Morph telco 2022-12-21, 13:00 CET

**Link:** [**https://meet.google.com/vvf-qdnd-xrq**](https://meet.google.com/vvf-qdnd-xrq) **[check here for link updates if it doesn’t work]**

**Latest Definitions:**  <https://github.com/ontolex/morph/blob/master/draft.md>

**Nexus:** https://nexuslinguarum.eu/the-action/join-us

**Participants [please add yourself]:**

Christian Chiarcos (CC)

Max Ionov (MI, excused)

Katerina Gkirtzou (KG)

Fahad Khan (FK) (excused: 20 min late)

Khadija Ait ElFqih (KAE) – excused

Matteo Pellegrini (MP)

Ciprian-Octavian Truică (CT)

Penny Labropoulou (PL)

Elena Simona Apostol (ESA)

Sina Ahmadi (SA)

Elena Benzoni (EB)

Petra Steiner (PS)

Theodorus Fransen (TF)

Thierry Declerck (DFKI)

Ranka Stanković

Gilles Sérasset (GS)

Mike Rosner

# 0. Module draft (4.17)



No changes

# 0. Organizational

Last two calls skipped, hence switching weeks with OntoLex-FrAC.

Next call, Jan 11

# 1. Publications

## MWE Chapter

* Initiated in FrAC
* Initial abstract: <https://docs.google.com/document/d/1HLTsYScMiZE4Nb6b7mozN-SoKAIWtisdiKLa0OoXPuU/edit?usp=sharing>
* Draft for chapter: <https://www.overleaf.com/8285444258rpfnbwgwbrdp>
  + Template applied (<https://www.overleaf.com/latex/templates/langsci-skeleton-for-edited-volumes-2022-06/vvjbpxnrbnkk>)
  + Based on globalex paper
  + First version of intro (need input on MWE definition and typology)
  + First version of ontolex description
  + Not sync’ed with initial abstract
  + Second half of the paper is in very rough shape
* Slack channel: #frac-mwe-chapter
* authors: CC KG MI BK FK CT EA (TD can also contribute, we have been doing some work on MWE in Elexis)
  + Matteo: Latin not a great fit, possibly Elena Benzoni, tbc. yet (not sure)
    - Not by Jan 15
  + FK: maybe 1,5 weeks in January
* **Tbc**: Jan 15, 2023
  + Very little time, ask for extension

## eLex 2023

<https://elex.link/elex2023/>

Submission deadline: 31.01.2023

* Abstract only

Dates: 27.06-29.06

Topic tbd:

* Overview of the model
  + LREC was just dynamic part
  + Focus on changes since the last presentation
    - Inflection type, grammatical meaning changed
    - Descriptive part: morph
* Use case: Semitic? [if we have a consensus on that]
  + We need a native speaker
  + Mike willing to help but not native
  + **TODO@CC**: ask Khadija (fallback solution: ask Ilan)
  + FK: maybe go ahead with modelling and then ask whether ok
  + Not realistic until January
* Another use case (different dataset?)
  + Theodorus Fransen: medieval Irish with LLOD, collab with LiLa, also involving phonemic transcriptions, background FST morphology
    - Timeline a bit tight (unless more conceptual)
    - FK: copyright? LLOD possible?
      * TF: full-form dictionary for a corpus: copyright cleared
      * TF: FST grammar: no copyright issues
  + Penny: featured example for trad. Dict + computational resource
    - Possibly a greek resource
    - Tbc: availability of resources
* Abstract only, paper later (full paper deadline 15.03.2023)

## LDK Workshop

* W3C language technology CG day at LDK-2023, probably Sep 12
* Designated session for OntoLex-Morph: need moderator
  + TF might support
* NB: emails to John via private email (gmail)

# 2. Handling of Semitic languages [postponed]

**TO-BE-DONE**: figure out how to include missing categories to LexInfo

Khadija: data prepared for Arabic (cf. last time):

* [Morph Module semitic.odt](https://drive.google.com/file/d/1NST-qdkxAw7am6F3vGAap5o4eQv0ulV1/view?usp=sharing)

Necessary features:

* lexinfo:POS extensions (solved? See last call minutes for procedure)
* Modelling diacritics in Arabic (cf. call minutes last time)
  + Also cf. Umlaut in German and vowel harmony in Turkish for similar challenges
    - Recommendation (< GS): NFD normalization in morph:Replacement
    - **TBC:** are we ok with modelling roots as Morphs (i.e., LexicalEntries)?
  + If not modelled as morphs, then they could be modelled as rules (replacements)
    - **DONE@CC**: model updated
      * morph:grammaticalMeaning and morph:baseConstraint as properties of morph:Rule, the grammatical meaning is the change in meaning or morphology of the word (root)
    - **TO-BE-DONE@FK:** model a Hindi/Urdu example, e.g., <https://en.wiktionary.org/wiki/%E0%A4%96%E0%A5%81%E0%A4%B2%E0%A4%A8%E0%A4%BE#Hindi>
    - **TO-BE-DONE@Khadija:** Modelling examples for Arabic entries
* Not discussed yet: Circumfix
  + Morph:CircumfixParadigm
  + Prefix+suffix combination

## 3. Interlinking Latin Language Data in the Linguistic Linked Open Data

* Thierry STSM@CIRCSE/LiLa Sep-Oct 2022
* Focus on morphology, esp. with Marco Passarotti and Matteo
* *Issue*: linking with the corpus is problematic if corpus words are not disambiguated (semantically and grammatically)
  + Introduced underspecified forms (multiple lexinfo values for the same lexinfo property, e.g. lexinfo:case)
  + Suggestion: class morph:UnderspecifiedMorph as a subclass of ontolex:Form (?)
    - Matteo: yes, maybe better morph:UnderspecifiedForm?
* *Issue:* frequency counts for fully and partially disambiguated morphs
  + Suggestion: use morph:UnderspecifiedMorph, as well
  + Should be a FrAC Observable (“underspecified observable”)
* Tbc: do we need fully specified lexical entries in the model?
  + If not, this class would be nice to have
* Blog post
  + Draft restricted, so far, **TODO@TD**: provide a public link
* Gilles: rather classical problem, any examples with multiple underspecified dimensions
  + Same problem for Greek, Serbian, German (also Latin), e.g., number+case combinations
    - Post-call: Russian *кушанье* “food”, gen.sg.=nom.pl=acc.pl *кушанья* (gen.pl *кушаний*, nom.sg.=akk.sg. *кушанье*)
  + Matteo: link forms to bundles of features (=> GrammaticalMeaning[s], optionally with lexinfo feats)
* Christian:
  + possibly forms can be underspecified in OntoLex (cf. one form for multiple lexical entries [not forbidden], forms in lexical resources that don’t disambiguate morphological features]); then, no new class needed
  + Possibly to be modelled with frac:Observable, cf. corpus queries (not fully discussed yet)
    - Gilles: not limited to FrAC use cases, but general feature of lexical resources; also not limited to morphology
    - Matteo: problem in the definition of ontolex:Form? [just string or string+features? Both definitions occur]
  + Question: do we need a specific class?
    - Thierry: was introduced for readability
    - Christian: We can infer underspecification from having multiple grammatical meanings
    - Matteo: in Lila helpful, but can be user class, useful feature is to state that something is unambiguous
    - Christian: model by property morph:soleGrammaticalMeaning (must be unique per form)
    - Matteo: or, specify at observable? Doesn’t necessarily require an extension of morph
  + Telco link: <https://meet.google.com/whn-efaq-cha>
  + **TODO@MP**: small data example

## 4. Definitions (old) [postponed]

See <https://github.com/ontolex/morph/blob/master/draft.md> and <https://github.com/ontolex/morph/issues> .

Old minutes:

* PS: how to represent order of morphemes (more than one affix)
  + MI: inflectionType next (only for inflection)
  + CC: for derivation, we can decompose it into a tree structure by nesting lexical entries in derivation then order is indirectly encoded
  + CC: we also have an aggregate relation between form and morph, that can be an rdf:Seq [**TODO**: check that this is explicitly mentioned in draft]
    - Proposal@CC (after telco): that should be the preferred answer
* PS: how to represent derivation trees? ((A B) C)
  + CC: has been one of the original requirements (https://www.w3.org/community/ontolex/wiki/Morphology#Representation\_needs, "N7: Multiple segmentation strategies")
  + Current solution: use decomp (note we have decomp:subterm already in diagram); if all intermediate steps are valid lexemes, this can also be done via multiple WordFormationRelations (Rules)
* Otherwise postponed, leaving more time for 3.
* CC (after telco): started to check and update requirements in wiki: <https://www.w3.org/community/ontolex/wiki/Morphology#Representation_needs>

## 5. AoB ?