Word Segmentation in Universal Dependencies

Free Morphs

- en nice (property)
- en work (action or object)
- en now (property)
- cs pes 'dog' (object)
- en ouch (not a root)

Roots (+Affixes)

- cs plyn 'gas' (free root)
- tr ev-ler 'house-Plur' (free root with affix)
- it alber-o 'tree' (bound root with required affix)
- en re-place-ment (non-required affixes)
- cs Josef-ov-ým 'Josef-Poss-Ins' (opt ⇒ req affix)

Clitics

- en **the** book
- cs Smál **se**. 'He laughed.'
- es **de la** escuel-a 'of the school';
- de **auf der** Brücke 'on the bridge'

Compounds (+Affixes)

- en flower-pot
- de Auto-bahn 'highway'
- cs straš-pytel 'scaredy-cat'
- el γεω-γραφ-ία / geô-graf-ía 'geography' (affix)

Non-Haspelmath Words in UD

- de Liebe-s-brief 'love letter'
- cs ruk-o-pis 'manuscript'

- de am ⇒ an dem 'at the', im ⇒ in dem 'in the'
- fr au ⇒ à le 'to the'

Kilian Evang, Daniel Zeman

evang@hhu.de, zeman@ufal.mff.cuni.cz

Heinrich Heine University Düsseldorf, Charles University

UniDive

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by
the European Union