**Participants:**

Bettina

Christian C.

Stefania Racioppa

Fahad Khan

Max

Julia

Ranka Stanković

**Agenda**

1. **Discussion and rating of old and new representation needs**

The representation needs will be discussed by choosing between the following two options:

1) The OntoLex morphology module should NOT be able to express the issue,

2) The OntoLex morphology module should be able to express the issue, as either

* a) being an explicit part of the model,
* b) being referenced to an existing external model or
* c) being no explicit part of the model but at least addressed in the model specification by providing a usage recommendation.

**N2: More Specifications of Affixes, Decision: open**

CC: no, if we express position information in another way

**N5: Morphemic Gloss, Decision: 1 (NOT)**

Max: not required

CC: can be modelled by FrAC attestation by reference to an example, e.g., an annotation in a corpus (e.g., using the Ligt vocabulary or whatever comes out of [LD4LT discussions on linguistic annotation](https://github.com/ld4lt/linguistic-annotation))

**N10: Suppletive forms replace a stem+affix combination with a wholly different word, Decision: 2a**

John: already covered (a lexical entry can have more than 1 root)

**N11: Morphology crosses part-of-speech boundaries (derivation), Decision:** postponed

Marco: Next meeting: LiLa sample data with a suggested extension

**N12: Morphs linked to Lexical Entries, Decision: 2a**

John: question is if we can underspecify the morphological pattern

**N13: Multiple segmentation strategies, Decision: 2a**

Way to allow more than 1 segmentation of a single ontolex:Form

**N14: The morphosyntax of a language describes how the morphemes in a word affect its combinatoric potential, Decision: 2a**

(= recursive generation)

**N15: The phones making up a morpheme don’t have to be contiguous, 2a**

**N16: The form of a morpheme doesn’t have to consist of phones, 2a**

**N17: Meanings of stems and roots, 2c**

Etymological dictionary for example, Marco can provide example

**Also N17: Link morphs and senses**

For roots or stems with lexical senses or lexical concepts, e.g., for semantic fields of roots , e.g., reconstructed protoforms (resp., their meaning) [why is Morph not a Lexical Entry?]

**N19: Derivational Meanings**

Cf. N11

**N21: “Missing” part of the stem becomes a separate token, 2a**

**N22: Variants of a single word form under a given lexical entry, 2a**

**N23: Word-form generation takes LexicalEntry and Form as input**

**N24: Coverage of morphological language types**

Incorporating languages (Inuktitut)

Agglutinating languages (Sumerian)

Polysythetic languages (Old High German, Old Irish)

Turkic languages

**N25: Suggestion for discussion: Agree on output data**

**Missing modelling needs:**

4 ways to represent inflected word-forms

* rules with transformations(Max)
* Inflection Tables with full inflected word form(Max)
* set of affixes/word endings (Max)
* Assimilation rules as additional rules on top of 1.

**Part of Speech transformation**

For lexical entries of categories that systematically come in two variants by “Zero Derivation”, e.g., every German adjective is an adverb, but this cannot be modelled with OntoLex vocabulary, and it is a productive process (so it’s within morphology)

**Recursive morphology**

Current focus is on inflectional morphology. But there are languages where after inflection is applied, additional layers of inflection can be applied (see data samples for Sumerian).

**Incorporation**

**I**n many languages, incorporation is a way for a verb to refer to a specific semantic role (usually THEME). This is a productive process and it corresponds in function to case inflection in standard average european, so within the realm of inflection (see data samples on Inuktitut).

**Weak noun/verb distinction**

Some languages are relatively flexible in “recasting” a verb into a noun, e.g., Standard Average European participles (verb => adjectives) but also finite verbs (Inuktitut qimmiuvuq “he has a dog” (verb) = “dog-owner” (noun), as a noun, this can be inflected as a noun, e.g., qimmiuvup “dog owner”.ERG). The mechanism here is that a particular type of inflection is “repurposed” for derivation. Technically, this can be treated like Zero Derivation.

**Derivation**

With recursion and incorporation, two morphological processes are addressed that share important characteristics with derivation: incorporation takes a noun and uses a verbal affix to produce a verb (so it involves a shift in parts of speech), recursive morphology is a recursive process (like most forms of derivation). As a generalization over both these aspects, derivation would require that an affix (morph?) posits constraints on the base form it is applied to (e.g., nominal for incorporation), and the grammatical features (“meaning”?) of the resulting form (e.g., verbal for incorporation). Sample data for derivation (plus the variants above) in Inuktit data sample.

**Compositional structure of compounds and derivations**

Basically, being able to represent morphological parses. See Old High German sample data.

**Slot grammars**

A number if agglutinative languages are described in terms of slots, i.e., morphemes that follow each other in a fixed order. This is not the same as a paradigm, because this is part of a paradigm only. See Sumerian sample data.

**Assimilation rules**

Depending on context, a morpheme can be serialized in different ways. So far, we decided to not model assimilation rules as productive rules, but we might want to capture interdependencies, e.g., vowel harmony in Turkic languages.

**Transliteration rules**

(technically similar to assimilation rules). Many languages have defective orthographies (e.g., writing a CCVCC language like Greek with a CV syllabary like Cypriot). Then, a morphologically derivable form must be mapped onto a particular orthography. To provide rules that directly produce orthographic forms may be too complex. Sample data from Sumerian (actual data sample use idealistic representations, not orthgraphic stzrings)

**2. Setup for testing pipeline**

Input data

* Ontolex dataset - [current test data](https://drive.google.com/drive/folders/1hNOj5qm34J7uGJrsudI4eDuQAMCFFfzv?usp=sharing) not in ontolex RDF- Preprocessing required?
* Extent? with/without morph:Morph resources

Module draft

* OWL file,
* SPARQL queries?!

Output data

* Separate file or integrated into ontolex input dataset
* Output integrated into existing OntoLex dataset already containing morphological data
* Granularity:
  + ontolex:LexicalEntry/ontolex:Form resources only
  + ontolex:LexicalEntry/ontolex:Form resources + morph:Morph resources
  + ontolex:LexicalEntry/ontolex:Form resources + morph resources + interrelation between ontolex:LexicalEntry/ontolex:Form and morph:Morph resources

Testing documentation

* Table with all modeling needs, stated input data and modeling draft, evaluation on results of output data against modeling needs
* Separate place to store input and output data?

**TODO:**

Bettina:

Check access to Ontolex Google folder for all

Example needed (for every need) with RDF example in module vocabulary

Include Latin word-formation extension into agenda-presentation by Marco