

# (Re)ranking Meets Morphosyntax: State-of-the-art Results from the SPMRL 2013 Shared Task



Anders Björkelund<sup>§</sup> Özlem Çetinoğlu<sup>§</sup> Richárd Farkas<sup>†</sup> Thomas Müller<sup>§,‡</sup> Wolfgang Seeker<sup>§</sup>

<sup>§</sup>Institute for Natural Language Processing, Stuttgart, Germany  
{anders,ozlem,muellets,seeker}@ims.uni-stuttgart.de

<sup>†</sup>Department of Informatics, Szeged, Hungary  
rfarkas@inf.u-szeged.hu

<sup>‡</sup>Center for Information and Language Processing, Munich, Germany

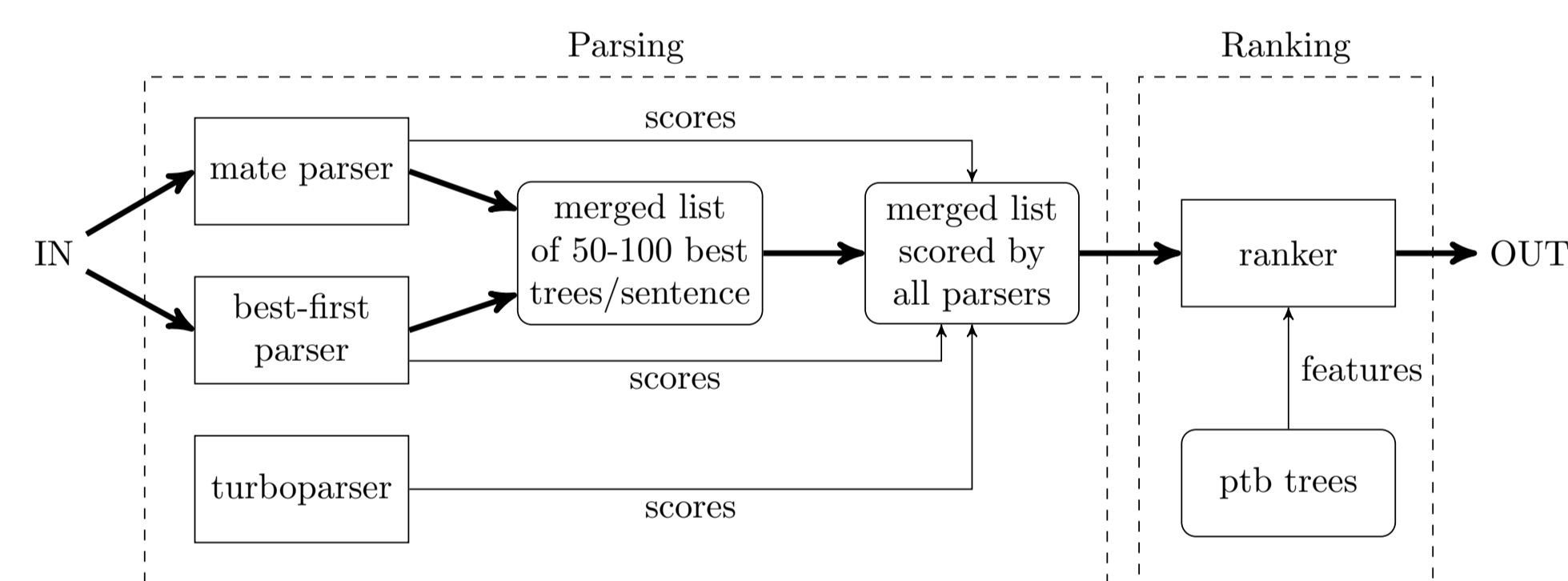
## Preprocessing

- ▶ Predicted POS and morphology using MarMoT (Müller et al., 2013)
- ▶ MarMoT extended to use features from morphological analyzers
- ▶ Provided predicted tags are also integrated into feature models
  - ▶ Improvements on Swedish, Polish, Basque
- ▶ Leads to better results for dependency parsing (mate parser) compared to provided predicted morphology:

	Arabic	Basque	French	German	Hebrew	Hungarian	Korean	Polish	Swedish
Shared Task	-	83.50	84.49	90.85	75.89	82.84	82.39	85.81	<b>77.16</b>
Δ	-	0.93	0.35	0.61	3.48	1.57	3.37	0.49	-0.11
MarMoT	<b>85.42</b>	<b>84.43</b>	<b>84.84</b>	<b>91.46</b>	<b>79.37</b>	<b>84.41</b>	<b>85.76</b>	<b>86.30</b>	77.05

Table: LAS for the mate parser using provided vs own POS/morphology predictions on the development sets.

## Dependency Parsing



- ▶ Parsers
  - ▶ The mate parser (Bohnet, 2010) (best)
  - ▶ TurboParser (Martins et al., 2010) (second)
  - ▶ In-house Best-first parser (Goldberg and Elhadad, 2010) (third)
- ▶ Modified the parsers to
  - ▶ give n-best output
  - ▶ output scores given a tree
- ▶ Merging n-best lists improves oracle scores
- ▶ (Re)ranking of merged n-best lists
  - ▶ Has not been done before (on merged lists)

	Arabic	Basque	French	German	Hebrew	Hungarian	Korean	Polish	Swedish
mate	85.42	84.43	84.84	91.46	79.37	84.41	85.76	86.30	77.05
Δ	1.32	1.18	0.77	1.22	1.65	0.36	1.36	1.39	1.52
Ranked	<b>86.74</b>	<b>85.61</b>	<b>85.96</b>	<b>92.68</b>	<b>81.02</b>	<b>84.77</b>	<b>87.12</b>	<b>87.69</b>	<b>78.57</b>
Oracle	90.71	91.91	90.43	97.44	87.18	88.76	94.65	95.29	84.96

Table: Baseline, ranker, and oracle LAS scores on the development sets.

- ▶ Ranker features (tuned for each language)
  - ▶ Scores from base parsers (and combinations)
    - ▶ Very important (cf. Zhang et al. (2009))
  - ▶ Projectivity features and ill-nestedness
  - ▶ Case agreement between heads and dependents
  - ▶ Function label uniqueness for certain labels
  - ▶ Phrase structure features based on paths in constituent trees

## Summary and Findings

- ▶ 1st in Dependency track
  - ▶ Good preprocessing improves results
  - ▶ Reranking n-best output from multiple parsers (1st time)
  - ▶ MarMoT and mate constitute strong off-the-shelf baseline
  - ▶ Our baseline ranked second
- ▶ 1st in Constituency track
  - ▶ Replacing rare words with morphology predictions improves
    - ▶ Particularly helpful for agglutinating languages
  - ▶ Known techniques from English parsing helps, e.g., Product-of-grammars and reranking

## Constituency Parsing

- ▶ Berkeley parser as a baseline (Petrov et al., 2006)
- ▶ Replacing rare words with their morphological tag (from MarMoT)
  - ▶ Improvements of up to 15% absolute (Basque)
- ▶ Product of 8 grammars (Petrov, 2010)
- ▶ Reranking following Charniak and Johnson (2005)

	Arabic	Basque	French	German	Hebrew	Hungarian	Korean	Polish	Swedish
Berkeley	78.24	69.17	79.74	81.74	87.83	83.90	70.97	84.11	74.50
Δ	0.46	<b>15.16</b>	-0.06	1.00	1.72	<b>5.18</b>	<b>11.87</b>	3.01	1.02
Replaced	78.70	84.33	79.68	82.74	89.55	89.08	82.84	87.12	75.52
Δ	1.60	<b>1.88</b>	1.74	1.82	0.94	0.72	1.31	1.20	3.73
Product	80.30	86.21	81.42	84.56	<b>90.49</b>	89.80	84.15	88.32	79.25
Δ	0.94	1.14	<b>1.07</b>	0.45	0.00	1.27	0.48	0.08	0.28
Reranked	<b>81.24</b>	<b>87.35</b>	<b>82.49</b>	<b>85.01</b>	<b>90.49</b>	<b>91.07</b>	<b>84.63</b>	<b>88.40</b>	<b>79.53</b>

Table: PARSEVAL scores on the development sets.

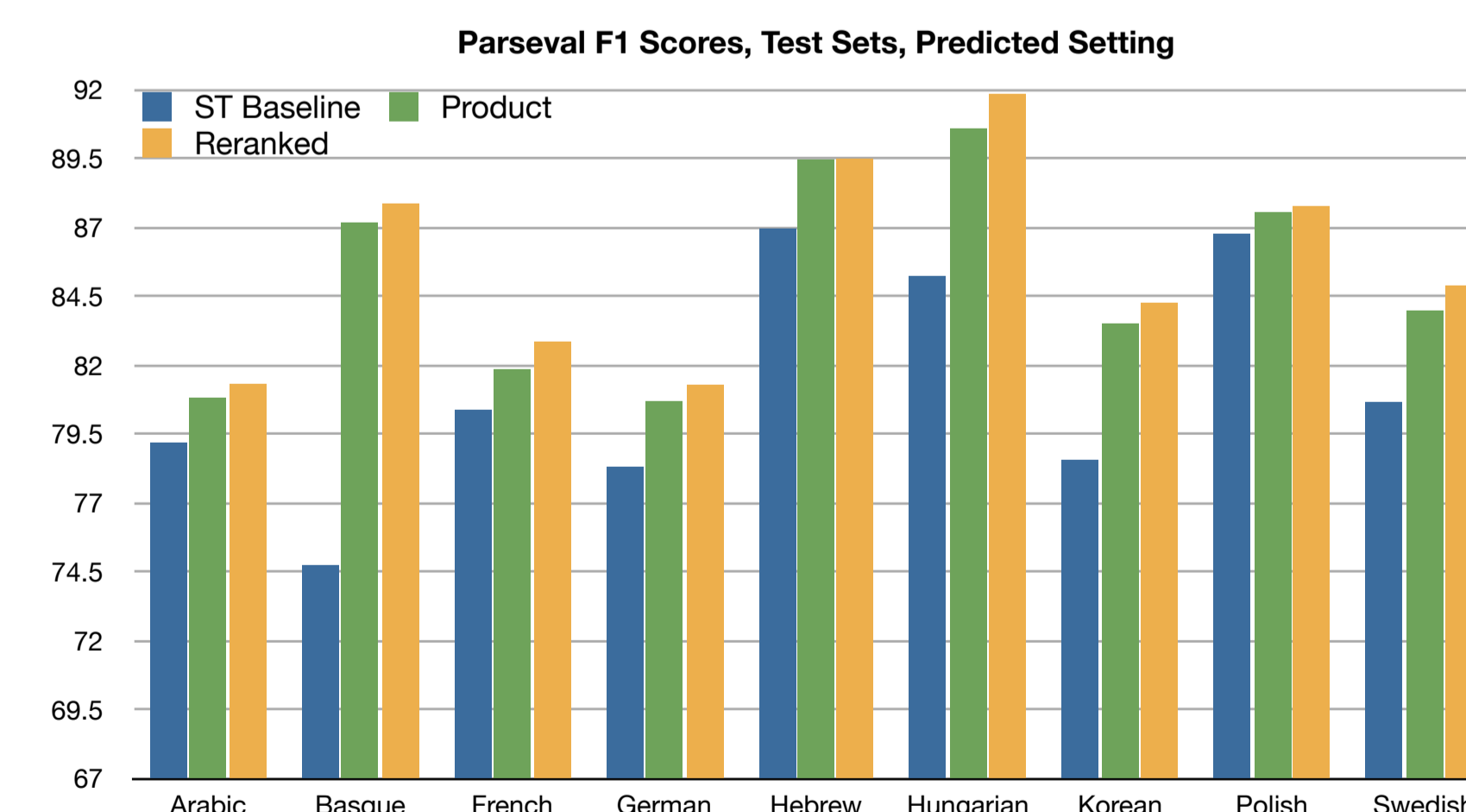
## Test Set Results

### Constituency Results:

- ▶ Our system outperforms the baseline on all languages
- ▶ Reranking consistently improves over 'Product' on all languages
  - ▶ up to 1.35% absolute (Hungarian)

	Arabic	Basque	French	German	Hebrew	Hungarian	Korean	Polish	Swedish
ST Baseline	79.19	74.74	80.38	78.30	86.96	85.22	78.56	86.75	80.64
Product	80.81	87.18	81.83	80.70	89.46	90.58	83.49	87.55	83.99
Reranked	<b>81.32</b>	<b>87.86</b>	<b>82.86</b>	<b>81.27</b>	<b>89.49</b>	<b>91.85</b>	<b>84.27</b>	<b>87.76</b>	<b>84.88</b>

Table: PARSEVAL F<sub>1</sub> scores for constituents on the test set.

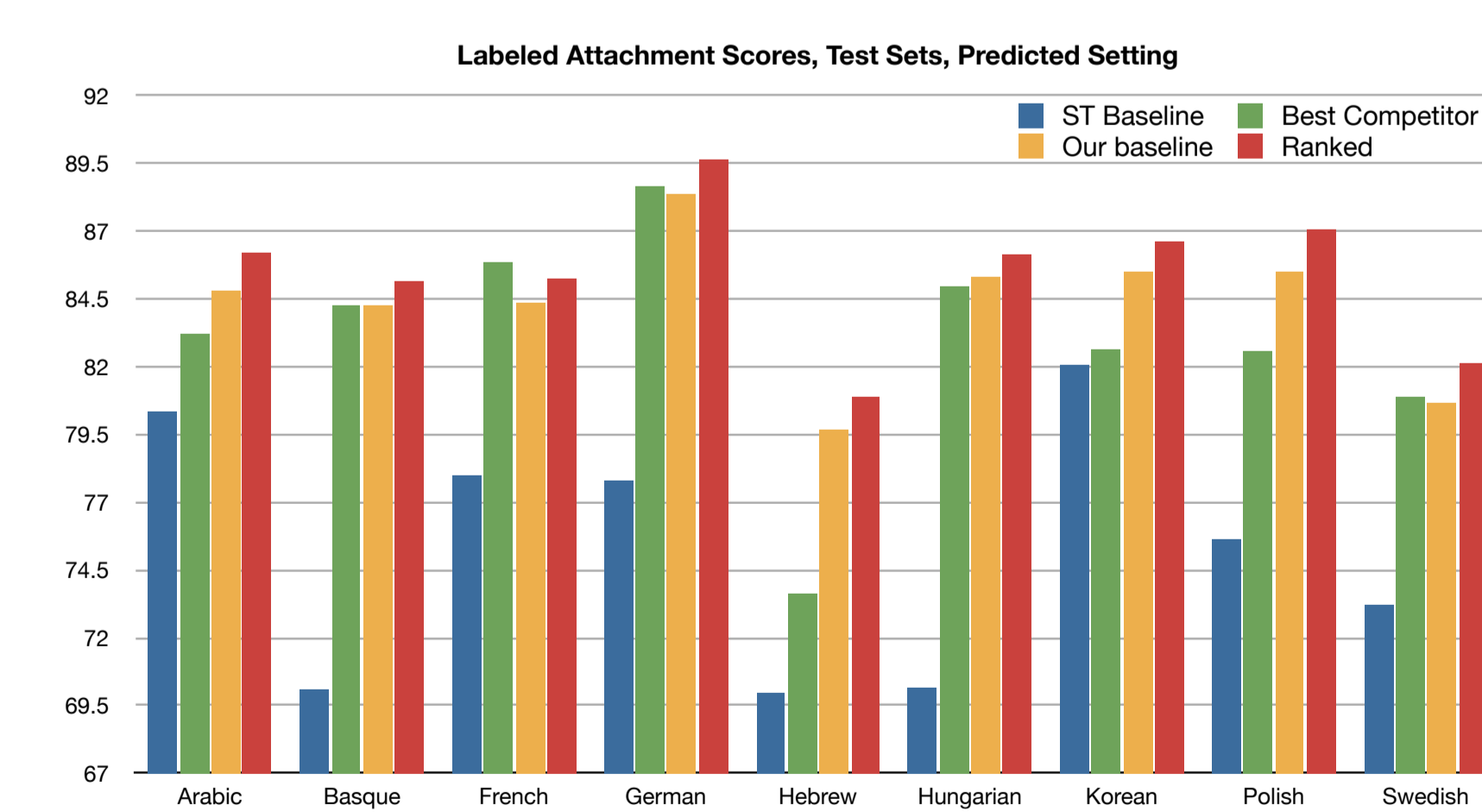


### Dependency Results:

- ▶ Achieved the best scores on all languages except French
- ▶ Our baseline is equal or better than the best competitor on 6 languages
- ▶ Ranking consistently improves over our baseline on all languages
  - ▶ up to 1.45% absolute (Polish)

	Arabic	Basque	French	German	Hebrew	Hungarian	Korean	Polish	Swedish
ST Baseline	80.36	70.11	77.98	77.81	69.97	70.15	82.06	75.63	73.21
Best Competitor	83.20	84.25	<b>85.86</b>	88.66	73.63	84.97	82.65	82.56	80.88
Baseline	84.81	84.25	84.37	88.37	79.67	85.31	85.51	85.51	80.67
Ranked	<b>86.21</b>	<b>85.14</b>	85.24	<b>89.65</b>	<b>80.89</b>	<b>86.13</b>	<b>86.62</b>	<b>87.07</b>	<b>82.13</b>

Table: LAS scores for dependencies on the test sets.



## References

- Bohnet, B. (2010). Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *COLING*.
- Charniak, E. and Johnson, M. (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *ACL*.
- Goldberg, Y. and Elhadad, M. (2010). An Efficient Algorithm for Easy-First Non-Directional Dependency Parsing. In *NAACL-HLT*.
- Martins, A., Smith, N., Xing, E., Aguiar, P., and Figueiredo, M. (2010). Turbo Parsers: Dependency Parsing by Approximate Variational Inference. In *EMNLP*.
- Müller, T., Schmid, H., and Schütze, H. (2013). Efficient Higher-Order CRFs for Morphological Tagging. In *EMNLP*.
- Petrov, S. (2010). Products of Random Latent Variable Grammars. In *NAACL-HLT*.
- Petrov, S., Barrett, L., Thibaux, R., and Klein, D. (2006). Learning accurate, compact, and interpretable tree annotation. In *COLING-ACL*.
- Zhang, H., Zhang, M., Tan, C. L., and Li, H. (2009). K-Best Combination of Syntactic Parsers. In *EMNLP*.

## Acknowledgments

Richárd Farkas is funded through the project FuturICT.hu (grant no.: TÁMOP-4.2.2.C-11/1/KONV-2012-0013). Thomas Müller is supported by a Google Europe Fellowship in NLP. The remaining authors are funded by the Deutsche Forschungsgemeinschaft (DFG) via the SFB 732, projects D2 and D8.