**Participants:**

Bettina Klimek (BK)

Jakub Šimek

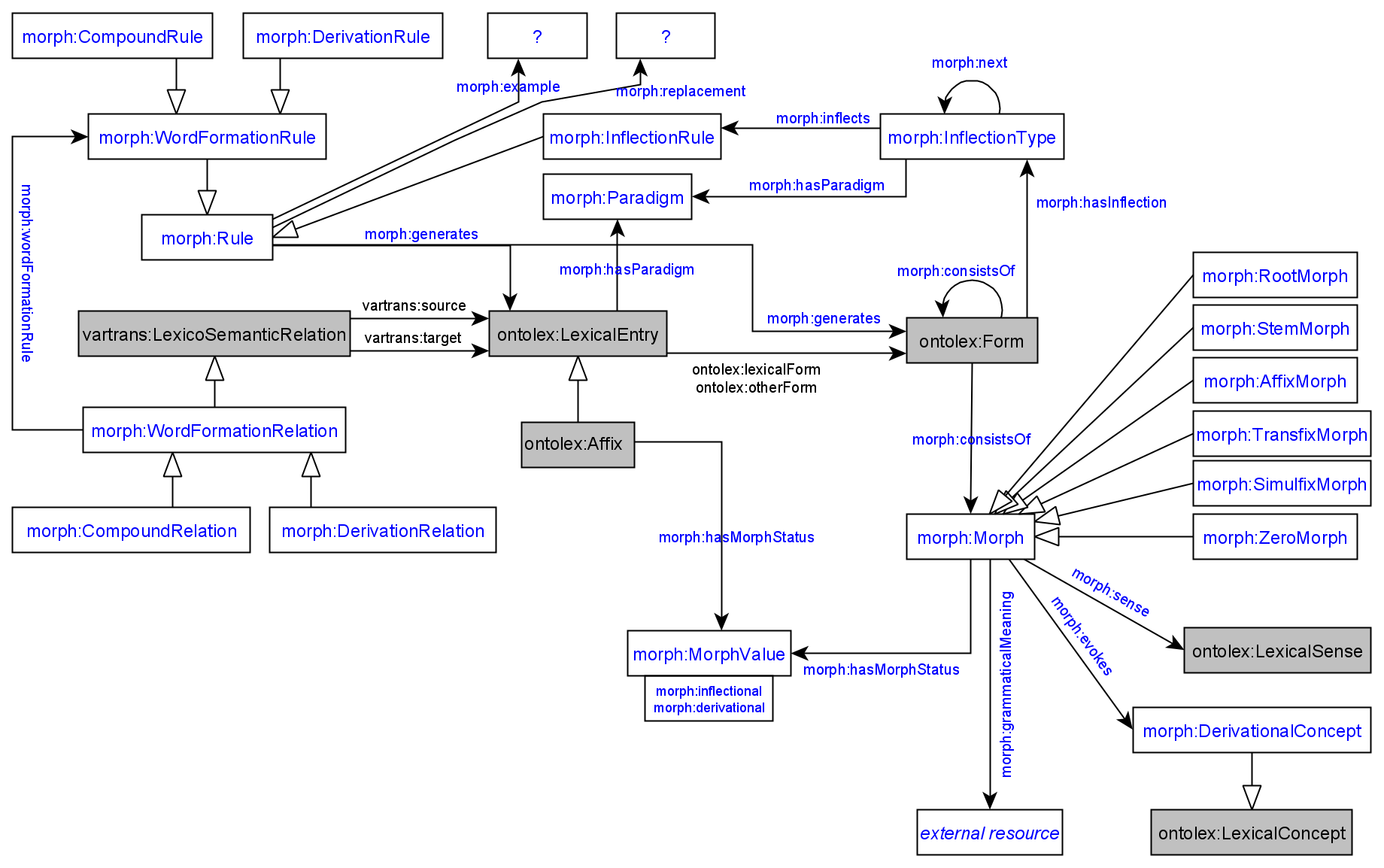
Max Ionov (MI)

Matteo Pellegrini

Fahad Khan

1. **Module draft 4.2**

Morph module status presentation at ontolex workshop: John and Max offered to find a solution

****

Adaptions of module draft 4.2 to be included for next telco:

* morph:generates: 1 rule generates ex. 1 ontolex:Form or ontolex:LexicalEntry resource
* extent domain and range of morph:consistsOf to morph:Morph

**2. Representation needs modeling**

**not modeled yet:**

**N7: Multiple segmentation strategies**

The segmentation of lexical entries or wordforms varies with different granularity:

German verb *jagte* "hunted"

Complete segmentation: root-stem-suffix

[[[jag]-t]-e] - [[[root]tense suffix]number suffix]wordform

Contracted segmentation: stem-suffix

[[jagt]-e] - [[past tense stem]number suffix]wordform

Christian: Does occur in Splett's Old High German dictionary (<https://brill.com/view/journals/abag/42/1/article-p264_28.xml>): Here, full morphological parses (tree structures) are being used. The other (main) use case is in language documentation (with Toolbox, from which dictionaries are being created): Linguistic glossing can operate on a superficial level or on a deep level, cf. German *fressen* ("to eat, of an animal") which superficially involves two morphemes (fress- + -en), but on a deep level involves three (\*ver- + ess- + -en, \*ver- contributing the derogative [non-human] meaning as in verwerfen "reject", lit. "cast away"). Normally, while one dictionary may chose one level of depth, another dictionary may chose another. Admitting more than one level of depth allows to merge information from different sources in a coherent representation. Wrt. morphological pattern: Isn't the idea that the morphological pattern describes a context for one given morph(eme)? So if have more than one (-t- and -e-) here, how will be formalize their combination?

Issue: the inflection rule forces the generation of two distinct ontolex:Form resources with different segmentation strategies (i.e. everything that is stated for ontolex:Form morph:consistsOf …) , but they are actually identical resources. Solution: interconnect both ontolex:Form resources with some “has segmentation variant” symmetric property?

BK suggestion: object property morph:hasSegVariant with ontolex:Form in domain and range. This would allow to have two ontolex:Form instances with separate segmentations indicated by their morph:consistsOf statements and to have only the desired ontolex:Form represented as a word form of an ontolex:LexicalEntry instance

Decision: not sure if this RN is worth the complexity that would have to be introduced, e.g. by a new property or a segment object container that could consist of separate segmentations with morph:consistsOf, maybe leave it open to the data creators to solve this issue

**N13: “missing” part of the stem becomes a separate token**

Description: I think there is a need to allow for morphology to break up a stem. I see John has raised a similar issue in N9, but what I am suggesting is that some tokens represent reduced forms of the stem/headword, but that the “missing” part of the stem becomes a separate token. Language example:

Eg. Old Irish verbs like do-beir:

1. Prototonic form is tabair (a verb), with the ta- mapping to the do- of the stem.

2. Deuterotonic form is do + beir (a particle + a verb).

In this case, while the headword, do-beir contains do-, the morphological form does not, and do- exists as a separate particle token. Pronouns can come between the particle and the verb and this is not considered tmesis.

BK:

How are particles, or grammatical words, treated in ontolex in general, as ontolex:LexicalEntry instances?! It is possible.

Is this an issue to be solved by the morphology module or by a prospective etymology module? In the latter case the morphology module would already provide the required morphological elements, i.e. stems.

Decision: created morph:consistsOf range for ontolex:LexicalEntry

Decision revised: a morph:Morph cannot consist of ontolex:LexicalEntry resources because these require per definition to have a lexical sense. Therefore, free morphs, such as particles, with grammatical meanings have to be morph:Morph resources and then this can be expressed by stating that a morph consists of another (free) morph that has a grammatical meaning: created morph:consistsOf range: morph:Morph

Adrian will create modeling example for what he needs in this RN:

**N14: variants of a single word form under a given lexical entry**

Description: It seems that the framework is built around having inflected variants of lexical entries. If I understand the data structure correctly that would imply that there is a lexical entry as a head, and each variant of that lexical entry is assumed to be an inflected variant, or word form.

Some languages, particularly historical varieties, will require the ability to have two or more tokens which look different, but represent the same word form under the one lexical entry.

Language example:

Eg. Latin length notation:

The Latin forms, nĕgōtĭum, negōtium, and negotium may all be found in different sources to represent the same word form, but are clearly different tokens. They should all be represented as variants of a single word form under a given lexical entry as they do not represent different morphology.

Eg. Old Irish verbs like ro-icc can have spelling variants, all representing the one word form. All of the following are attested present indicative forms of ro-icc:

1. 1 Sg.

a. roiccu

b. riccu

c. riccim

d. ruïcim

2. 2 Sg.

. rricci

a. rrici

3. 3 Sg.

. ric

a. ricc

4. 3 Pl.

. roeccat

a. roecat

b. recat

ro n ecat (verb split into particle, ro, and verb, ecat, by pronoun, n. See N13)

BK: Would it be sufficient to use ontolex:otherForm and extend the domain of this property to ontolex:Form as well in order to “represent the same word form under the one lexical entry”?

Decision: remove as RN, because this is addressed by ontolex vocabulary by assigning multiple written representations to one ontolex:Form that is related to the ontolex:LexicalEntry (i.e. headword)

John: the issue that the written representations are only strings and no resources has to be addressed by the ontolex core working group

**N15: Lexeme generation takes LexicalEntry and Form as input**

| Description: The generation of ontolex:LexicalEntry resources should allow to take resources of the type ontolex:LexicalEntry as well as ontolex:Form as input sources. This is required for languages which form new lexemes with inflected word-forms. |
| --- |
|  |
| BK suggestion: extend morph:consistOf range to ontolex:LexicalEntry next to morph:Morph and ontolex:Form- no, keep vartrans:source and vatrans:target restricted to ontolex:LexicalEntry |