# Developing an OntoLex Morphology Module

**Telco: Feb 3rd, 2021**

**Participants:**

Bettina

Max

Francesco

Stefania

John

Thierry

Matteo

Eleonora

Fahad

Maria Pia

Adrian

Saurav

Christian C.

Julia

John

**Agenda**

1. [**Modeling Needs**](https://docs.google.com/document/d/1z1HjQ0_WmdQwdRoktSFPGewOkzcHesdJiz0O4CFX1Q8/edit?usp=sharing)

None collected

The 4 needs from telco 20.01.2021 are also missing:

4 ways to represent morphology

1. rules with transformations(Max)
2. Inflection Tables with full inflected word form(Max)
3. set of affixes/word endings (Max)
4. Assimilation rules as additional rules on top of 1.
5. [**Test data collection**](https://drive.google.com/drive/folders/1hNOj5qm34J7uGJrsudI4eDuQAMCFFfzv?usp=sharing)

So far three test datasets submitted by CC.

Repeat call for test data submission via mailing list.

1. **Mail to Nexus Linguarum**

We do not send the mail, but Max will function as a bridge between NL and the morph module and address the NL interests with regard to the module development.

Max reach out to Hugo to inform on collaboration with NL Task 1.5.

@CC, Fahad: What is the status?

FK: Hasn’t been sent yet. We can send the following draft (based on Bettina’s proposed mail) today:  
*Dear Nexus Linguarum researchers,*

*We are reaching out to you to invite you to participate in the development of the new OntoLex-lemon Morphology Module.*

*After a long break, the regular work on the module was taken up again with the first telco last Wednesday. This module aims to enable the representation of morphological language data by addressing computational linguistic as well as language descriptive needs. Moreover, it aims to allow for the automatic generation of inflectional paradigms and word-forms of lexemes and the generation of lexemes that result from word-formation processes. A preliminary report of the current modeling status can be consulted here:*

*https://elex.link/elex2019/wp-content/uploads/2019/09/eLex\_2019\_33.pdf*

*Right now we are working on drafting a model for the automatic generation of inflected word-forms and derived lexemes. Since there is considerable overlap between the OntoLex-lemon community and the efforts being discussed in this COST action, we are hoping to work together in building a Morphology Module that fulfills the modeling needs current ontologies and vocabularies lack to express.*

*There are several ways of a collaborative exchange and participation in the module development:*

*- You are welcome to join our bi-weekly telcos on Wednesday at 1-2 pm CET (the next telco will be on February, 3rd): https://meet.google.com/rrs-garw-oxu?hs=122&authuser=0 We would also like to encourage you to share language data with us that addresses a specific modeling need you would like to be representable with the Morphology Module or that you like to provide as test data. Further ways for collaboration/participation?!*

*- You can enter your representation needs into this document: https://docs.google.com/document/d/1z1HjQ0\_WmdQwdRoktSFPGewOkzcHesdJiz0O4CFX1Q8/edit?usp=sharing.*

*- For the contribution of test data please use the folder: https://drive.google.com/drive/folders/1hNOj5qm34J7uGJrsudI4eDuQAMCFFfzv?usp=sharing.*

*We believe that a collaboration between our research efforts will lead to more accurate language data as Linked Data and are looking forward to hearing back from you soon.*

*Kind regards,*

*Christian, Fahad and Bettina*

[CC: fine with me, one reason to wait was to discuss the possible relation with T1.5 \*today\*]

[FK: There is no mailing list that includes everyone (I believe) so we will probably have to send it separately to each WG]

1. **Latest Morphology Module Draft (for purpose 2.1 and 2.2)**



More detailed fragment:



InflectionType corresponds to a allomorphy-free paradigm for one inflection slot (e.g. number). For non-agglutinative languages there will mostly be only one InflectionType per LexicalEntry.

Paradigm is an abstract type of inflection (e.g. 1st Latin declension). It is a container for all the allomorphy variants presented by InflectionTypes.

A Rule corresponds to a single cell in the table (allomorphy-free paradigm), and presents a way to construct forms with a FST/regular expression mechanism. In case it is not something needed in the dataset, morph:example should be specified.

More precisely, classes and properties for form generation are defined [here](https://docs.google.com/document/d/1SKSfj2lGcC5PeB2lexTRZDpCKNY2SxBEOAP7p58SwvI/edit)

MI: Morph and Rule seem to be in parallel. Can they be connected or can Morph be instead of Rule?

**ToDos**

* Max: add the 4 modeling needs from 20.01.2021 telco to document
* All: Collect more modeling needs with regard to purpose 2.1 and 2.2
* Bettina: repeat call for test data submission via mailing list