Old School vs. New School: Comparing Transition-Based Parsers with and without Neural Network Enhancement

Miryam de Lhoneux, Sara Stymne & Joakim Nivre

UPPSALA UNIVERSITET

TLT15, 20 January 2016

Overview

Introduction

Motivation

Introduction

Motivation

- UD treebanks
  - are growing in number and size at a fast pace
  - make our models generalisable across languages and domains

Introduction

Motivation

- UD treebanks
  - are growing in number and size at a fast pace
  - make our models generalisable across languages and domains
- Parsing models enhanced by neural networks
  - have seen a large boost in accuracy
  - are expensive to optimise

Introduction

Motivation

- UD treebanks
  - are growing in number and size at a fast pace
  - make our models generalisable across languages and domains
- Parsing models enhanced by neural networks
  - have seen a large boost in accuracy
  - are expensive to optimise

Our proposal

Work on a small but representative sample of UD treebanks

Introduction

Introduction

Motivation

Straka et al. (2015) trained Parsito, a neural network parser for UD
Limited comparison with MaltParser (Nivre et al., 2007).

Introduction

Motivation

Straka et al. (2015) trained Parsito, a neural network parser for UD
Limited comparison with MaltParser (Nivre et al., 2007).

This Work

Carry out a comparison of the 2 on a small sample

Treebank Sampling

Selection criteria

Treebank Sampling

Selection criteria

- typological variety:

Treebank Sampling

Selection criteria

- typological variety:
  - 8 different fine-grained and 5 coarse-grained families

Treebank Sampling

Selection criteria

- typological variety:
  - 8 different fine-grained and 5 coarse-grained families
  - isolating, morphologically-rich and inflecting languages

Treebank Sampling

Selection criteria

- typological variety:
  - 8 different fine-grained and 5 coarse-grained families
  - isolating, morphologically-rich and inflecting languages
  - non-projectivity

Treebank Sampling

Selection criteria

- typological variety:
  - 8 different fine-grained and 5 coarse-grained families
  - isolating, morphologically-rich and inflecting languages
  - non-projectivity
- variety in treebanks sizes and domains

Treebank Sampling

Selection criteria

- typological variety:
  - 8 different fine-grained and 5 coarse-grained families
  - isolating, morphologically-rich and inflecting languages
  - non-projectivity
- variety in treebanks sizes and domains
- availability of morphological features

Treebank Sampling

Selection criteria

- typological variety:
  - 8 different fine-grained and 5 coarse-grained families
  - isolating, morphologically-rich and inflecting languages
  - non-projectivity
- variety in treebanks sizes and domains
- availability of morphological features
- quality of the treebank (according to UD validation tests)

Treebank Sampling

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	sino-tibetan	sinitic	isolating language
Finnish	uralo-altaic	finno-ungric	many different domains
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	dravidian	tamil	small; language family
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	**indo-eur**	**balto-slavic**	**largest UD treebank**
Chinese	sino-tibetan	sinitic	isolating language
Finnish	uralo-altaic	finno-ungric	many different domains
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	dravidian	tamil	small; language family
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	**sino-tibetan**	**sinitic**	**isolating language**
Finnish	uralo-altaic	finno-ungric	many different domains
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	dravidian	tamil	small; language family
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	sino-tibetan	sinitic	isolating language
Finnish	**uralo-altaic**	**finno-ungric**	**many different domains**
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	dravidian	tamil	small; language family
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	sino-tibetan	sinitic	isolating language
Finnish	uralo-altaic	finno-ungric	many different domains
English	**indo-eur**	**germanic**	**largest + full manual check**
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	dravidian	tamil	small; language family
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	sino-tibetan	sinitic	isolating language
Finnish	uralo-altaic	finno-ungric	many different domains
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	**indo-eur**	**hellenic**	**amount of non-projectivity**
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	dravidian	tamil	small; language family
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	sino-tibetan	sinitic	isolating language
Finnish	uralo-altaic	finno-ungric	many different domains
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	**uralo-altaic**	**turkic**	**smallest; full manual check**
Tamil	dravidian	tamil	small; language family
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	sino-tibetan	sinitic	isolating language
Finnish	uralo-altaic	finno-ungric	many different domains
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	**dravidian**	**tamil**	**small; language family**
Hebrew	afro-asiatic	semitic	language family

Table 1 : Treebank Sample

Treebank Sampling

language	coarse	fine	main argument for inclusion
Czech	indo-eur	balto-slavic	largest UD treebank
Chinese	sino-tibetan	sinitic	isolating language
Finnish	uralo-altaic	finno-ungric	many different domains
English	indo-eur	germanic	largest + full manual check
AncientGreekPR	indo-eur	hellenic	amount of non-projectivity
Kazakh	uralo-altaic	turkic	smallest; full manual check
Tamil	dravidian	tamil	small; language family
Hebrew	**afro-asiatic**	**semitic**	**language family**

Table 1 : Treebank Sample

Comparing Parsing Accuracy

Comparing Parsing Accuracy

UDPipe (Straka et al., 2016) :

- Pretrained models
- Optimised using UDPipe's random hyperparameter search

Comparing Parsing Accuracy

UDPipe (Straka et al., 2016) :

- Pretrained models
- Optimised using UDPipe's random hyperparameter search

MaltParser:

- Arc-standard swap system with lazy oracle
- Feature model optimised with MaltOptimizer (Ballesteros and Nivre, 2016).

Comparing Parsing Accuracy

UDPipe (Straka et al., 2016) :

- Pretrained models
- Optimised using UDPipe's random hyperparameter search

MaltParser:

- Arc-standard swap system with lazy oracle
- Feature model optimised with MaltOptimizer (Ballesteros and Nivre, 2016).

UDPipe used for tagging POS and morphological features

Comparing Parsing Accuracy

UDPipe (Straka et al., 2016) :

- Pretrained models
- Optimised using UDPipe's random hyperparameter search

MaltParser:

- Arc-standard swap system with lazy oracle
- Feature model optimised with MaltOptimizer (Ballesteros and Nivre, 2016).

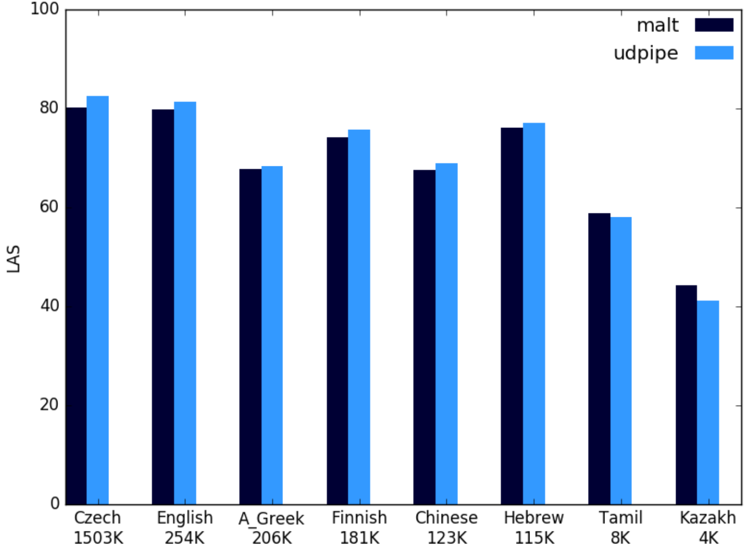
UDPipe used for tagging POS and morphological features

SyntaxNet (Andor et al., 2016) :

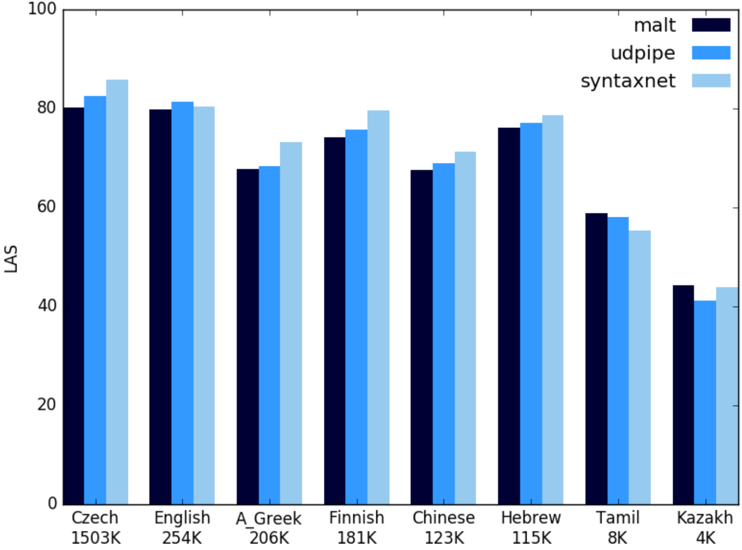
- *reported results*

Comparing Parsing Accuracy

Comparing Parsing Accuracy



Comparing Parsing Accuracy



Impact of Training Size on Neural Network Parsing

Learning curve experiment:

Impact of Training Size on Neural Network Parsing

Learning curve experiment:

MaltParser: arc-standard swap with lazy oracle and extended feature model
UDPipe: arc-standard swap with lazy oracle and default hyperparameters

Impact of Training Size on Neural Network Parsing

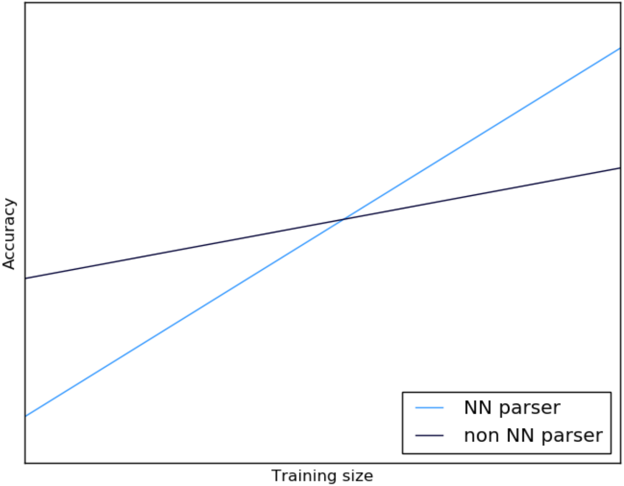
Learning curve experiment:

MaltParser: arc-standard swap with lazy oracle and extended feature model
UDPipe: arc-standard swap with lazy oracle and default hyperparameters

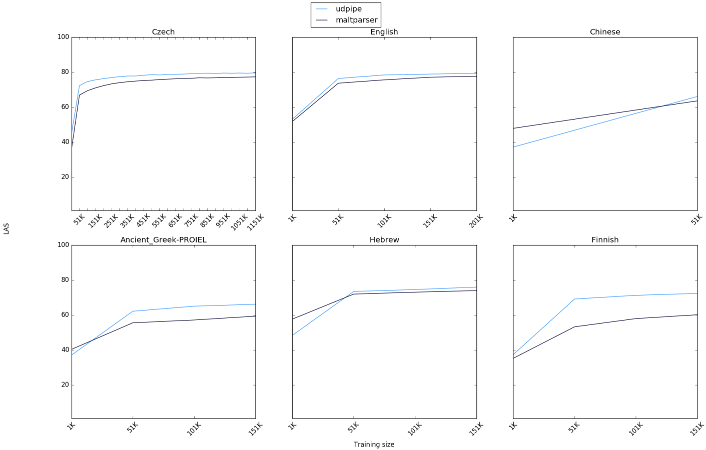
Splits: 1K to max with 50K splits
zoom on small data sizes: 1K to 15K

Impact of Training Size on Neural Network Parsing

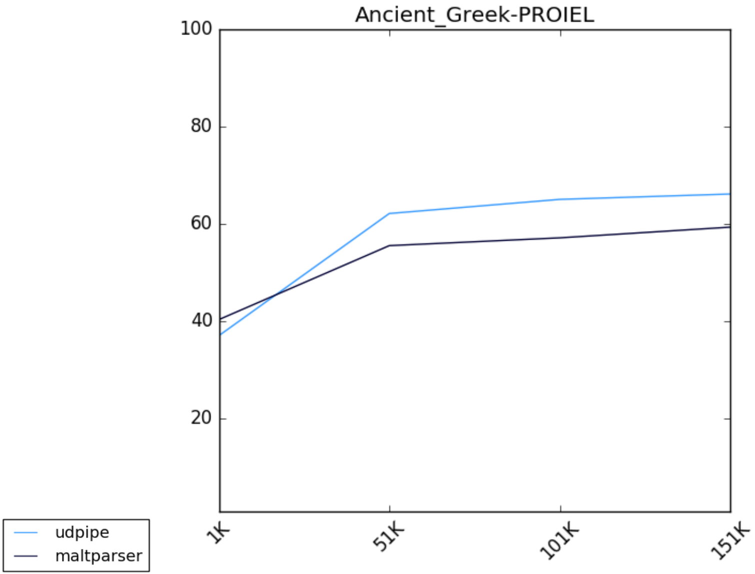
Expectation



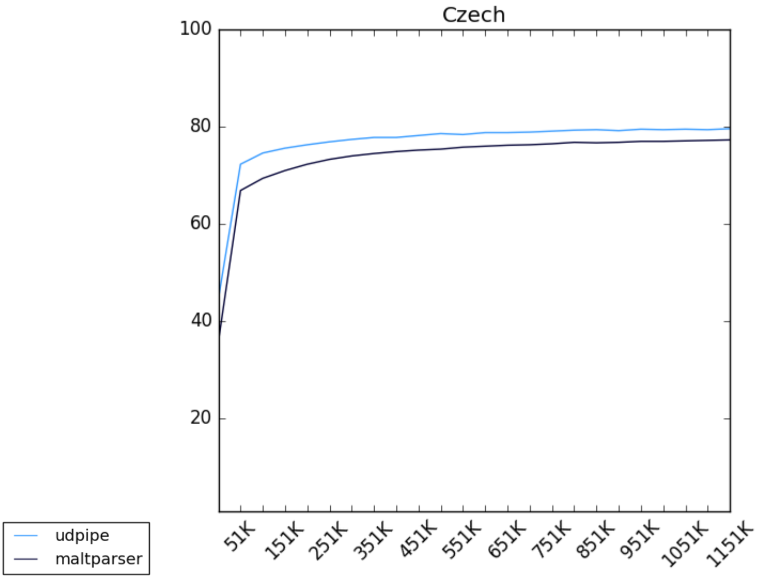
Reality


udpipe
maltparser
Czech
English
Chinese
Ancient_Greek-PROIEL
Hebrew
Finnish
LAS
Training size
100
80
60
40
20
51K
151K
251K
351K
451K
551K
651K
751K
851K
951K
1051K
1151K
1K
101K
201K

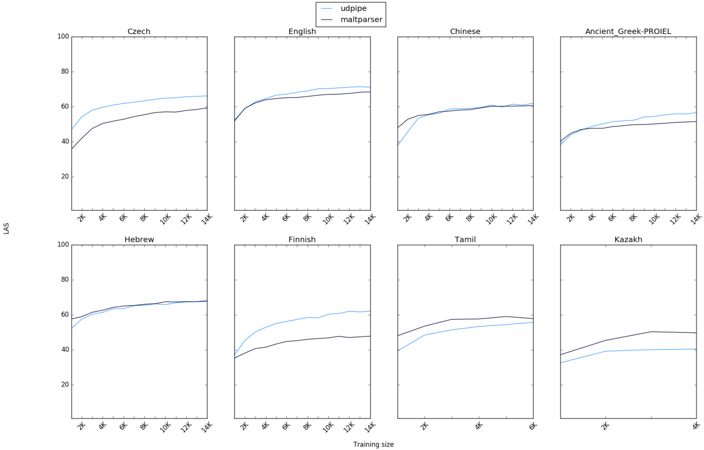
Impact of Training Size on Neural Network Parsing



Impact of Training Size on Neural Network Parsing



Impact of Training Size on Neural Network Parsing



Error Analysis

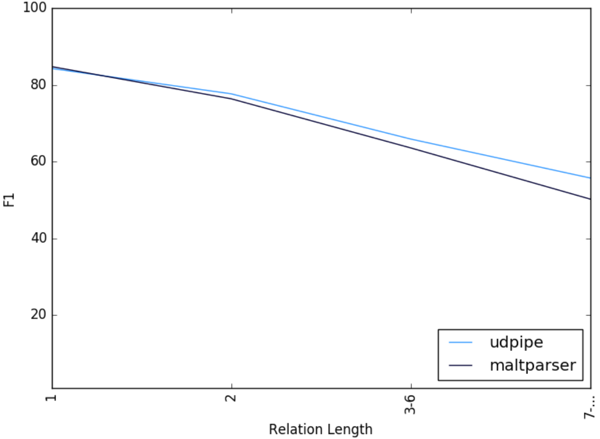
Error Analysis

- Inspired by McDonald and Nivre (2007): comparison of the accuracy of 2 parsers on a variety of graph and linguistics factors

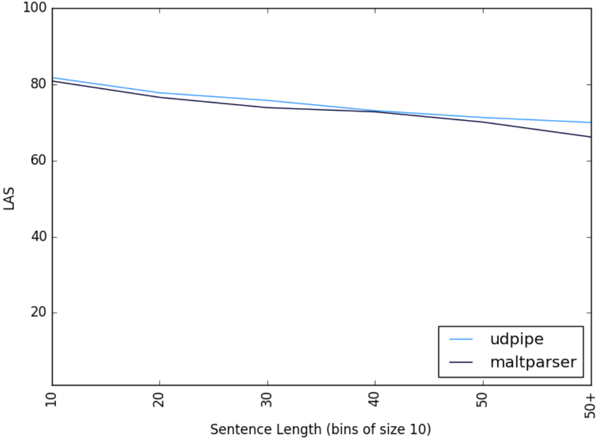
Error Analysis

- Inspired by McDonald and Nivre (2007): comparison of the accuracy of 2 parsers on a variety of graph and linguistics factors
- Concatenating 9K of all development sets + all development set for Kazakh and Tamil

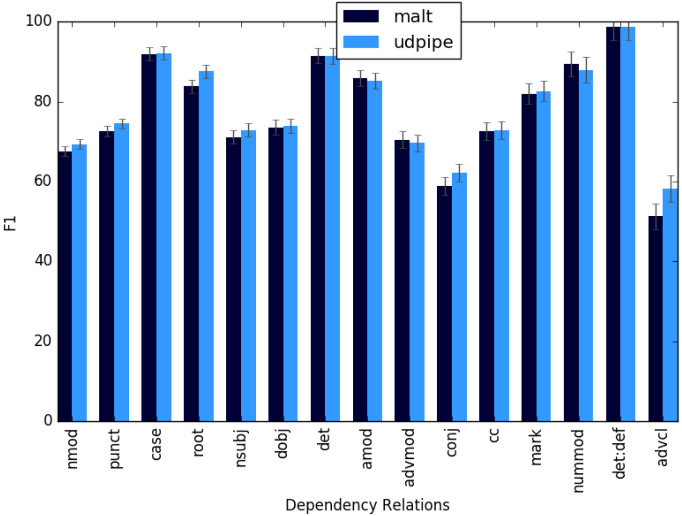
Error Analysis: Relation Length



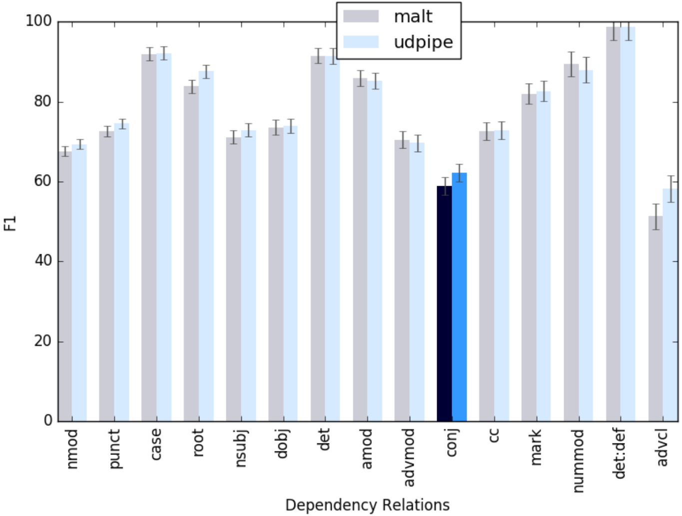
Error Analysis: Sentence Length



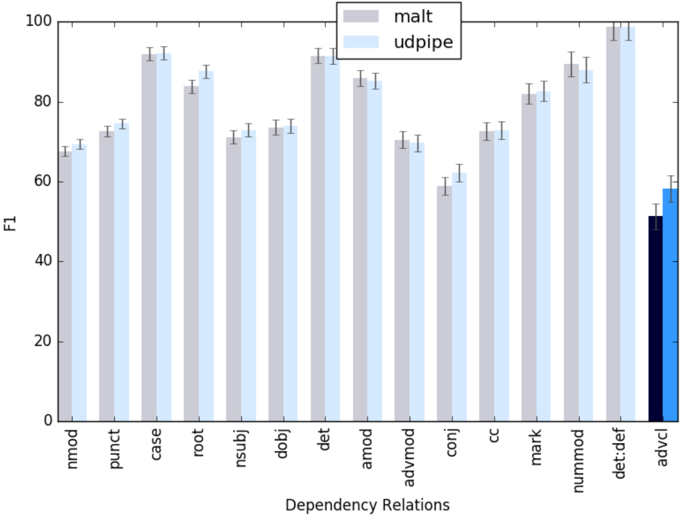
Error Analysis: Dependency Relations



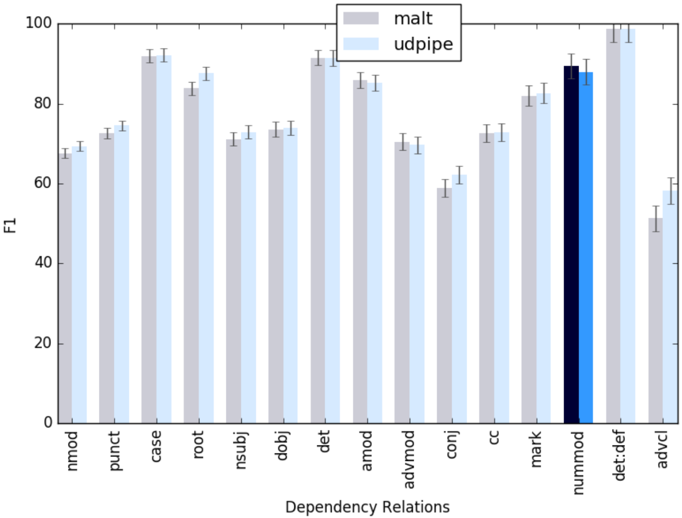
Error Analysis: Dependency Relations



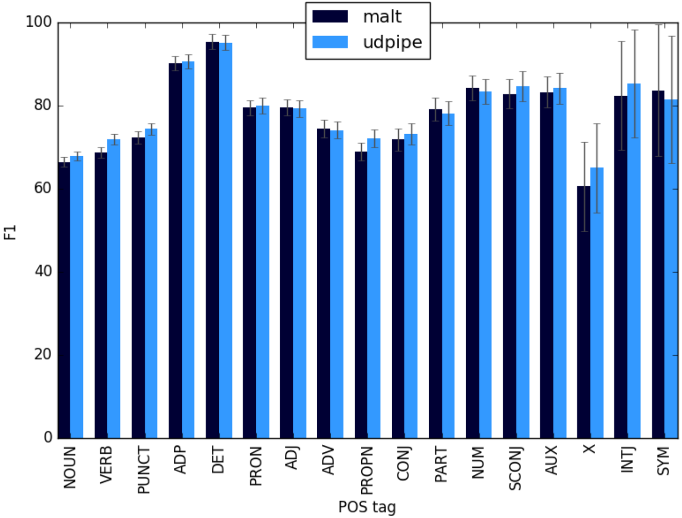
Error Analysis: Dependency Relations



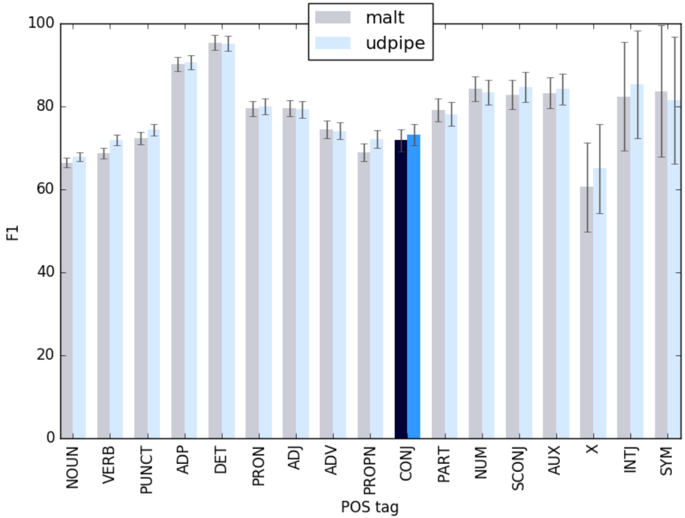
Error Analysis: Dependency Relations



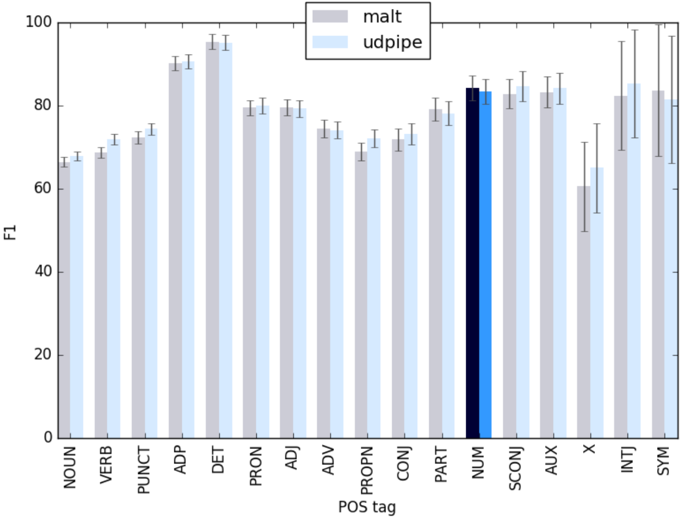
Error Analysis: POS tags



Error Analysis: POS tags



Error Analysis: POS tags



Conclusion and Future Work

Conclusion and Future Work

Conclusion

Conclusion and Future Work

Conclusion

- UDPipe outperforms MaltParser with large training size

Conclusion and Future Work

Conclusion

- UDPipe outperforms MaltParser with large training size
- The learning curve of UDPipe is steeper than MaltParser

Conclusion and Future Work

Conclusion

- UDPipe outperforms MaltParser with large training size
- The learning curve of UDPipe is steeper than MaltParser
- But flattens out in the same way as MaltParser

Conclusion and Future Work

Conclusion

- UDPipe outperforms MaltParser with large training size
- The learning curve of UDPipe is steeper than MaltParser
- But flattens out in the same way as MaltParser
- MaltParser suffers more than UDPipe from longer sentences/dependencies

Conclusion and Future Work

Conclusion

- UDPipe outperforms MaltParser with large training size
- The learning curve of UDPipe is steeper than MaltParser
- But flattens out in the same way as MaltParser
- MaltParser suffers more than UDPipe from longer sentences/dependencies

Future Work

Conclusion and Future Work

Conclusion

- UDPipe outperforms MaltParser with large training size
- The learning curve of UDPipe is steeper than MaltParser
- But flattens out in the same way as MaltParser
- MaltParser suffers more than UDPipe from longer sentences/dependencies

Future Work

- Effect of beam search

Conclusion and Future Work

Conclusion

- UDPipe outperforms MaltParser with large training size
- The learning curve of UDPipe is steeper than MaltParser
- But flattens out in the same way as MaltParser
- MaltParser suffers more than UDPipe from longer sentences/dependencies

Future Work

- Effect of beam search
- More fine-grained error analysis

References

Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. 2016. Globally normalized transition-based neural networks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.

Miguel Ballesteros and Joakim Nivre. 2016. MaltOptimizer: Fast and effective parser optimization. *Natural Language Engineering* 22(2):187–213.

Ryan McDonald and Joakim Nivre. 2007. Characterizing the errors of data-driven dependency parsing models. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. pages 122–131.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülşen Eryiğit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(2):95–135.

Milan Straka, Jan Hajič, and Straková. 2016. UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. European Language Resources Association (ELRA), Paris, France.

Milan Straka, Jan Hajič, Jana Straková, and Jan Hajič jr. 2015. Parsing universal dependency treebanks using neural networks and search-based oracle. In *Proceedings of Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT 14)*.