Morph telco 2022-04-20, 13:00 CET

**Link:** [**https://meet.google.com/nsj-tbcy-yop**](https://meet.google.com/nsj-tbcy-yop)

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

**Latest Paper (submitted to LDL-2022):** <https://www.overleaf.com/4868363189kczjzdndgxwc> (folder submission/)

**Participants [please add yourself]:**

Christian Chiarcos (CC) [christian.chiarcos@gmail.com](mailto:christian.chiarcos@gmail.com)

Max Ionov (MI)

Katerina Gkirtzou (KG)

Fahad Khan (FK)

Matteo Pellegrini (MP)

Ciprian-Octavian Truică (CT)

Penny Labropoulou (PL)

Elena Simona Apostol (ESA)

Agenda (please add, *but* do not edit table of contents directly, but add sections below and then update here):

[**0. Module draft**](#_1kaaos1w4ub1) **2**

[**1. Publication plans**](#_5q4h03dxajtd) **3**

[2. definition consolidation](#_n0e2ll1nl5iu) **3**

[**2.1 InflectionType**](#_letz9qce1sim) **4**

[**2.2 morph(eme) order**](#_ki1gty3ur827) **4**

[**3 open problems/other data**](#_jwteu367w2rp) **5**

[3.1 Samples to be modelled (all)](#_gvwqhfae33jq) 5

[3.2 inflection tables (Fahad, others?)](#_rzzsrcthlbm3) 5

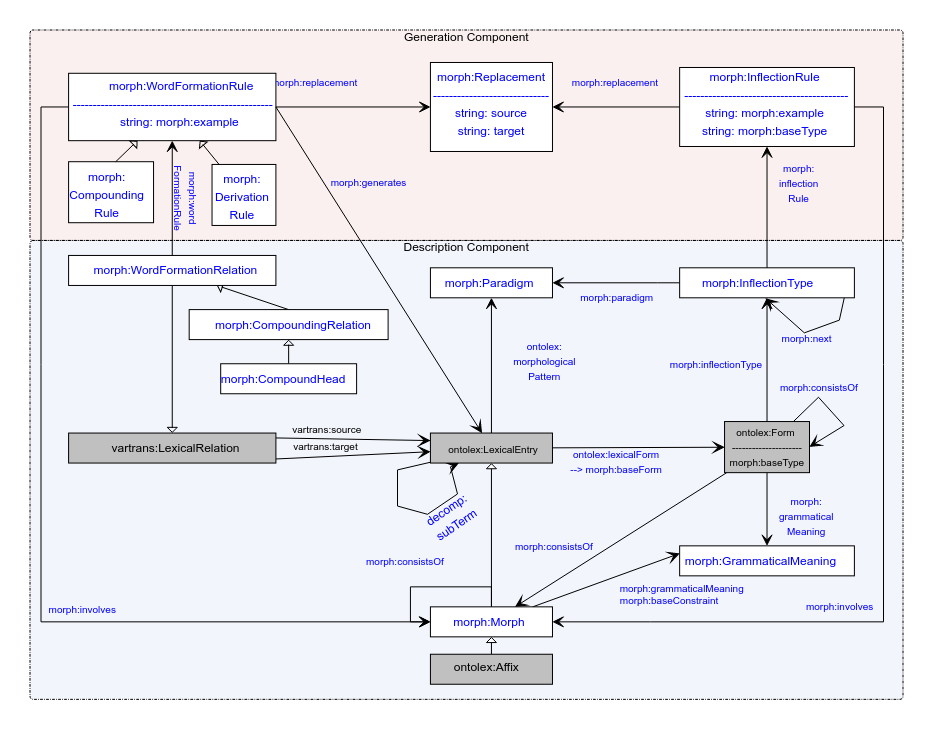
[3.3 semitic consonantal roots (unassigned)](#_w3ikvzzgmbyd) 6

[**4. AOB**](#_4u89668ejc7q) **6**

# 

# 0. Module draft

**draft 4.14**



**implemented updates:**

* morph:inflects => morph:involves (for parallel with word formation)
* morph:paradigm (between lexical entry and paradigm) => ontolex:morphologicalPattern (we forgot about that ;)
* vartrans:LexicoSemanticRelation => vartrans:LexicalRelation (was an error)
* CompoundRelation => CompoundingRelation (in analogy to derivation)
* CompoundRule => CompoundingRule (tentative consensus)
* CompoundHead stays
* decomp:subTerm added (TODO: add at definitions, too)
* text: CC restructured “other data” section below

**Model draft 4.15 updates (to be discussed):**

* inflection type to be discussed
* consistsOf?

# 1. Publication plans

* COLING?
  + may 17
  + could be an altewrnative venue for the LDL submission (if rejeted or revoked) or a novel paper (if we have novel content)
    - decision postponed
  + idea for novel paper: word formation in OntoLex-Lemon
    - not original content, but more like a survey and documentation of best practices?
    - can be helpful to consolidate/revise word formation part of the module
    - possible input from LiLa
  + paper@COLING:
    - FK: update of OntoLex, incl. FrAC, Morph, MModality
* LLODREAM? (Matteo, Marco) <https://easychair.org/cfp/llodream2022>
  + 500 words
  + deadline june 15th
  + ?publication
* later journal paper

# 2. definition consolidation

* internal deadline: before May 17 (suggested at last call)
* <https://github.com/ontolex/morph/blob/master/draft.md>
* you can contribute suggestions by creating issues (<https://github.com/ontolex/morph/issues>), via pull requests, or by direct editing (share your GitHub username)
* definition refinement
  + procedure:
    - open an issue
    - pull request + close the issue
* **old todos**:
  + @CC: update definition of inflects in GitHub draft
  + @Matteo: read definitions until next telco
  + @Penny: read definitions => issue
  + @CC: define morph subclasses in LexInfo rather than OntoLex-Morph, cf. <https://github.com/ontolex/lexinfo/issues/21> (needs update: not as TermElement; also add equivalence axioms (lexinfo:Prefix subclassOf [ lexinfo:termElement lexinfo:prefix ])
  + @unassigned: describe the relation between decomp and CompoundRelation
    - * morph:CompoundingRelation is a reification of decomp:subterm
      * there is an alternative reification with decomp:Component, but this is less well-suited for compound analysis, because it doesn’t relate to lexicosemantic relations.
      * the current modelling of decomp is oriented towards an analysis of synsem (semantic) roles within a compound. in morphology, we normally don’t have that, what we have, instead, are relations between lexemes and morphemes.
  + @unassigned: describe grouping of lexical (sub-) entries
    - LiLa: “flexeme”, sub-entries with different paradigms, but identical in meaning, etc.
      * + suggestion: model the grouping by lexicog, have both the overarching lexical entry and the flexemes as separate lexical entries, no vocabulary extension needed, but a usage note in the report
    - Penny: sub-entries of the same lexical entry to mark contracted and non-contracted versions of the same paradigm
      * + can be partially modelled by means of “markers”, i.e., lexinfo usage properties, instead
        + todo@Penny: tbc. whether lexinfo needs to be extended for that
  + @all: think about metadata properties for LexInfo (hypothetical/unattested form, etc.) => tentative consensus, but details to be discussed

## 2.1 InflectionType

* revise/abandon inflection type? (needs to be checked on data => discuss German FST data)
  + Max: no, this is a complete revision of generation model
  + we don’t seem to have a common understanding of what InflectionType is meant to be
    - Max: hard to explain to other people
  + to be discussed next time
* <https://link.springer.com/chapter/10.1007/978-3-030-98876-0_34>: MMoOn modelling for Greek morphology. Worth a comparison?

## 2.2 morph(eme) order

* CC: this is an old problem we always postponed, forgot to mention last time
* consistsOf
  + currently
    - ambiguous: form -> form, form -> morph, morph -> morph
    - doesn’t represent order
  + suggestion:
    - replace by aggregation (rdfs:Seg or rdfs:Bag)
    - define a form as an aggregate of morphs
    - no such relation between form -> form and morph -> morph
    - instead of morph -> morph: because morph is a lexical entry, we go from morph -> form, and then use “consistsOf” equivalent at form level
    - instead of form -> form: use form -> morph and assign the “sub-forms” as lexicalForm to the morph
  + has the original “consistsOf” ever been used by anyone before?

# 3 open problems/other data

## 3.1 Samples to be modelled (all)

* most sample data originally on GDrive (where is the link?)
  + now (also) on GitHub: <https://github.com/ontolex/morph/tree/master/data/gdrive>
    - CC: can we fully move there?
* samples @ GitHub
  + Latin (word formation<LiLa: tbc: is that covered?)
  + Sumerian (agglutination<CC: open requirement: slots)
  + Old High German (word formation<CC: open requirement: tree structures)
  + Italian (word formation<Stefania; tbc: is that covered?)
  + Italian (inflection<WHOM?; tbc: is that covered?)
  + Inuktitut (incorporation, polypersonal agreement, assimilation/allomorphy; generation/parsing<CC: open requirement
  + UniMorph (inflection<CC: todo: apply modelling), cf. <https://github.com/acoli-repo/acoli-morph/tree/main/unimorph>
  + Finnish (generation<Max: todo: to be updated): <https://github.com/ontolex/morph/blob/master/data/generation/dataset-generation-example.ttl>
  + ?FST (FOMA, Quechua): <https://github.com/ontolex/morph/tree/master/data/foma/quechua>
* external:
  + Finnish@Max (where?)
  + LEXIS (Greek Parole-Simple dict@Penny)
  + DeriNet/UDer/Universal Derivations
    - Latin@LiLa ?
    - German@Christian: <https://github.com/acoli-repo/acoli-morph> (UDer 0.5 only)
  + SFST: <https://github.com/acoli-repo/acoli-morph> (Morphisto@Christian, German; inflection only)
  + GermaNet compounds: <https://github.com/acoli-repo/acoli-morph> (German@Christian)
  + Morph@Thierry (= Italian samples?)
  + **TODO**: (please list your data, unless described in separate section)
  + **Open requirements**: IGT/ToolBox/FLeX data, inflection tables!

## 3.2 inflection tables (Fahad, others?)

* Latin (?)
* Old English (Fahad): sample data: coman/quoman example, cf. <https://en.wiktionary.org/wiki/cuman#Old_English>
  + issues with dialects (reference dialect vs. other dialects) and diachrony (phonological processes); treatment of syncopation, suppletion, fusion of different roots [=> variants?] ?
* postponed until Fahad has some progress on modelling

## 3.3 semitic consonantal roots (unassigned)

* + from the same consonant cluster, we can generate different POSes
  + cf. <https://en.wikipedia.org/wiki/K-T-B>, <https://en.wiktionary.org/wiki/%D9%83_%D8%AA_%D8%A8>)
    - this cannot (always) be modelled as inflection, as OntoLex requires (at most) one POS per lexical entry
    - note that this page describes vowelized words as “derivatives”: can we model this as derivation ? (but the process occurs in inflection, too)
      * given a real dictionary, can be easily distinguish derivation and inflection?
  + cf. Arabic example from <https://en.wikipedia.org/wiki/Dictionary_of_Modern_Written_Arabic> (from Max)
    - dictionary organized by roots, but root is not made explicit
    - **todo@unassigned**: put an example into GitHub
* discussion postponed until we have a Semitic speaker
  + - Ilan?
      * but first, check Bettina’s conversion of KDictionaries’ Hebrew dict

# 4. AOB

next call in two weeks