Natalia Klyueva and Daniel Zeman
Institute of Formal and Applied Linguistics,
Faculty of Mathematics and Physics,
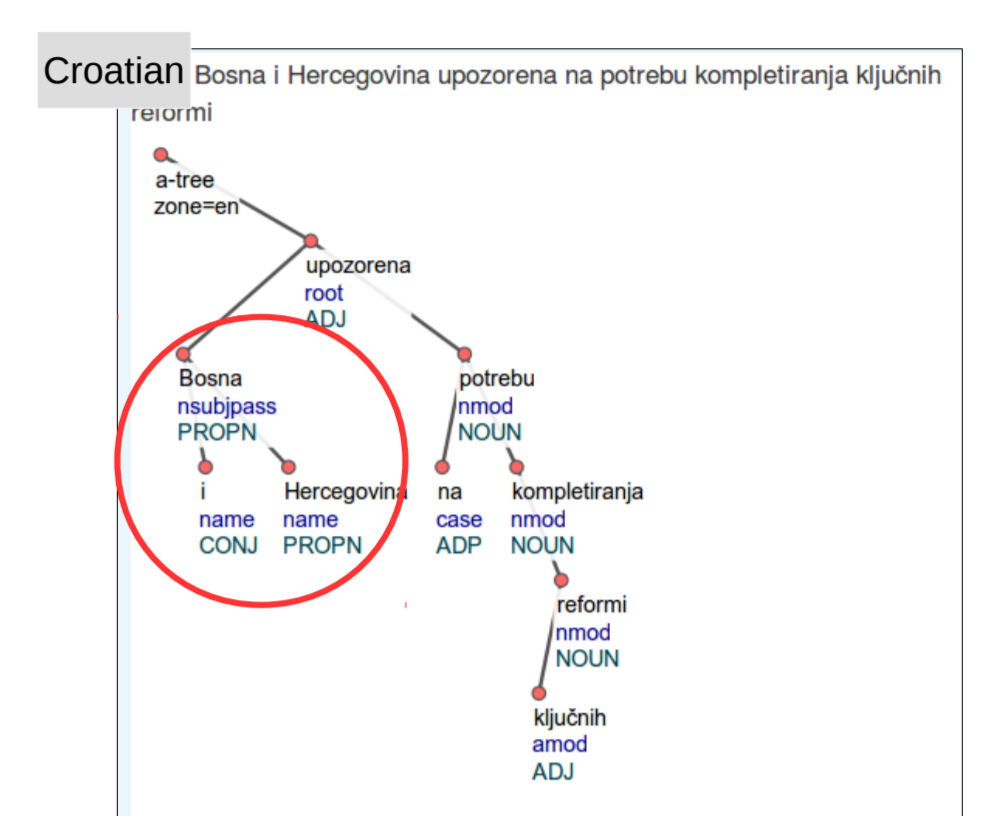Charles University, Czech Republic
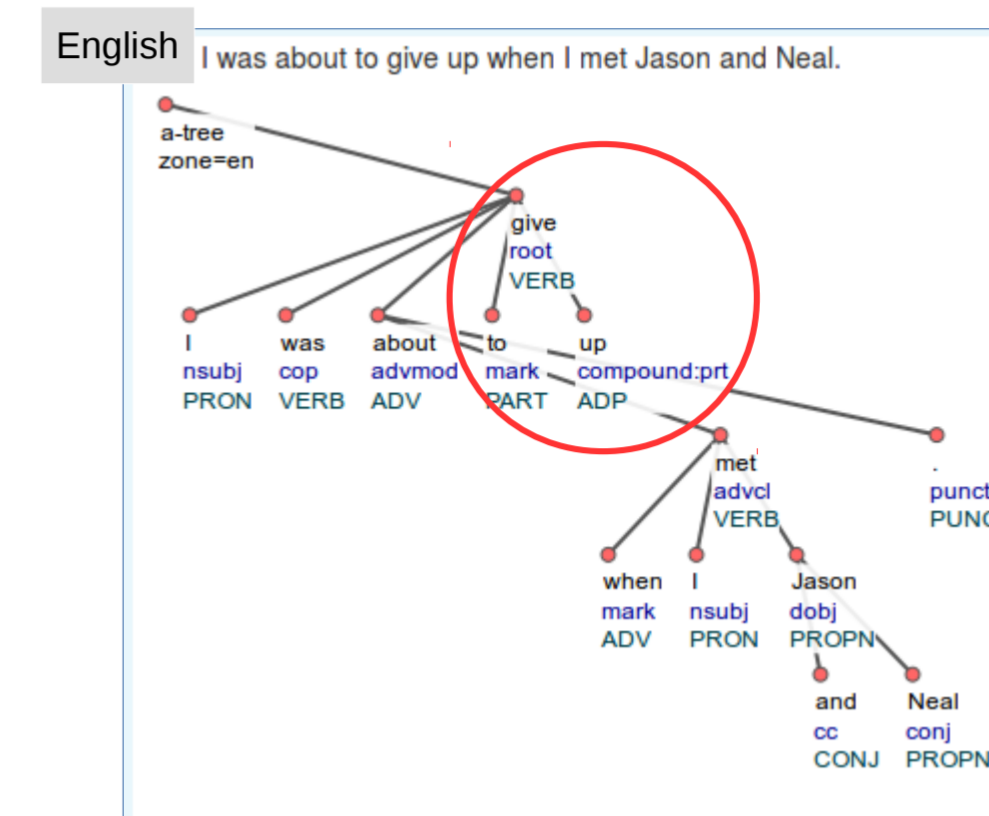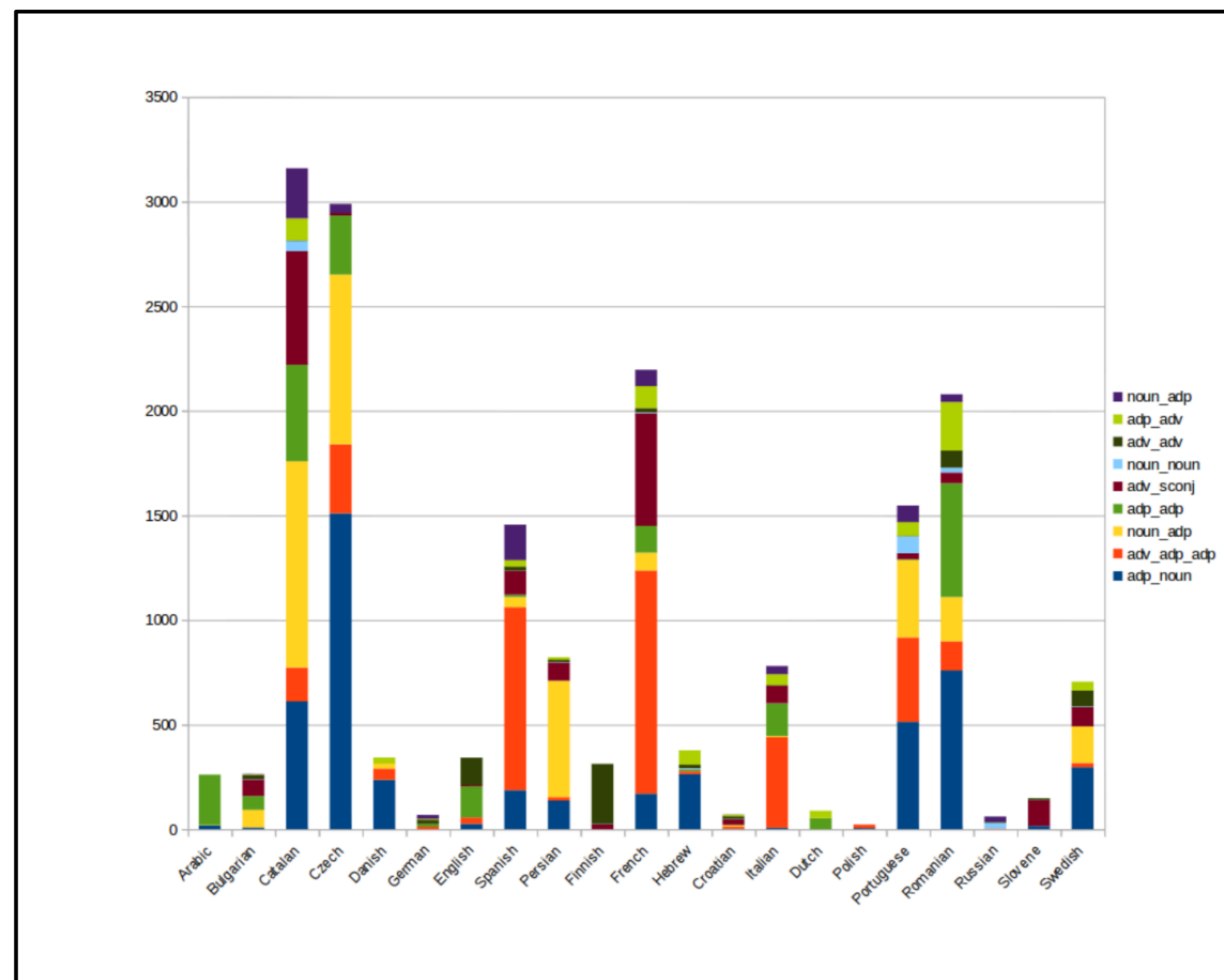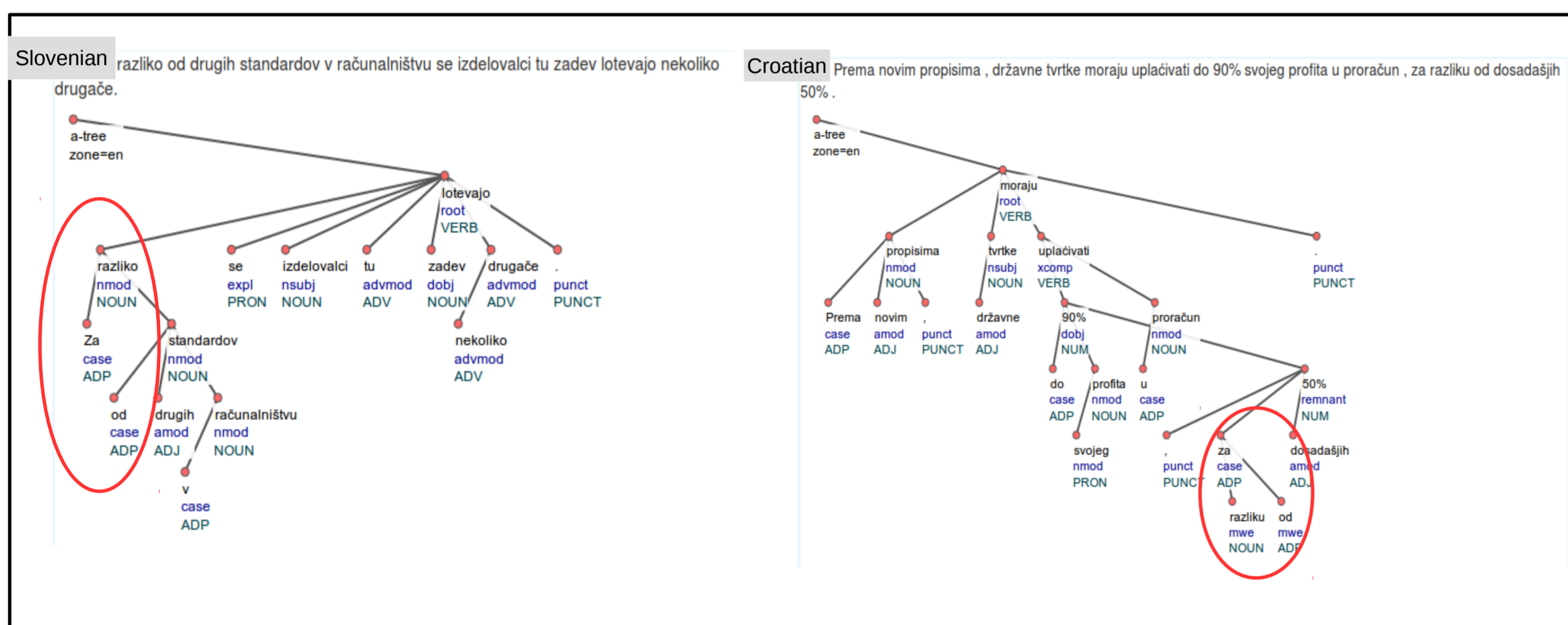{klyueva|zeman}@ufal.mff.cuni.cz

MWEs in Universal Dependencies

- morphologically and syntactically annotated treebanks
- 40 languages
- Stanford dependencies
- Google universal part-of-speech
- Interset interlingua
- Dependency relations for MWEs (until UD 1.3):
  - mwe
  - compound
  - name
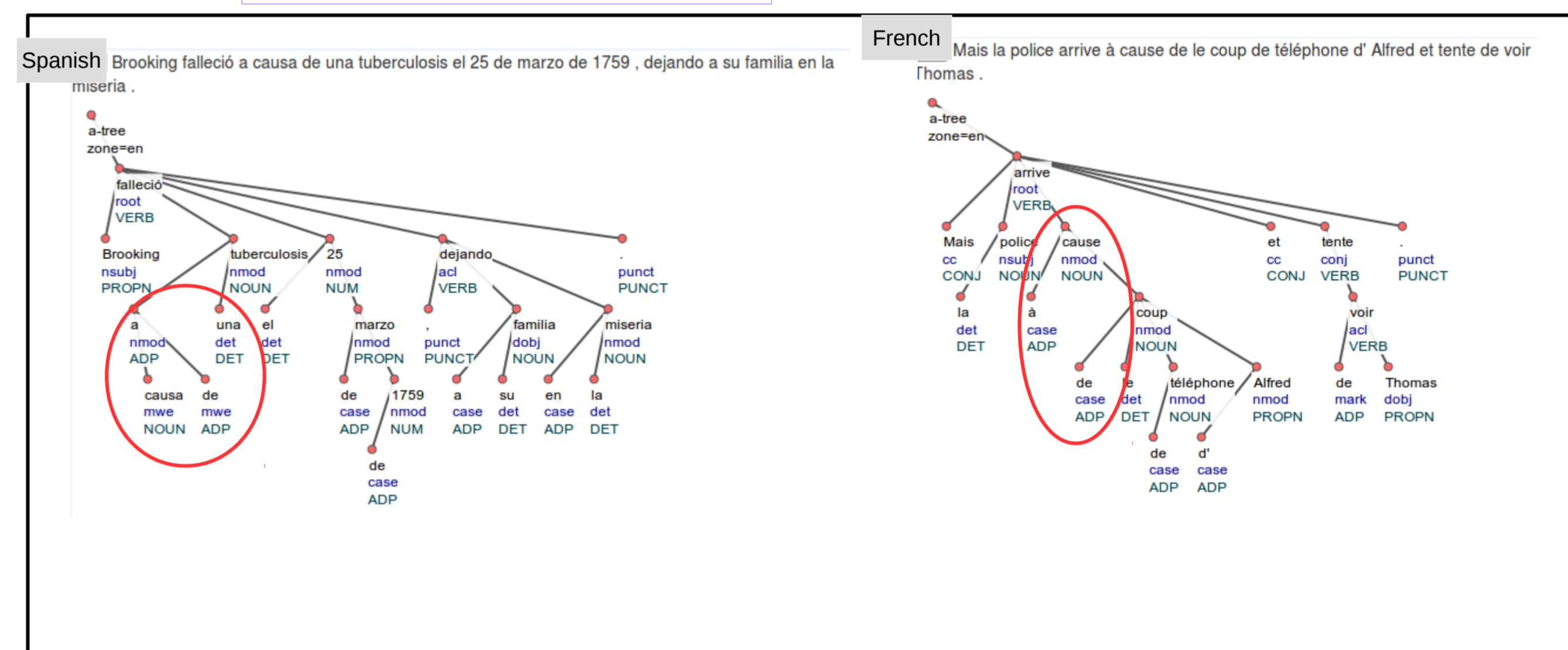- Changes under consideration:
  - **fixed**
  - **flat**

English — I was about to give up when I met Jason and Neal.

Croatian — Bosna i Hercegovina upozorena na potrebu kompletiranja ključnih

mwe → fixed

Inconsistency in Slavic: lit. 'in difference from' - unlike

Slovenian — razliko od drugih standardov v računalništvu se izdelovalci tu zadev lotevajo nekoliko drugače.

Croatian — Prema novim propisima , državne tvrtke moraju uplaćivati do 90% svojeg profita u proračun , za razliku od dosadašjih 50% .

Inconsistency in Romance: 'because of'

Spanish — Brooking falleció a causa de una tuberculosis el 25 de marzo de 1759 , dejando a su familia en la miseria .

French — Mais la police arrive à cause de le coup de téléphone d' Alfred et tente de voir l'homas .
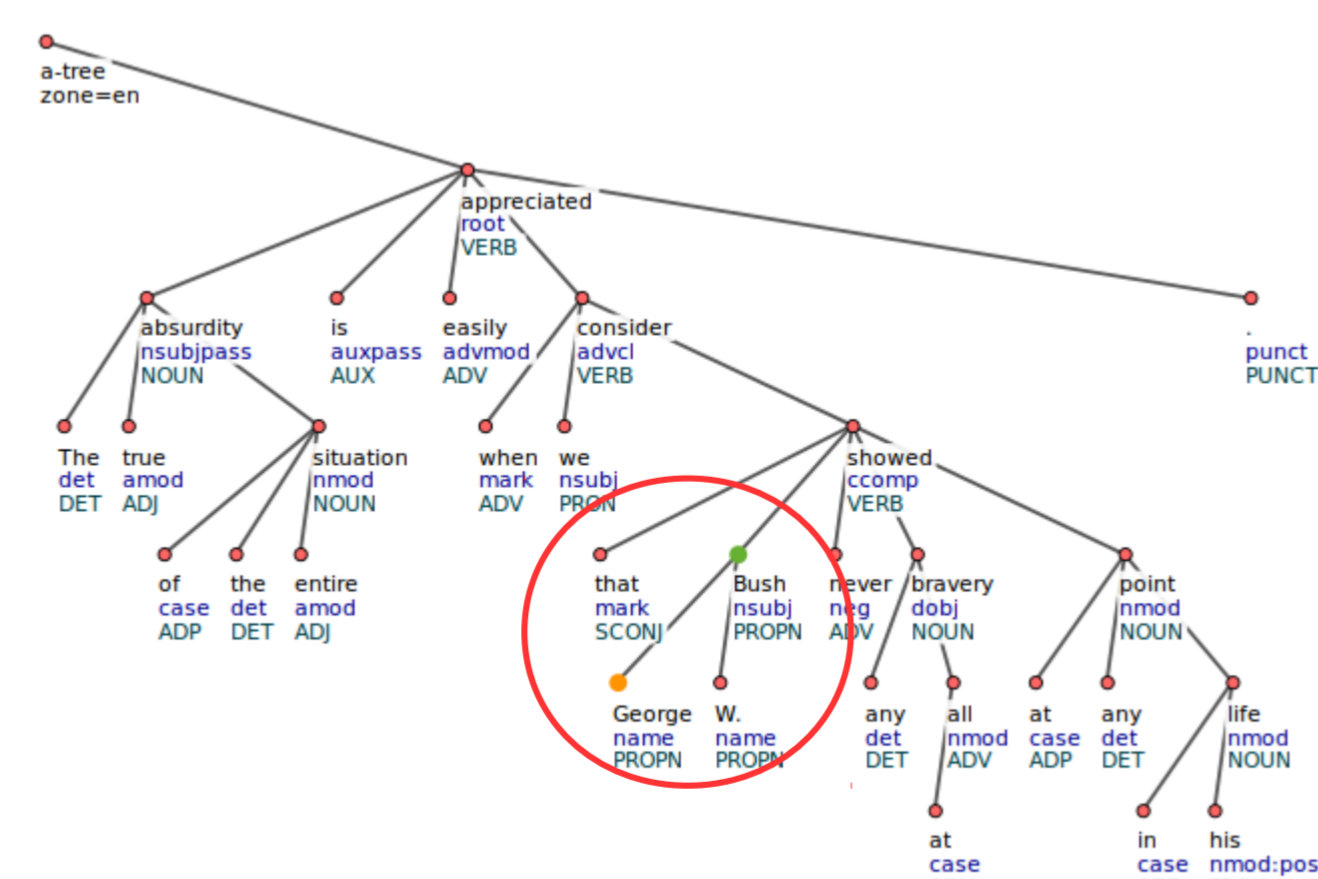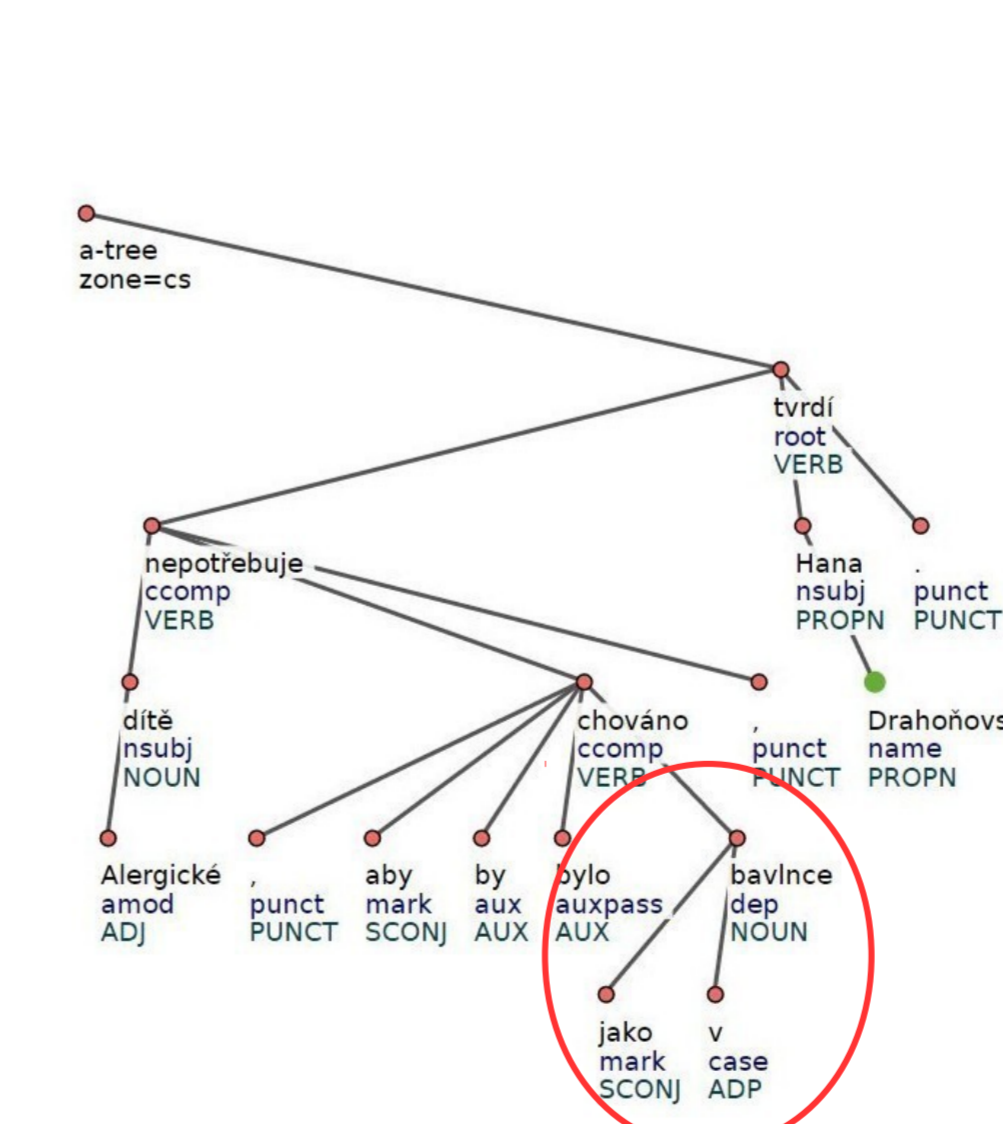
name vs. compound vs. flat

- compound:
  - not only noun phrases
  - the parent is evident
- flat:
  - When one is not sure which is the head (*president Barack Obama*)

Idioms in UD?

UD does not provide any means for special treatment of idioms:

However, the layer on top of UD will be highly welcomed

Alergické dítě nepotřebuje, aby bylo chováno jako v bavlnce, tvrdí Hana Drahoňovská.

(Czech idiom 'jako v bavlnce': *like in the silk*)

Universal Dependencies:
http://universaldependencies.org/
Search and visualization:
http://lindat.mff.cuni.cz/services/kontext/
https://lindat.mff.cuni.cz/services/pmltq/

LINDAT CLARIN

ÚFAL
ÚSTAV FORMÁLNÍ
A APLIKOVANÉ LINGVISTIKY