



# Dependency Relations of Light Verb Constructions in Four Languages

WG4



Veronika Vincze and István Nagy T.  
University of Szeged, Hungary  
{vinczev,nistvan}@inf.u-szeged.hu

## Motivation

- MWE representation in NLP resources and treebanks
- What syntactic features do MWEs have?
- Light verb constructions (LVCs) in treebanks
- Syntactic characteristics
- Interlingual comparisons

## The 4FX corpus

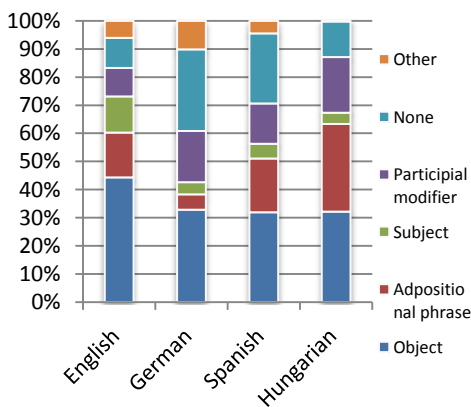
- Texts from JRC-Acquis
- Legal domain
- English, German, Spanish and Hungarian
- Approximately 100K tokens for each language
- Manual annotation for LVCs [4]

## Dependency parsing

- Automatic parsing
- Training datasets:
  - English: CoNLL-2008 shared task data [5] – Bohnet parser [1]
  - German: TIGER treebank [2] – Bohnet parser
  - Spanish: IULA treebank [3] – Bohnet parser
  - Hungarian: Szeged Dependency Treebank [6] – magyarlanc 2.0 [7]

## Standardized dependency relations

- Different treebanks use different labels although the grammatical relations are the same
- We standardized the dependency relations across treebanks and languages, e.g.:  
**dobj – OA – DO – obj**
- Prepositional phrases in English, German and Spanish are paired with a noun + oblique case suffix in Hungarian



	English		German		Spanish		Hungarian	
Object	280	44.30%	213	32.87%	273	31.89%	272	32.19%
Adpositional phrase	101	15.98%	35	5.40%	164	19.16%	263	31.12%
Subject	81	12.82%	28	4.32%	45	5.26%	34	4.02%
Participial modifier	64	10.13%	118	18.21%	122	14.25%	167	19.76%
<b>Sum</b>	<b>526</b>	<b>83.23%</b>	<b>394</b>	<b>60.80%</b>	<b>604</b>	<b>70.56%</b>	<b>736</b>	<b>87.10%</b>
None	68	10.76%	188	29.01%	214	25.00%	107	12.66%
Other	38	6.01%	66	10.19%	38	4.44%	2	0.24%
<b>Sum</b>	<b>632</b>	<b>100%</b>	<b>648</b>	<b>100%</b>	<b>801</b>	<b>100%</b>	<b>845</b>	<b>100%</b>

## Extracting potential LVCs

- Earlier approaches concentrated on verb-object pairs
- This restriction limits the range of LVCs to a great extent
- No syntactic constraints should be applied

## Possibilities for constructing treebanks

- Separate MWE label can mark the relationship among the members of the MWE
  - LVC label between *decision* and *make* in *make a decision*
  - When LVCs exhibit a uniform syntactic behaviour in the data
- A standard dependency label can be enhanced with MWE notation
  - dobj label at the dependency layer + LVC label at a different layer
  - Complex label: dobj+LVC
  - When there are several syntactic types of LVCs in the data
  - The inner structure of the LVC is transparent – modifiers can be added

## References

1. Bernd Bohnet. 2010. Top accuracy and fast dependency parsing is not a contradiction. In Proceedings of Coling 2010, pages 89–97.
2. Sabine Brants, Stefanie Dipper, Peter Eisenberg, Silvia Hansen-Schirra, Esther König, Wolfgang Leizius, Christian Rohrer, George Smith, and Hans Uszkoreit. 2004. TIGER: Linguistic interpretation of a German corpus. Research on Language and Computation, 2(4):597–620.
3. Montserrat Marimon, Beatriz Fisas, Núria Bel, Marta Villegas, Jorge Vivaldi, Sergi Torner, Mercè Lorente, Silvia Vázquez, and Marta Villegas. 2012. The IULA Treebank. In Proceedings of LREC-2012, pages 1920–1926, Istanbul, Turkey. ELRA.
4. Anita Rácz, István Nagy T., and Veronika Vincze. 2014. 4FX: Light Verb Constructions in a Multilingual Parallel Corpus. In Proceedings of LREC'14, Reykjavik, Iceland. ELRA.
5. Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Márquez, and Joakim Nivre. 2008. The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In Proceedings of the CoNLL-2008, pages 159–177. ACL.
6. Veronika Vincze, Dóra Szauter, Attila Almási, György Móra, Zoltán Alexin, and János Csirik. 2010. Hungarian Dependency Treebank. In Proceedings of LREC 2010.
7. János Zsibrita, Veronika Vincze, and Richárd Farkas. 2013. magyarlanc: A toolkit for morphological and dependency parsing of Hungarian. In Proceedings of RANLP, pages 763–771.