

# The Word

Martin Haspelmath (2023) Defining the Word

# Survey: words in languages and UD treebanks

- Theoretical definition vs. practice in corpora (here: in UD)
- Asked people to use concepts as defined by Martin in his paper
  - Sometimes difficult because people know other definitions
  - Some difficulties with details when applying Martin's definition (**main topic for this session**)
- Two rounds (refined questions)
  - 41 languages (44 responses) in first round
  - 35 languages (40 responses) in second round
  - 47 languages total
- Now analyzing in online meetings

# Root category vs. derivation

- Root category (object-property-action) is important for affix/clitic distinction
- It is not equivalent to NOUN-ADJ-VERB because of nominalizations etc.
  - [en] *protect* (VERB) and *protect-ion* (NOUN): action root in both cases
- However, the direction of derivation is not always clear
  - [cs] *plyn* ‘gas’ (NOUN / object?)
    - [cs] *plyn-árn-a* ‘gasworks, gas company’
    - [cs] *plyn-ař* ‘gasman’
  - [cs] *plyn-out* ‘flow’ (VERB / action)
    - [cs] *plyn-ul-ý* ‘fluent’ (ADJ / property)
    - [cs] *plyn-ul-e* ‘fluently’ (ADV / property)

# Phonological changes

- Is it still “the same morph”, or a different one (with the same meaning)?
  - [de] *Baum* ‘tree’
  - [de] *Bäum-e* ‘trees’
  - [es] *dorm-ir* ‘to sleep’
  - [es] *duerm-o* ‘I sleep’
- Consequences if not the same:
  - Some affixes go with one root, other affixes with the other
  - One of the roots may require affixes, the other may not
  - It probably does not change wordhood of either the affixes or the root+affixes

# Clitics vs. affixes: define “combining with a root”

- *Clitics combine with roots of different classes.*
- *Affix must occur on a root; cannot occur on roots of different root classes.*
  
- Closest root? To the left or to the right?
- Adjacent to the root? Not necessarily!
  - Other affixes can occur between an affix and its root.
    - But not other roots or clitics.
  - Can other clitics and roots occur between a clitic and “its” root?
  - **How do we know which root the bound morph belongs to?**
  - Is it the root that is needed in a free form in which the bound morph occurs? (Would allow other roots in between.)

# Contractions

- Adposition + DET / PRON (Indo-European languages)
  - [de] *zum* = *zu* + *dem* ‘to the’
    - Could we say that *m* is just a phonological variant of *dem*, i.e., *zum* is still two morphs (clitics)?
  - [fr] *au* = *à* *le* ‘to the’
  - [pt] *à* = *a* *a* ‘to the’
    - Even if German above is two morphs, we can hardly delimit two morphs here.
- Auxiliary + negation (English)
  - [en] *don't*, *can't*, *cannot*
    - Should we split *cannot* in corpora? (It is done in GUM.)

# Compounds

- In many languages, linking morphs are frequent (but the compound is considered one word => not split)
  - [de] *Liebe-s-brief* 'love letter'
  - [pl] *Biał-y-stok* (city name) lit. 'white slope' ... *biał-y* 'white' needs a regular adjectival suffix
  - [pl] *polsk-o-niemieck-ich* 'Polish-German'; *Kiribati-polsk-ich* 'Kiribati-Polish'
- Compounds are the biggest source of divergence of corpora from the definition of word by Martin

# Roots or not?

- Pronouns: full forms vs. clitics
  - Do the full forms count as object roots?
  - (As clitics, they are words, too; but we may need them as roots when recognizing other words.)
  - Adjectival / adverbial pro-forms: property roots?
- Cardinal numerals: property roots?



# Auxiliary and modal verbs

- Not contentful (they do not denote an action)
  - But they may consist of multiple morphs (“root” + inflectional affixes)
    - Clitic is single morph
  - Some of them may act as main verbs (in different context)
- 
- [pl] *będz-ie rob-ił* ‘he will do’ (*będzie* = be.3Sg.Fut)
  - [cs] *bud-e děl-at* ‘he will do’ (*bude* = be.3Sg.Fut)
- 
- We treat them as exceptional verbs (action root + affix)
  - Similar: *kick the bucket* ... *bucket* lost its original meaning, still a root?

# Slavic negative prefixes / clitics

- [bg] **he** *moga* / **ne** *moga* ‘I cannot’
- [cs] **nemohu** ‘I cannot’
- Also for adjectives (and other parts of speech):
  - [cs] **ne***byl velký* ‘he was not big’
  - [cs] *byl* **ne***velký* ‘he was not big’ (lit. he was unbig)
- Not an affix: combines with roots of different classes
- Not a clitic: may go between root and its prefix:
  - [cs] *nej*-**ne**-*pří-jem-n-ějš-í* ‘most unpleasant’

# Reflexive morphs

- Clitics in west and south Slavic (*se, si, so, sa, się, ce, cu*)
  - It does not have to occur “on” the root.
- Suffixes in east Slavic (*-ся/-сь*)
  - It always immediately follows the verb (the root + possibly inflectional suffix)
- 
- Italian: also always on the verb?
- Spanish: other clitics or auxiliaries in between
  - [es] *se la vi comiendo (vi a usted comiendo la cena para usted)* ‘I saw you eating it’

# Derivational prefixes vs. verbal particles

- [de] **an**kommen ‘arrive’ ... separable verb “prefix” ... clitic?
  - *Sie kommt sofort **an***
  - *Sie wird sofort **an**kommen*
- [cs] *měn-it* ‘to change’ – *měn-a* ‘currency’; *z-měn-it* ‘to change’ – *z-měn-a* ‘change’, *s-měn-it* ‘trade’, *od-měn-it* ‘reward’, *ob-měn-it* ‘alter’, *vy-měn-it* ‘exchange’ (**vý-měn-a!**), *za-měn-it* ‘mix up’ (**zá-měn-a!**), *pro-měn-it* ‘transform’, *pře-měn-it* ‘convert’, *roz-měn-it* ‘break (about money)’ (no \**rozměna* here)
  - *od-měň-ov-at*, *ob-měň-ov-at*, *vy-měň-ov-at*, ...
- The derivational prefixes would have to be clitics if the root *měn/měň* is not always the same root from the same (action) category.