)
[2] Discoparset (Coavoux and Cohen, 2019)
[3] BERT (Devlin et al., 2019)

Errors in LDD prediction

Predicted LDDs	**DiscoDOP**	**Discoparset**	**ParTAGe**
	test	test	test
# true positives	13	14	**22**
# false positives	7	0	0
# false negatives	14	13	**5**

Table 3: Prediction of long-distance dependencies (LDDs) on test data.

Error analysis for LDD prediction

- Errors in POS-tag predictions.

  What is one ***to think*** *of all this?* (*is* tagged AUX instead of V)

- Wrong attachment of an LDD part (if the relative or wh phrase of the LDD is an adjunct).

  And why do you imagine that we ***bring*** *people to this place?*

- Cases where the embedding verb also has a strong tendency to take a wh-element as argument.

  [. . .] slip of paper which they said ***was the bill***

Perspectives

Linguistic resources

- Corpus-based wide-coverage RRG grammar for English
- Proof of concept for TWG, in particular wrt. non-local dependencies

Parsing

- Wide-coverage probabilistic parsing with TWGs
- Capturing LDDs in constituency parsing

THANK YOU VERY MUCH FOR YOUR ATTENTION!

References I

Coavoux, M. and Cohen, S. B. (2019). Discontinuous constituency parsing with a stack-free transition system and a dynamic oracle. 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 204–217, Minneapolis, Minnesota. Association for Computational Linguistics.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). 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. Association for Computational Linguistics.

Kallmeyer, L., Osswald, R., and Van Valin, Jr., R. D. (2013). Tree Wrapping for Role and Reference Grammar. In Morrill, G. and Nederhof, M.-J., editors, *Formal Grammar 2012/2013*, volume 8036 of *LNCS*, pages 175–190. Springer.

Osswald, R. and Kallmeyer, L. (2018). Towards a formalization of role and reference grammar. In Kailuweit, R., Künkel, L., and Staudinger, E., editors, *Applying and Expanding Role and Reference Grammar.*, pages 355–378. Albert-Ludwigs-Universität, Universitätsbibliothek. [NIHIN studies], Freiburg.

van Cranenburgh, A., Scha, R., and Bod, R. (2016). Data-oriented parsing with discontinuous constituents and function tags. *Journal of Language Modelling*, 4(1):57–111.

Van Valin, Jr., R. D. and LaPolla, R. (1997). *Syntax: Structure, meaning and function*. Cambridge University Press.

Van Valin Jr, R. D. (2005). *Exploring the syntax-semantics interface*. Cambridge University Press.

Xia, F. (1999). Extracting tree adjoining grammars from bracketed corpora. In *Proceedings of the 5th Natural Language Processing Pacific Rim Symposium (NLPRS-99)*, pages 398–403.