

Morph telco 2022-03-08, 13:00 CET

Link: <https://meet.google.com/jxf-kvsv-ibz> [one-time-link; check here for link updates if it doesn't work]

next : <https://meet.google.com/ymg-cauc-ezc>

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)

Katerina Gkirtzou (KG)

Besim Kabashi (BK)

Fahad Khan (FK)

Khadija Ait ElFqih (KAE)

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)

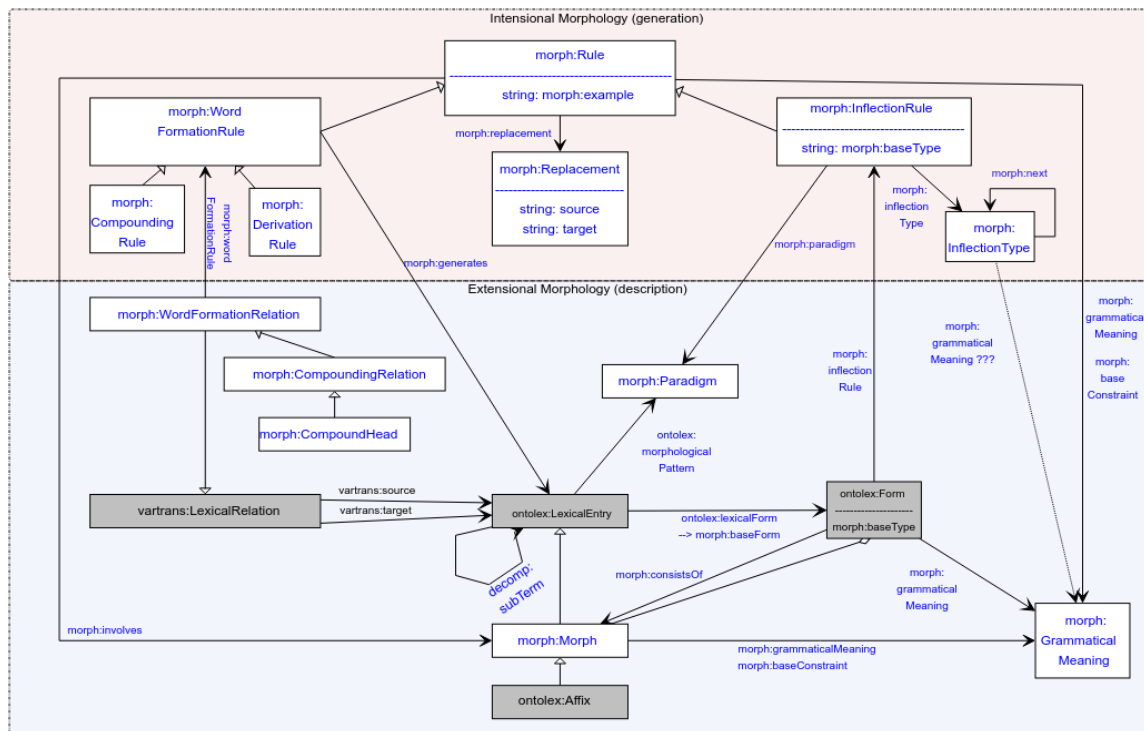
Thierry Declerck (DFKI)

Ranka Stanković

Gilles Sérasset (GS)

Mike Rosner (MR)

0. Module draft (4.17)



No changes

Organisational

What are open topics?

- Character / sound classes
- morph:baseConstraint
- Finite state terminology
- Issues with semitic?
- MP: Naming "paradigm" => morphological patterns? Inflection type? Inflection Class?
- TODO@CC: Tbc: issues on wiki page
- TODO@CC: Tbc: issues on specific data sets (github)

Suggestion:

- Can we do a requirement freeze in two weeks?
 - Sending email announcement, asking people to raise objections or to propose new data
 - Not: freezing vocabulary, just freezing requirements (and to propose new data)
- Vote:
 - Yes: 11
 - No: 0
 - Abstain: 1

- **TODO@CC**: send email about that freeze (DONE 2023-03-13)

Publications

MWE Chapter

- Submitted: <https://www.overleaf.com/8285444258rpfnbwggwbrdp>
- Slack channel: #frac-mwe-chapter

LDK

- Mike and Max on Maltese
- Possibly Ranka on football dictionary
 - Contributors?

LDK workshop

- Fahad worked on the website
- Call for participation, program needed
- In contact with sara carvalho (on terminology)
- To figure out how to do registration
- Sep 12, full-day
- Catering in doubt
- There will be parallel events
 - Not online, yet
 - proling knower (half-day) and the Frame Net (half day) and Disinformation and Toxic Content analysis (full day) will be in parallel on the 12th
 - proling morning, framenet afternoon

=> so any synsem stuff in our workshop into the morning and any metadata into the afternoon

- We need a separate call
 - Prep to be coordinated by CC and FK on slack

Character / sound classes

Max and Mike, LDK paper on morph on maltese/semitic

Problem: / (K) ([aeiou]{1,2}) (K) ([e]{1,2}) (K) /\1\3i\5t/

- Can be done without, but illegible

Gilles:

- Does it resemble two-level morphology?

- Max: basically named groups, can just be expanded, but illegible. Also problem with functionality might be digraphs that represent a single unicode character

Proposal: to add a class representing a character class (e.g. vowels, consonants, “sun consonants” — Maltese)

```
consonants morph:SoundClass ;
    rdf:label "Sun" ;
    ?:contains "d", "n", "r", "s", "t", "x" .
```

```
consonants morph:SoundClass ;
    rdf:label "Sun" ;
    ?:contains "[dnrstx]" .
```

- Data is the Maltese data discussed here before, will be partially converted
- CC: separate module? Together with signs?
- MI: not about phonology, more about representation; workaround: precompile within rule set
- GS: proposal contains two things:
 - Mechanism and terminology
- TF: what is the usecase of mimicking finite state paradigms
- CC: interoperability with dictionaries and “text book rules”
- MI: rules motivated from minimizing the inventory (rules instead of full forms). That was a requirement from the beginning
- Mike: question of what we want morph, necessary for languages where orthography mapping is not 1:1
- GS: strip out heuristics, focus on description how language changes strings
- MI: not implementing FST, just using the means (a very small part) to encode rules for generating wordforms
- GS: morphology describes transformations, so we need that
- CC: let the proposal sink in for a few weeks, if there are additional use cases that really require this *within morph*, we could implement it, the need and motivation is clear. Not fully convinced that modelling should be *restricted to morph*
- Mike: will be discussed in paper

EVERYTHING ELSE POSTPONED

morph:baseConstraint

- morph:baseConstraint — found only in one example with Inuktitut generation. In a nutshell, morph:baseConstraint can be used to provide prerequisites for a morph to be compatible. For example


```
:m1 morph:baseConstraint [ :pos "v" ] .
```

sets the requirement for the word that this morph can be added to. And morph:grammaticalMeaning holds grammatical categories for the morph itself, as before. But shouldn't it be only for a Rule, not for a Morph?

It is 2023, we need to finish the module

- MI: I think we achieved a level where we freeze everything and write it up
- Is there anything we are missing?
 - KG: LREC paper discusses the limitations of the module, we need to look at this, discuss the asap. **TODO@KG**: look into it and add this to the agenda or report it
 - MI: I think that we can still have limitations unsolved for the final module, we just need to explicitly decide
- Should we vote in Slack+mailing list since not everyone is in the call (*prediction*)
 - CC and MI should decide, then send to OntoLex chairs to approve
 - **TODO@MI,CC**: discuss
- Thoughts?
 - FK: maybe we still need more discussions, especially w.r.t. Semitic languages
 - FK: maybe we can take a couple of Wiktionary examples, model them (naively) and then show to experts (Mike, Khadija, Ilan)
 - FK: it is very important that we deal with non-European languages (non-Indo-European), so we need to make sure that we *can* model at least a good sample of languages of the world
 - MR are we going to try to handle Maltese data of [this kind](#)? If so I would be glad to help.

Handling of Semitic languages [maybe today]

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

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

-  Morph Module semitic.odt

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

AoB ?