Verbal Multiword Expressions in Treebanks

Veronika Vincze and István Nagy T.

University of Szeged, Hungary
Department of Informatics
Introduction

• Verbal multiword expressions in treebanks
• 4 languages:
  – English (PENN Treebank)
  – German (TIGER Treebank)
  – French (French Treebank)
  – Hungarian (Szeged Treebank)
• 2 types:
  – Verb-particle constructions (VPCs)
  – Light verb constructions (LVCs)
Comparisons

LVCs

- Annotated in 3 languages: German, Hungarian and French
- Level of annotation:
  - French: morphological level
  - Hungarian & German: syntactic level
- Coverage:
  - Hungarian & French: full coverage
  - German: restricted definition of LVC
Comparisons

VPCs

• Annotated in 3 languages: German, Hungarian and English
• Level of annotation:
  • morphological level
  • syntactic level
• Coverage:
  • English: full coverage
  • Hungarian & German: only if separated from the verb