Morph telco 2022-04-05, 13:00 CEST

**Link:** [**https://meet.google.com/unn-ofrg-hdb**](https://meet.google.com/unn-ofrg-hdb) **[one-time-link; check here for link updates if it doesn’t work]**

**next :** [**https://meet.google.com/erm-dktq-knp**](https://meet.google.com/erm-dktq-knp)

**Latest Definitions:** [**https://github.com/ontolex/morph/blob/master/draft.md**](https://github.com/ontolex/morph/blob/master/draft.md)

**Nexus:** [**https://nexuslinguarum.eu/the-action/join-us**](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 (RS)

Gilles Sérasset (GS)

Mike Rosner (MR)

Table of Contents

[0. Module draft (4.17)](#_v4addnpz2vtu)

[Publications](#_1y5nmh4j3qo9)

[MWE Chapter](#_vlm6501m5uv2)

[UniDive](#_ovh2lj8i7l42)

[LDK workshop](#_g7lldhrmp0qh)

[Requirement freeze](#_3viriepxuvr)

[Status](#_qwbmiadubr87)

[Current open topics](#_vz8i95sgcpae)

[Datasets to be addressed](#_tk3lzsx0lee7)

[Character / sound classes](#_pfcfhjk2wtc7)

[morph:baseConstraint](#_29bag1tyi8ay)

[LDL challenges](#_nm87emzh5bnr)

[Semitic languages](#_f5mzkeb24i9j)

[Requirements (from wiki)](#_b77jzkoaql9r)

[Appendix: Organizational](#_cljq1d14u0f5)

[Procedure for Feature freeze](#_pre0dch5io6n)

# 0. Module draft (4.17)



No changes

Question: What is the difference between baseConstraint and grammaticalMeaning?

* **TODO**@Katerina: find a new wording in the definition ;)
* CC: original idea
  + Derivational (and inflection) morphemes can only be applied under certain conditions, base constraint is the precondition for a morpheme to apply, grammatical meaning is a feature of the derived (inflected) form/lexical entry, e.g.,

# Publications

## MWE Chapter

* Submitted: <https://www.overleaf.com/8285444258rpfnbwgwbrdp>
* Feedback delayed

## UniDive

* Christian has been asked to co-lead a task on modelling MWE dictionaries in Cost Action UniDive, Fahad would like to contribute [ppt](https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:2023-saclay:2023-03-16-unidive-paris-saclay-opening.pdf) [googledoc for wg2 meeting](https://docs.google.com/document/d/1NsBBXn5Z0YdknrwwbbJxBhVIrcrIpINBg_eje0t81Dc/edit)
* WG2 (corpus-lexicon interface) call tomorrow
  + Todo: add link to mailing list registration and registration procedure

## LDK workshop

* Prep call next week: April 12
* TODO@CC: dig out internal planning document (until 16:00)
* TODO@FK: contact people (today)
* TODO@CC: contact everyone else not reached by today (LD4LT, OntoLex chairs)

Last time:

* Fahad worked on the website
* Call for participation, program needed
* In contact with sara caravalho (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

# Requirement freeze

## Status

* requirement/dataset freeze as of March 22, 2023
* Last-minute additions (March 22)
  + Old Irish
  + nothing else

## Current open topics

* Character / sound classes
* morph:baseConstraint
* Finite state terminology
* Issues with semitic?
* MP: Naming “paradigm” => morphological patterns? Inflection type? Inflection Class?
* Wiki: requirements (see below)
* github/gdrive: data sets (see below)
* GitHub issues, comments from draft
* Matteo: generates to lexical forms
* Katerina: discuss the challenges from LDL if they are still challenges or not, whether they fall under morph. => separate section
* Besim: Clitics, e.g., Macedonian
  + Max: idea was to treat them like separable prefixes in German, it’s ok to have a space in a written representation of a form
  + **TODO**: find example data of lexical entries in 5 languages
    - Ranka: Samples for Serbian (link below)
* MWEs
  + As addressed in chapter
  + Gilles: describe decomposition of MWEs in DBnary w. Thierry
* Gilles: new derivational morphology for French
  + Technically, this is too late, but if there is a volunteer to take a look, we can take that feedback into account
* Non-concatenative morphology
  + 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> (Fahad)

## Datasets to be addressed

After the requirement freeze, this should be data previously discussed here or in joint papers.

NB: We don’t discuss the datasets per se, only the problems they pose to the model as of now. So the process is the following:

1. People responsible for the dataset (possibly, with the help of others) model the problematic part in OntoLex-Morph (example, not full dataset) offline
   1. For problems in modelling or finding information: reach out to moderators or mailing list (or Nexus Slack)
2. Deposit data in GitHub (or contact moderators)
3. Readme in GitHub: describe problematic points (if any)
   1. Problematic points can be: something you cannot model or something you made a design decision that you’d like to discuss
4. There should be a session devoted to each of the problematic datasets, presented by a person responsible for the dataset;
5. Datasets without problems (or readmes) are not presented, but can be used as examples when writing up the specification.

* Clitics examples: find example data of lexical entries in ~5 languages (**TODO**, partially on Besim)
  + E.g., Old Irish (TF). See [here](https://docs.google.com/document/d/1r7zbFt56zuYiySMmIsFtob-0U-7rSSLooeM2g2grQpU/edit#). The contents in the document and accompanying spreadsheet are possibly equally (or more) pertinent to “Current open topics” above.
  + Greek (Penny)
  + Serbian ([Ranka](https://docs.google.com/document/d/1d1-gQ1R_2WuIBsCrKiAIC9lqHZGA0wgJPAPJqu52sLA/edit))
  + Romanian (Elena & Ciprian) / Italian (?)
    - **TODO**@EA: particular examples for each clitic, not finished yet
    - Fahad on Italian: clitic example in Johanna and Maria Pia Buono's poster@unidive, <https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:2023-saclay:abstracts:wg2-3-monti-dibuono-poster.pdf>, in Ontolex on the poster
  + Any non-IE? E.g., as used in Toolbox data (**TODO**@Max)
    - Max: looking into this, no problematic cases so far
    - Written representations can contain spaces
    - Elena: I actually hadn't been thinking about italian, but maybe I have a suggestion about non-IE languages. Austronesian languages have clitics (for example samoan), I'm looking for dictionaries to check if clitics appear in lexical entries
* Maltese (Mike): [sample](https://docs.google.com/document/d/1Z7z36kjzgTjLCB3YJswqzh1lLhN9GDKZ2ecACcXYeqI/edit?usp=sharing)
* Serbian (Ranka): [paper for LDK - football use case, work in progress](https://www.overleaf.com/read/sdcqvrmgqzfw) [ttl](https://drive.google.com/file/d/1eZksRhCe0t7ffVLfdIJj4Tk1Ux5KtpbS/view?usp=sharing) (work in progress)
* German (Petra Steiner)?
  + CELEX
* Old Irish (Dorus, illustrative examples needed, cf. [here](https://docs.google.com/document/d/1r7zbFt56zuYiySMmIsFtob-0U-7rSSLooeM2g2grQpU/edit#) mentioned under clitics above)
  + mutations
  + Stem allomorphy / stem alternation
  + Polysynthetic constructions
* <https://github.com/ontolex/morph/tree/master/data>
  + agglutinating/Turkish (Christian)
  + foma/quechua (to be skipped, suggested by Christian)
  + fusional/lexis (Mod. Greek, Katerina/Penny)
  + gdrive/Latin\_Word\_Formation (Matteo)
  + gdrive/Agglutinative\_Sumerian.docx (Christian)
  + gdrive/Composition\_Derivation\_Old\_High\_German.docx (Christian)
  + gdrive/Italia\_wordform\_generation\_Stefania\_24062021.docx (Stefania Raciotta)
  + gdrive/Italian.docx (?)
  + gdrive/Polysynthetic\_Inuktitut.docx (Christian)
  + gdrive/Unimorph.docx (Christian)
* data on Latin inflection (Matteo Pellegrini)
  + **TODO**: link
* All data from our papers
  + LiLa: inflectional data
  + <https://github.com/acoli-repo/acoli-morph> (German, incl. UniMorph, UniDer, FST)
  + ?Data for the 2017 paper — individual examples from lexicographic resources, see paper (no additional data apart from what was in it)
  + Is there more?
    - 2017: Ancient Greek, Spanish, one more
    - 2019: Thierry & Stefania
* other data mentioned in minutes on demand
* From *latest* minutes (there is much more, we dig out/add at demand):
  + Arabic: [Morph Module semitic.odt](https://drive.google.com/file/d/1NST-qdkxAw7am6F3vGAap5o4eQv0ulV1/view?usp=sharing) (Khadija)
    - Other Semitic: not necessary
    - examples from Arabic and Hebrew from Ilan
      * **TODO**@Max: model, and let specialists take a look
  + Non-concatenative morphology
    - 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> (Fahad)
  + Tbc: coverage on non-European/non-Indo-European
    - Wiktionary: which languages? (no description of morphology, but programs to generate forms, LUA [!], cannot be extracted from DBnary data)
* To be updated: <https://github.com/ontolex/morph/blob/master/data/gdrive/Vocabulary_Tests_Evaluation.xlsx>

## Procedure

the process for datasets is the following:

1. People responsible for the dataset (possibly, with the help of others) model the problematic part in OntoLex-Morph (example, not full dataset) offline
   1. For problems in modelling or finding information: reach out to moderators or mailing list (or Nexus Slack)
2. Deposit data in GitHub (or contact moderators)
3. Readme in GitHub: describe problematic points (if any)
   1. Problematic points can be: something you cannot model or something you made a design decision that you’d like to discuss
4. There should be a session devoted to each of the problematic datasets, presented by a person responsible for the dataset;
5. Datasets without problems (or readmes) are not presented, but can be used as examples when writing up the specification.

As for requirements: try to move as much into offline mode as possible

* **Tbc. next time (suggestion following)**

FK: examples should reflect non-IE/non-European languages

CC: maybe two documents: tech spec (minimal examples) and best practices (prose, more languages)

**To be discussed**

# Topics for next time

## Confirm procedure for discussion requirements

## morph:baseConstraint

OLD:

* 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?

## ?Semitic languages

**DONE**: figure out how to include missing categories to LexInfo

=> vit GitHub issues under <https://github.com/ontolex/lexinfo/issues>

**OLD:**

Khadija: data prepared for Arabic:

* [Morph Module semitic.odt](https://drive.google.com/file/d/1NST-qdkxAw7am6F3vGAap5o4eQv0ulV1/view?usp=sharing)

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@Khadija:** Modelling examples for Arabic entries
* Not discussed yet: Circumfix
  + Morph:CircumfixParadigm
  + Prefix+suffix combination

## List of Requirements from Wiki

See below

# Topics not for next time

## Character / sound classes [NOT YET]

**OLD:**

**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

**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 dictrionaries 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

# LDL challenges [postponed]

OLD

* **TODO**@Katerina: check draft versions  
  - Variation in inflexion/Flexemes :: use to dialectal, diachronic or simply orthographic variation. Examples
  + In Latin, lavo `wash can be inflected according to either the 1st (lavare)

or the 3rd (lavere) conjugation.

* + Suppletion, eg in Old English, the verb wesan `to be' whose infinitive

represents one underlying root, whereas its indicative present singular forms

are based on two other roots (eom 1.sg. `(I) am'; bist 2.sg. `(you) are').

* + Modern Greek, `τραίνο' and `τρένο' have the same meaning and syntactic behaviour, so they can be modeled as the same ontolex:LexicalEntry, where the inflected forms of each are grouped together instead of all of them being represented as simply ontolex:otherForm. The current proposal is to introduce a new relation for orthographic variation (\onto{lexis:OrthVariant}) as a subclass of \onto{vartrans:LexicalRelation} and relate the orthographic variants through this relation.

- Markers of morphological variation :: labels of style, dating, dialect, etc.

* + Resolved by transfering the issue to LexInfo Vocabulary
  + Note :: it would be desirable if the OntoLex-Morph vocabulary would eventually be accompanied by best practice recommendations for the assignment of markers and provenance.

- Challenges in word formation ::

* + not fully predictable phonological processes like assimilation or apophony, which prevent the simple juxtaposition of formative elements from generating the actual surface form of derivatives;
  + formal and semantic constraints that make a word formation rule not applicable to all the words in the lexicon with a specific part of speech.

(morph:GrammanticalMeaning??)

# Requirements (from wiki)

| **N1: Morph resources** |
| --- |
| **Description**: In order to represent morphemic elements that do not apply to the restrictive definition of ontolex:Affix as being ontolex:LexicalEntry resources, a distinct class morphMorph is required as another top-level class next to ontolex:LexicalEntry and ontolex:Form. Moreover, with regard to a future etymology OntoLex module, it could serve as a means to represent data that has been identified and should be pointed to but to which no further detailed knowledge exists yet but might be added later. |
| **Required vocabulary**: owl:Class |
| **Initial consensus**: *approved modeling:*  morph:Morph a owl:Class ; rdfs:subClassOf owl:Thing . |
| **Status Updates**: as of 2021, we shifted towards modelling morph:Morphs as subclasses of ontolex:LexicalEntry. This was done to eliminate redundancy in morph-level form and sense attributes. |

| **N2: Specific morph resources** |
| --- |
| **Description**: Next to main morph:Morph class, more specific morph resources should be representable. For morphological representation, the elements root and stem should be assignable to classes. Further a morph:Affix class is required in parallel to ontolex:Affix to enable the representation of morphs that are not considered ontolex:LexicalEntry resources. Further, more specific affix types such as transfix (a discontinuous affix), simulfix (change or replacement of vowels or consonants (usually vowels) which changes the meaning of a word) and zero morph (a morpheme that has a morphological meaning that corresponds to no overt form) which are not covered by other existing RDF vocabularies are required as well. |
| **Language example**:  English Simulfix: a-->e in man (singular) vs. men (plural)  Hebrew Transfix:grammatical information is encoded in a discontinuous vowel pattern that is applied to a consonantic root pattern. E.g. the transfix a-a-a (third person, singular, past) is inserted into the root k-t-b 'all concepts evolving around writing' to render the word-form kataba 'he wrote'.  German Zero Morph: case and gender are not overtly marked in the German noun Herr 'master' and, thus, correspond to no overt form. The morpheme NOM.SG is realized by the zero morph Ø (i.e. Herr-Ø (at morph level) vs. ‘master-NOM.SG’ (at morpheme level)). |
| **Required vocabulary**: owl:Class |
| **Initial consensus**: *approved modeling:*  current modeling with fixed set of classes  mor ph:RootMorph, morph:StemMorph, morph:AffixMorph, morph:TransfixMorph, morph:SimulfixMorph, morph:ZeroMorph rdfs:subClassOf morph:Morph . |
| **Status Updates**: The need is agreed upon, but as of early 2022, we decided to move the subclassification of morphs into Lexinfo. This is because this hierarchy is partially provided in LexInfo v. 3.0, already, and users should not be confused with having multiple namespaces for information of the same kind (e.g., lexinfo:Suffix alongside morph:Simulfix). |

| **N3: Differentiation between derivational and inflectional morph resources** |
| --- |
| **Description**: With regard to representing the morphological content of lexical data the destinction between word-form forming (inflectional) and lexeme-forming (derivational) morph:Morph resources should be expressible and extractable. Concomitantly, the existing limitation of ontolex:Affix resources to represent only the latter type of morphs (due to its subclass relation to ontolex:LexicalEntry) will be overcome. |
| **Language example**:  German (homonym) suffixes:  1) *-er*: an inflectional affix forming comparative adjectives, e.g. *schön* 'beautiful' --> *schöner* 'more beautiful'  2) *-er*: a derivational affix forming agent nouns from verbs, e.g. *fahren* 'to drive' --> *Fahrer* 'driver' |
| **Required vocabulary**: Explicit identification of morph:Morph resources as being an inflectional or derivational morph. |
| **Initial consensus**: *initial modelling:*  morph:Morph morph:hasMorphStatus morph:Value .  morph:derivational a morph:Value .  morph:inflectional a morph:Value . |
| **Status Updates**: 2021/2022: The need is agreed upon, but with the inclusion of data from the LinkingLatin project, we shifted towards class-based modelling, i.e., WordFormationRule (resp. WordFormationRelation) vs. InflectionRule. Furthermore, we encode the difference between compounding and derivation in subclasses of WordFormationRule, resp. (partially) WordFormationRelation. |
| **current modelling**:  [a morph:WordFormationRule ] morph:involves [a morph:Morph ].  [a morph:CompoundingRule ] morph:involves [a morph:Morph ].  [a morph:DerivationRule ] morph:involves [a morph:Morph ].  [a morph:WordFormationRelation ] morph:wordFormationRule [ a morph:DerivationRule; morph:involves [ a morph:Morph ]] .  Note that here, we don't model the difference as a property of the morph, but as a property of the analysis and via morph:WordFormationRelation |

| **N4: Inflectional paradigm** |
| --- |
| **Description**: Lexical data contains pointers to and/or tables of inflectional paradigms or classes including the respective stem affixes or the full word-forms. Both, the pointers to paradigms and the interconnection of word-forms that belong to a paradigm, should be representable. |
| **Language example**:  Greek assignment of a lexical entry to an inflection class: λόγος:  mounce-morphcat: n-2a  Greek inflectional class paradigm: (with reconstructed underlying stem endings and desinence) n-3e(3):  NS: -ευς {-εϝ+ς}  GS: -εως {-εϝ+ος}  DS: -ει {-εϝ+ι}  AS: -εα {-εϝ+α}  VS: -ευ {-εϝ+}  NP: -εις {-εϝ+ες}  VP: -εις {-εϝ+ες}  GP: -εων {-εϝ+ων}  DP: -ευσι {-εϝ+σι}  AP: -εις {-εϝ+ας}  Examples for inflection tables with the inflectional paradigm structure and the inflected word-form. Latin:<https://en.wiktionary.org/wiki/Appendix:Latin_third_conjugation>  German:<https://de.wiktionary.org/wiki/Flexion:jagen> |
| **Required vocabulary**:  ontolex:LexicalEntry [object property] [morph:Paradigm] . ontolex:Form [object property] [morph:Paradigm] . |
| **Tested on data**: |
| **Status**: *agreed (version 4.16)*  ontolex:LexicalEntry lexinfo:morphologicalPattern morph:Paradigm .  ontolex:Form morph:inflectionRule morph:InflectionRule .  ontolex:InflectionRule morph:hasParadigm morph:Paradigm . |

| **N5: Morphology crosses part-of-speech boundaries (derivation)** |
| --- |
| **Description**: John (Issue derived from "Linguistic Fundamentals for Natural Language Processing" by Emily Bender, Source: <https ://www.morganclaypool.com/doi/abs/10.2200/S00493ED1V01Y201303HLT020>) |
| **Language example**:  Morphological processes can turn one part-of-speech into another, effectively creating a distinct LexicalEntry  English   * "to play" (verb) => "played" (adjective) * "to play" (verb) => "the playing" (noun) |
| **Required vocabulary**:  ontolex:LexicalEnty ontolex:lexicalForm ontolex:Form .  ontolex:Form morph:consistsOf morph:ZeroMorph . |
| **Tested on data**: |
| **Status**: *agreed modelling*  CC: This should include "zero derivation", where one word receives another part-of-speech without any difference in form or meaning. As an example, every German adjective can be used as adverb, most English prepositions also occur as subordinating conjunctions (complementizers) and verbal particles, etc. For "zero morphology", a distinct LexicalEntry is necessary only if differences in sense can be established. The underlying issue is that OntoLex does not permit more than one part-of-speech per LexicalEntry (which would be the natural modeling here).  Bettina: Derivation should be expressable at least as the underlying word-formation process. Whether the three different types of derivation (i.e. 1) zero derivation, 2) word-class changing derivation with no additional meaning and 3) word-class changing derivation with additional meaning) should be expressable depends on the needs of the lexicographers.  Current draft: use established means for derivation to represent conversion and specify zero morph, e.g. “play” (noun):  descriptive/extensional modelling:  ex:play\_v\_rel\_play\_n a morph:WordFormationRelation ;  vartrans:source ex:lex\_play\_verb ;  vartrans:target ex:lex\_play\_noun .  ex:lex\_play\_noun ontolex:lexicalForm ex:form\_play\_noun\_sg .  ex:lex\_play\_noun rdfs:member|morph:consistsOf ex:lex\_play\_verb, [a morph:ZeroMorph ].  or generative/intensional modelling:  ex:play\_v\_rel\_play\_n a morph:WordFormationRelation ;  vartrans:source ex:lex\_play\_verb ;  vartrans:target ex:lex\_play\_noun .  ex:lex\_play\_noun ontolex:lexicalForm ex:form\_play\_noun\_sg .  ex:play\_v\_rel\_play\_n morph:wordFormationRule/morph:involves [a morph:ZeroMorph ]. |

| **N6: Morphs linked to Lexical Entries** |
| --- |
| **Description**: Many dictionaries contain information about the morphology of a headword. This is typically given relative to the lemma. A possibility should be provided that enables an explicit statement of word-forms or morphemic elements that are given as part of the lexical entry. |
| **Language example**:  German(from "Langenscheidt Taschenwörterbuch Deutsch als Fremdsprache":   * **Bedingung** die; -, -en * **Bedürfnis** das; -ses, -se * **Beitrag** der; -(e)s, Beiträge   Note the does not cover all forms of the German noun, e.g., "Bedürfnissen", "Beiträgen"  It should be possible to model this information with two conditions:   1. It is not necessary to materialize all forms of the word, instead only the relevant stems and minimal set of inflected forms or inflectional morphemes 2. It is possible to generate any form in a programmatic manner   JMC: question is if we can underspecify the morphological pattern |
| **Required vocabulary**: 1. reuse vocabulary for automatic generation of word-forms and 2. create new property with ontolex:LexicalEntry in its domain to explicitly state which word-forms and/or morphs or grammatical information are considered custom extensions of a lemma. |
| **Tested on data**: |
| **Status**: *unclear if this representation need should be kept*  Look up TEI representation: <h ttps://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html> and<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>  Telco 09.06.2021:  Proposal: Object property morph:morphologicalForm could be created (in parallel to ontolex:lexicalForm) with domain ontolex:LexicalEntry and range morph:Morph  → different positions on whether this should be representable in the module at all because all information/data is already covered with the vocabulary and it is a need of space-restricted print dictionaries - discuss again later!  Status 07.09.2022: This can be done via morph:morphologicalPattern and morph:paradigm. However, THERE IS NO DIRECT LINK between morph:InflectionRule and morph:Morph, so this would be represented as string replacements, only, not as morphs. |

| **N7: Multiple segmentation strategies** |
| --- |
| **Description**: Way to allow more than 1 segmentation of a single ontolex:Form |
| **Language example**:  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 |
| **Required vocabulary**: |
| **Tested on data**: |
| **Status**: *to be discussed*  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?  Petra Steiner (7.9.2022): need for modelling derivation trees ((A B) C) confirmed.  Current recommendation: model with decomp, no designated vocabulary needed HOWEVER: not clear whether this supports multiple concurrent segmentations in a single data structure. |

| **N11: Meanings of stems and roots** |
| --- |
| **Description**: 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?] |
| **Language example**:  The meaning of stems and roots differ in the former are language-specific and the latter language-independent concepts. Stems have a word-class affiliation and often also entail grammatical information like tense and number (inherent inflectional meanings). As they function as the underlying semantic core of the lexical entry they occur in, the meanings of stems could be treated as the meanings of lexical entries. Roots, however, comprise very unspecific meanings from which words of various wordforms can be built.  Hebrew root k-t-b conveys the concept "anything related to writing". From this root nouns and verbs can be build, e.g. to write, journalist, author. |
| **Required vocabulary**:  ` ‘sense’ property  domain: ontolex:LexicalEntry, morph:StemMorph and morph:RootMorph `  range: 'sense' concept class |
| **Tested on data**: |
| **Initial proposal**: *modeled as draft:*  Bettina: The description of meanings of stems and roots could be realized in the same way as the description of meanings of lexical entries as given in ontolex. For the representation of roots maybe external resources such as Concepticon could be recommended or the possibility of a plain textual definition could be established in addition.  Discussed proposal: Extend domain of ontolex:sense with ontolex:LexicalEntry and morph:StemMorph and morph:RootMorph.  JMC: not in favour of extending ontolex:sense domain with morph:Morph, proposes new property morph:sense with ontolex:LexicalSense and another Concept class.  JBG: With the use of ontolex:LexicalSense we are assuming an ontological reference, so we might run into the same problems as the ones we found when converting dictionaries (which ontological references to point to?). Since in the lexicog specification we opted to stick to ontolex:LexicalConcepts for the meaning of lexical entries in the conversion of dict entries to LLD, why would we want to point to LexicalSense in this case, instead of Concept?  Current draft: property morph:sense with morph:Morph in domain and ontolex:LexicalSense in range  object property: morph:sense  domain: morph:Morph  range: ontolex:LexicalSense |
| **Status**: *solved*: use OntoLex core vocabulary, as morph:Morph is now a LexicalEntry |

| **N12: Derivational Meanings** |
| --- |
| **Description**: Issue derived from "Linguistic Fundamentals for Natural Language Processing" by Emily Bender, Source: <http s://www.morganclaypool.com/doi/abs/10.2200/S00493ED1V01Y201303HLT020> |
| **Language example**:  Diminutives create a new noun with a meaning of being smaller, this could be modelled by means of adding a small classes to the meaning of a noun. Three types of derivational meanings should be considered: Conversion: word-class change with no affxal marking and no additional meaning, e.g. play (v) → play (n)  Derivation 1: word-class change with affxal marking and no additional meaning, e.g. play (v) → playing (n)  Derivation 2: with or without word-class change with affxal marking and additional meaning, e.g. book (n) → booklet (n), play (v) → player (n) |
| **Required vocabulary**: class for representing derivational meanings, e.g. morph:DerivationalConcept |
| **Tested on data**: |
| **Status**: *modeled as draft:*  Diminuitives are not an ideal example because they are sometimes considered to be inflectional rather than semantic features (a form of degree, such as comparative). A better example might be the English morpheme "-er" which attaches to a verb to form a noun that represents the agent. The classic representation is by means of a rule: V + "-er" => N\_ag (CC)  John: Model derivational meanings as concepts and link morph instances to this concept.  Fahad: Ignore examples with lexicalized words (e.g. computer). We do not need to model too deeply - just state “diminutive”.  John: Proposes to have DerivationalConcept as subclass of ontolex:Concept (but no need for InflectionaConcept subclass).  Current draft: property morph:evokes with morph:Morph in domain and ontolex:LexicalConcept in range  object property: morph:evokes  domain: morph:Morph  range: ontolex:LexicalConcept  morph:DerivationalConcept rdfs:subClassOf ontolex:LexicalConcept .  **Current status**: NOT MODELLED: instead of morph:evokes, we can use ontolex:evokes. TBC: what is the added value of morph:DerivationalConcept |

| **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. |
| **Required vocabulary**: class for representing free and/or grammatical morphs and an object property that allows statements to express that a free/grammatical morph is part of an ontolex:Form or a complex morph:Morph resource |
| **Tested on data**: |
| **Status**: *consensus on modelling:*  object property: morph:consistsOf  domain: morph:Morph  range: morph:Morph  ontolex:Form morph:consistsOf morph:Morph . |

| **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. One example is compounding in German, where the modifier takes on inflected forms (e.g., *Gäste+haus* "guest house", lit. "guests' house" [plural]). |
| **Language example**: |
| **Required vocabulary**: morph:consistOf range: ontolex:Form |
| **Tested on data**: |
| **initial proposal**: *modeled as draft:*  The object properties vartrans:source and vartrans:target are reused and the range of morph:consistOf will not be extended to ontolex:Form. Any word-forms involved in the source or target of a generated ontolex:LexicalEntry have to be expressed by using morph:WordFormationRule.  vartrans:source  vartrans:target  morph:WordFormationRule |
| **current status**: to be droppen? no real data. extension of vartrans:source is possible but beyond scope (in vartrans). We'd need to suggest a vartrans:LexicalRelation between forms.  In German linguistics, an alternative view on compounding with inflected modifiers has been advocated, i.e., that the (diachronic) inflection now serves as interfix. This is supported by the fact that these "inflections" lost their grammatical meaning, so there is German *Gästehaus* (guest house) along with *Gasthaus* (restaurant), but the difference in meaning has nothing to do with the singular or plural morpheme that acts as interfix. |

# Appendix: Organizational

No discussion, just as a reminder, add other organizational things here

## Procedure for Feature freeze

Not yet, but the following *procedure* was discussed before:

* MI: I think we achieved a level where we freeze everything and write it up. I think that we can still have limitations unsolved for the final module, we just need to explicitly decide
* Should we vote in mailing list since not everyone is in the call
* CC and MI should decide when we’re ready to run the vote, then send to OntoLex chairs to approve