

Subword Relations

Superword Features

WG1, WG2

Daniel Zeman

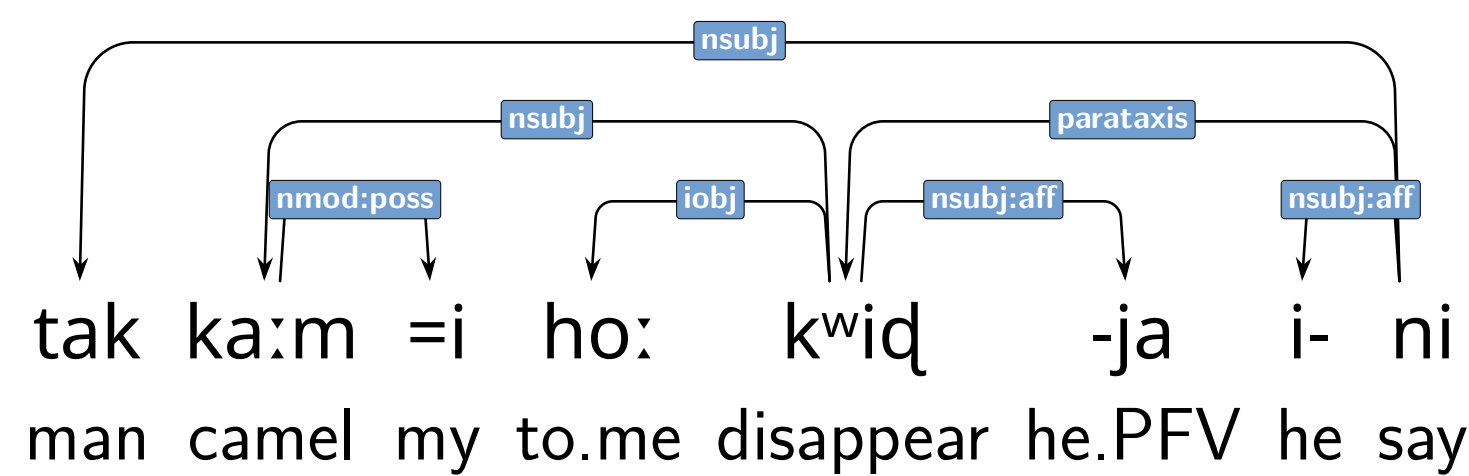
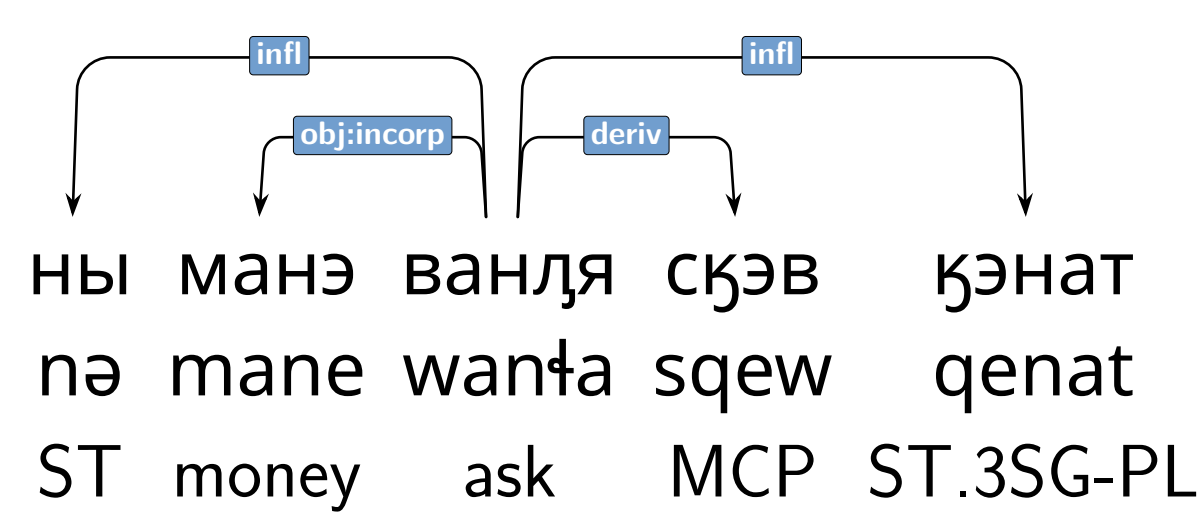
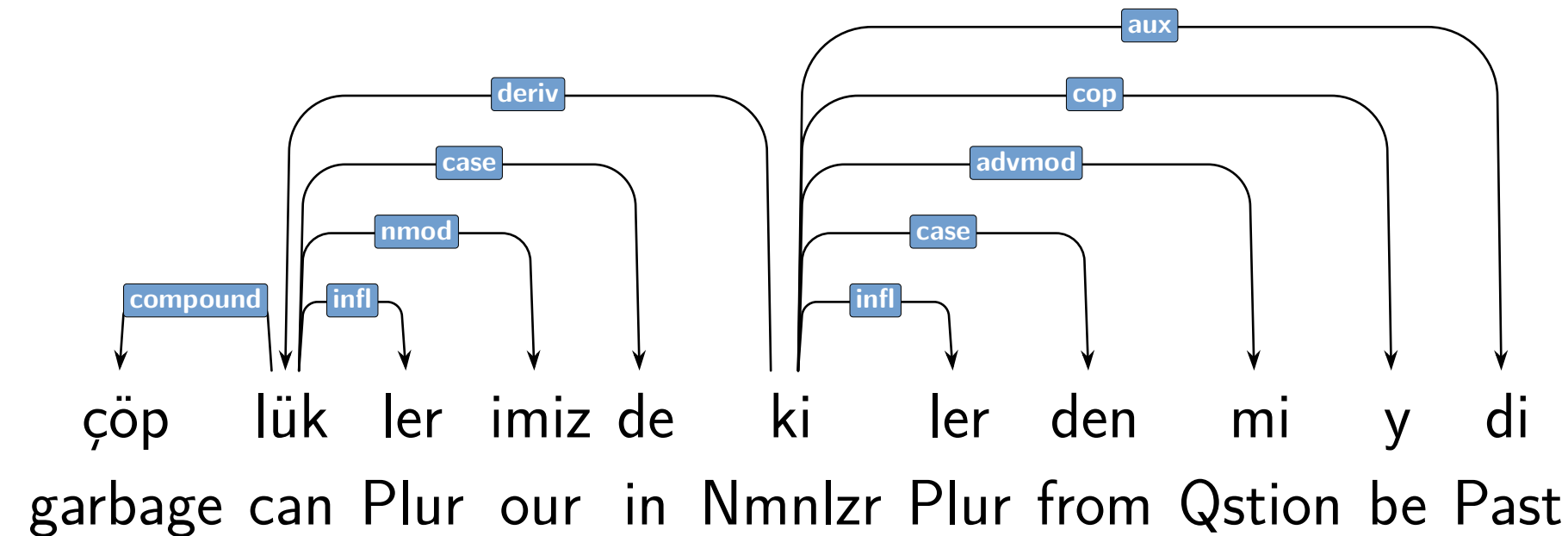
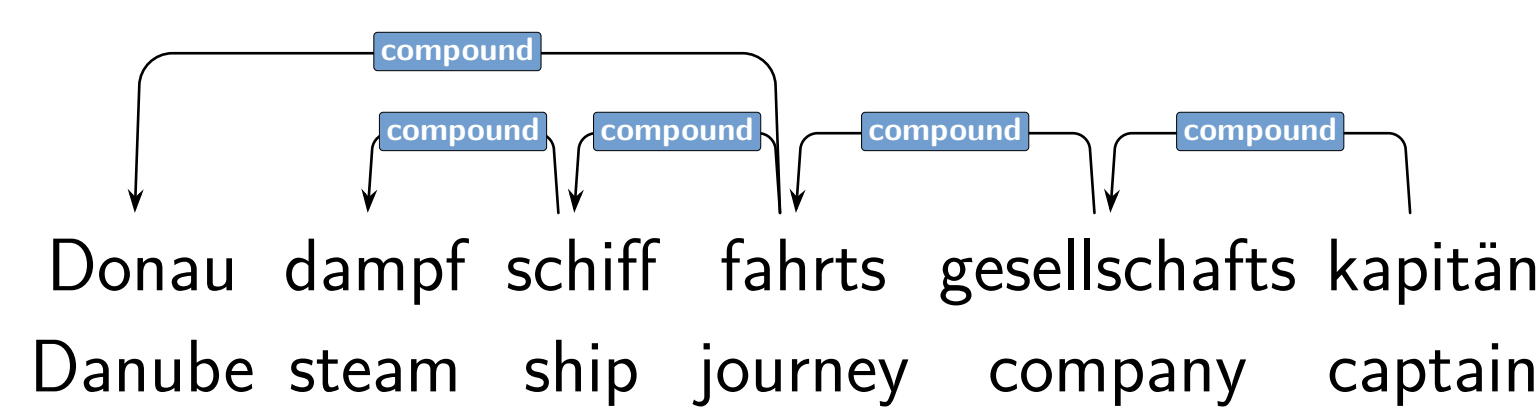
zeman@ufal.mff.cuni.cz

Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Prague, Czech Republic



Subword Relations

- German compounds are represented by one tree node in UD. Parallel English compounds are typically written as multiple words ⇒ multiple nodes connected by compound relations. Moreover, other annotation may pertain just to one part of a compound: We may want to annotate the MWE *Rolle spielen* “to play a role” in the compound *Hauptrolle spielen* “to play the main role”.
- Turkish words may combine several derivational and inflectional steps. Traditional analysis would break them up to *inflection groups* but in UD they are mostly kept together and the internal structure is not visible.
- Chukchi transitive verbs may incorporate their objects and switch to intransitive inflection.
- Fieldworkers may prefer morpheme-based analysis when documenting a language; a UD example is the treebank of Beja.



```
# text = Er spielt die Hauptrolle im Haus.
# text_en = He plays the main role in the house.
ID FORM LEM UPOS XPOS FEATS H DEPREL
1 Er er PRON - Case=Nom|PronType=Prs 2 nsubj
2 spielt spielen VERB - Mood=Ind|VerbForm=Fin 0 root
3 die der DET - Case=Acc|PronType=Art 4 det
4-6 Hauptrolle - - - - -
4 Hauptrolle Hauptrolle NOUN - Case=Acc|Number=Sing 2 obj
5 haupt haupt ADJ - Degree=Pos 6 amod
6 Rolle Rolle NOUN - Case=Acc|Number=Sing 4 wroot
7-8 im - - - - -
7 in in ADP - - - 9 case
8 dem der DET - Case=Dat|PronType=Art 9 det
9 Haus Haus NOUN - Case=Dat|Number=Sing 2 obl
10 . . PUNCT - - - 2 punct
```

Superword Features

```
# text = Er wird in München einsteigen.
# text_en = He will get on in Munich.
ID FORM LEM UPOS FEATS H DEPREL
1 Er er PRON Case=Nom|Person=3|PronType=Prs 5 nsubj
2,5 wird einsteigen einsteigen VERB Mood=Ind|Tense=Fin|VerbForm=Fin - -
2 wird werden AUX Mood=Ind|Tense=Pres|VerbForm=Fin 5 aux
3 in in ADP - - - 4 case
4 München München PRON Case=Dat|Gender=Neut|Number=Sing 5 obl
5 einsteigen einsteigen VERB VerbForm=Inf 0 root
6 . . PUNCT - - - 5 punct
```

MISC:

```
MWspan[1]=2,5 |
MWLemma[1]=einsteigen |
MWUPOS[1]=VERB |
MWMood[1]=Ind |
MWTense[1]=Fin |
MWVerbForm[1]=Fin
```

```
# text = Sie steigt in Frankfurt aus.
# text_en = She gets off in Frankfurt.
```

```
ID FORM LEM UPOS FEATS H DEPREL
1 Sie sie PRON Case=Nom|Person=3|PronType=Prs 2 nsubj
2,5 steigt aus aussteigen VERB Mood=Ind|Tense=Pres|VerbForm=Fin - -
2 steigt steigen VERB Mood=Ind|Tense=Pres|VerbForm=Fin 0 root
3 in in ADP - - - 4 case
4 Frankfurt Frankfurt PRON Case=Dat|Gender=Neut|Number=Sing 2 obl
5 aus aus ADP - - - 2 compound:prt
6 . . PUNCT - - - 2 punct
```

MISC:

```
MWspan[1]=2,5 |
MWLemma[1]=aussteigen |
MWUPOS[1]=VERB |
MWMood[1]=Ind |
MWTense[1]=Pres |
MWVerbForm[1]=Fin
```

```
# text = Hat man sich gut erholt?
# text_en = Did you recover well?
```

```
ID FORM LEM UPOS FEATS H DEPREL
1,3,5 hat sich sich erholen VERB Aspect=Perf|Mood=Ind|Tense=Pres|VerbForm=Fin - -
1 hat haben AUX Mood=Ind|Tense=Pres|VerbForm=Fin 5 aux
2 man man PRON PronType=Ind 5 nsubj
3,5 sich erholt sich erholen VERB VerbForm=Part - -
3 sich sich PRON PronType=Prs|Reflex=Yes 5 expl:prt
4 gut gut ADJ Degree=Pos 5 advmod
5 erholt erholen VERB VerbForm=Part 0 root
6 ? ? PUNCT - - - 5 punct
```

MISC:

```
MWspan[1]=1,3,5 | MWLemma[1]=sich erholen | MWUPOS[1]=VERB |
MWAspect[1]=Perf | MWMood[1]=Ind | MWTense[1]=Pres | MWVerbForm[1]=Fin |
MWspan[2]=3,5 | MWLemma[2]=sich erholen | MWUPOS[2]=VERB |
MWVerbForm[2]=Part |
```

