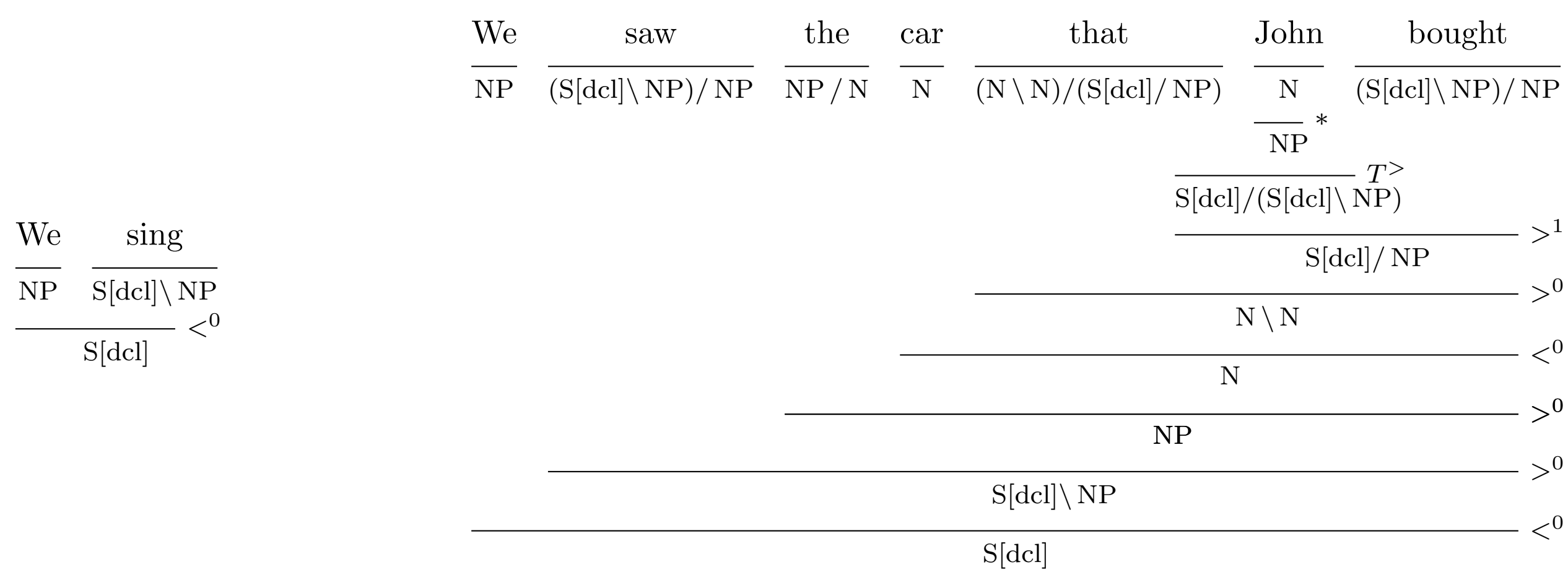


Combinatory Categorical Grammar (CCG)



- ✓ elegant handling of coordination
- ✓ universal rules
- ✓ widely used in semantic parsing
- ? what about languages without CCGbanks?

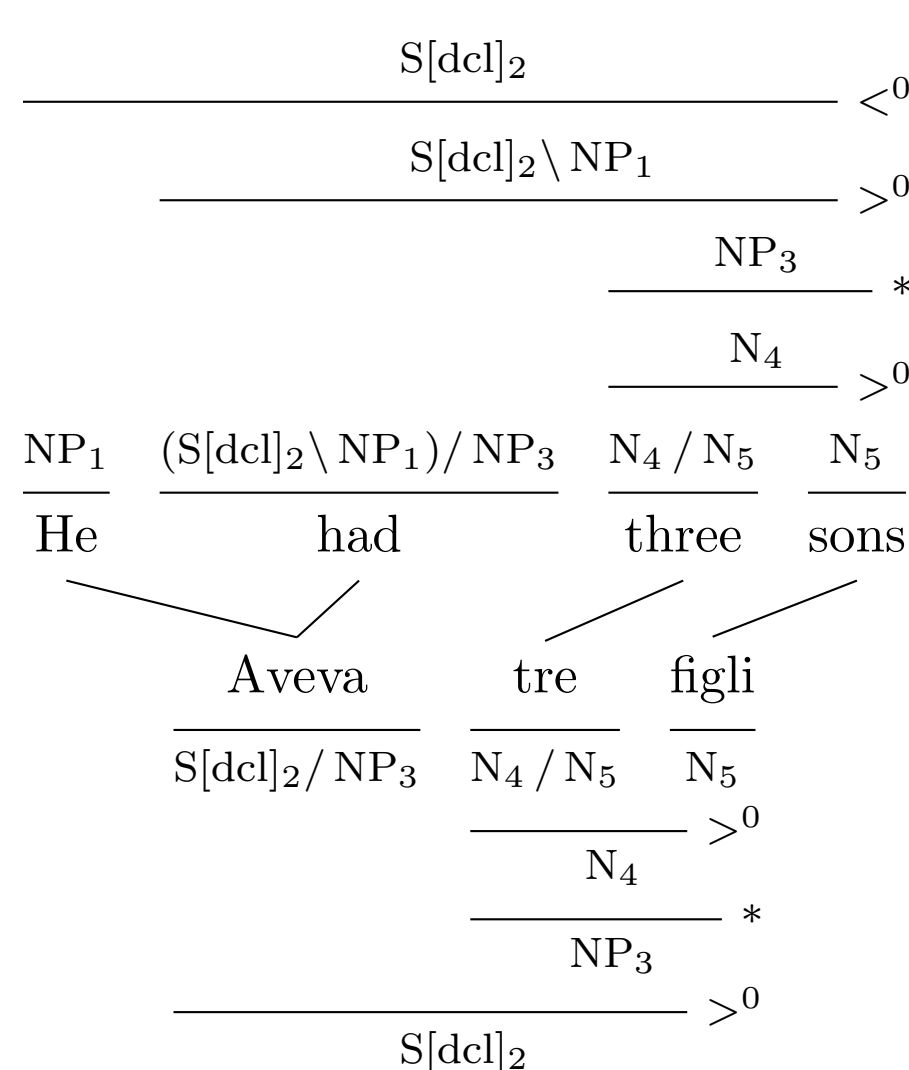
Parser Accuracy (UAS)

Language	ara	ces	dan	eus	nld	por	slv	swe
<i>Monolingual training on PASCAL (BH13: with gold POS tags)</i>								
Train tokens	5K	436K	25K	81K	79K	159K	54K	62K
BH13	.651	.507	.585	.450	.544	.629	.464	.669
BCH15	.437	.324	.377	.352	.438	.516	.236	.529
<i>Cross-lingual training on Tatoeba</i>								
Train tokens	20K	11K	21K	2K	44K	161K	835	24K
this work	.468	.449	.630	.290	.614	.678	.350	.637

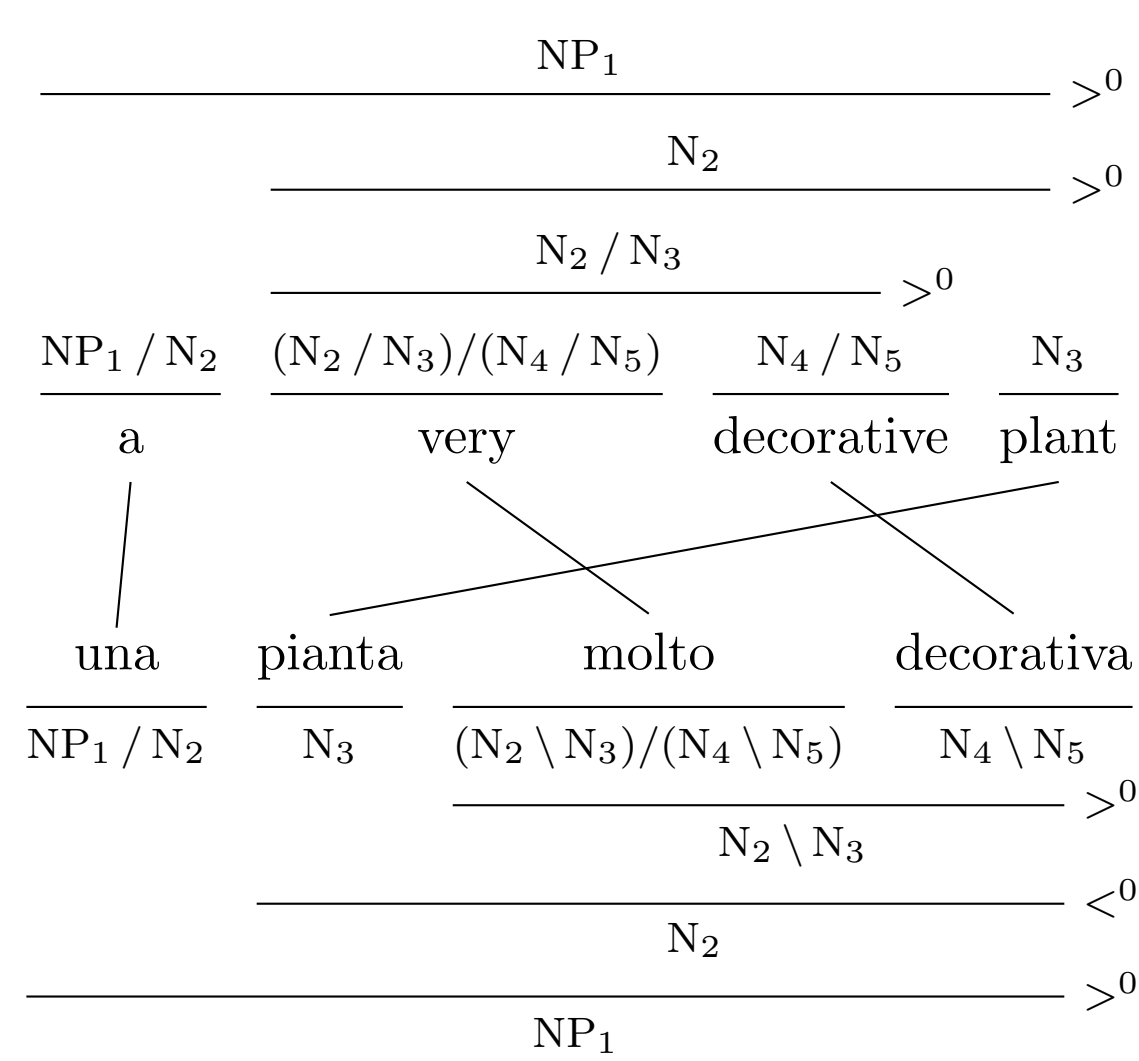
Training Strategy

- ♦ use parallel training corpus (Tatoeba)
- ♦ auto-parse source sentences
- ♦ unsupervised word alignment
- ♦ project derivations to target sentences
- ♦ train target parser

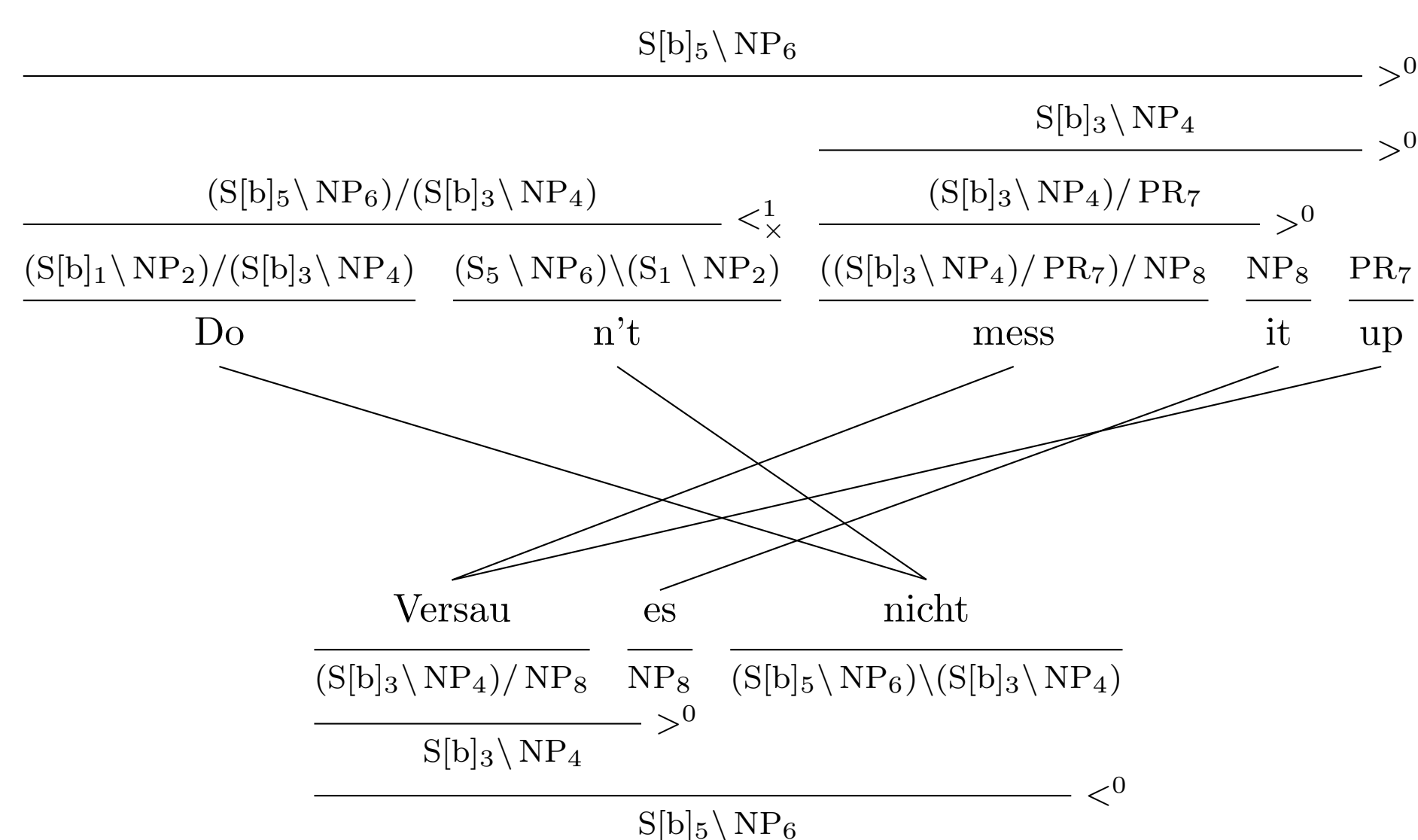
Annotation Projection (1)



Annotation Projection (2)



Annotation Projection (3)



Induced Categories (Frequency)

Category	eng	deu	ita	nld
<i>SVO and SOV</i>				
$(S[decl] \ NP) / NP$.0366	.0445	.0256	.0389
$(S[decl] \ NP) \ NP$.0000	.0056	.0046	.0061
$(S[b] \ NP) / NP$.0284	.0032	.0147	.0044
$(S[b] \ NP) \ NP$.0000	.0169	.0043	.0151
<i>do-support</i>				
$(S[decl] \ NP) / (S[b] \ NP)$.0237	.0184	.0150	.0180
<i>infinitives of particle verbs</i>				
$(S[b] \ NP) \ PR$.0000	.0000	.0000	.0000
$(S[b] \ NP) / PR$.0004	.0000	.0000	.0000
<i>prenominal vs. postnominal adjectives</i>				
N / N	.0309	.0299	.0213	.0316
$N \ N$.0013	.0018	.0099	.0018
$(N / N) / (N / N)$.0016	.0018	.0003	.0012
$(N \ N) / (N \ N)$.0001	.0000	.0008	.0000
<i>pro-drop</i>				
$S[decl]$.0000	.0000	.0012	.0001
$S[decl] / NP$.0004	.0013	.0115	.0007

Conclusions

- ✓ cross-lingual signal helps, compared to unsupervised induction
- ✓ induced categories reflect linguistic descriptions

References

- Bisk, Y., Christodoulopoulos, C., and Hockenmaier, J. (2015). Labeled grammar induction with minimal supervision. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 870–876. Association for Computational Linguistics.
- Bisk, Y. and Hockenmaier, J. (2013). An HDP model for inducing combinatory categorial grammars. *Transactions of the Association for Computational Linguistics*, 1:75–88.
- Steedman, M. (2001). *The Syntactic Process*. The MIT Press.
- Tatoeba (2019). Tatoeba: Collection of sentences and translations. <https://tatoeba.org/>. Accessed: 2019-07-16.