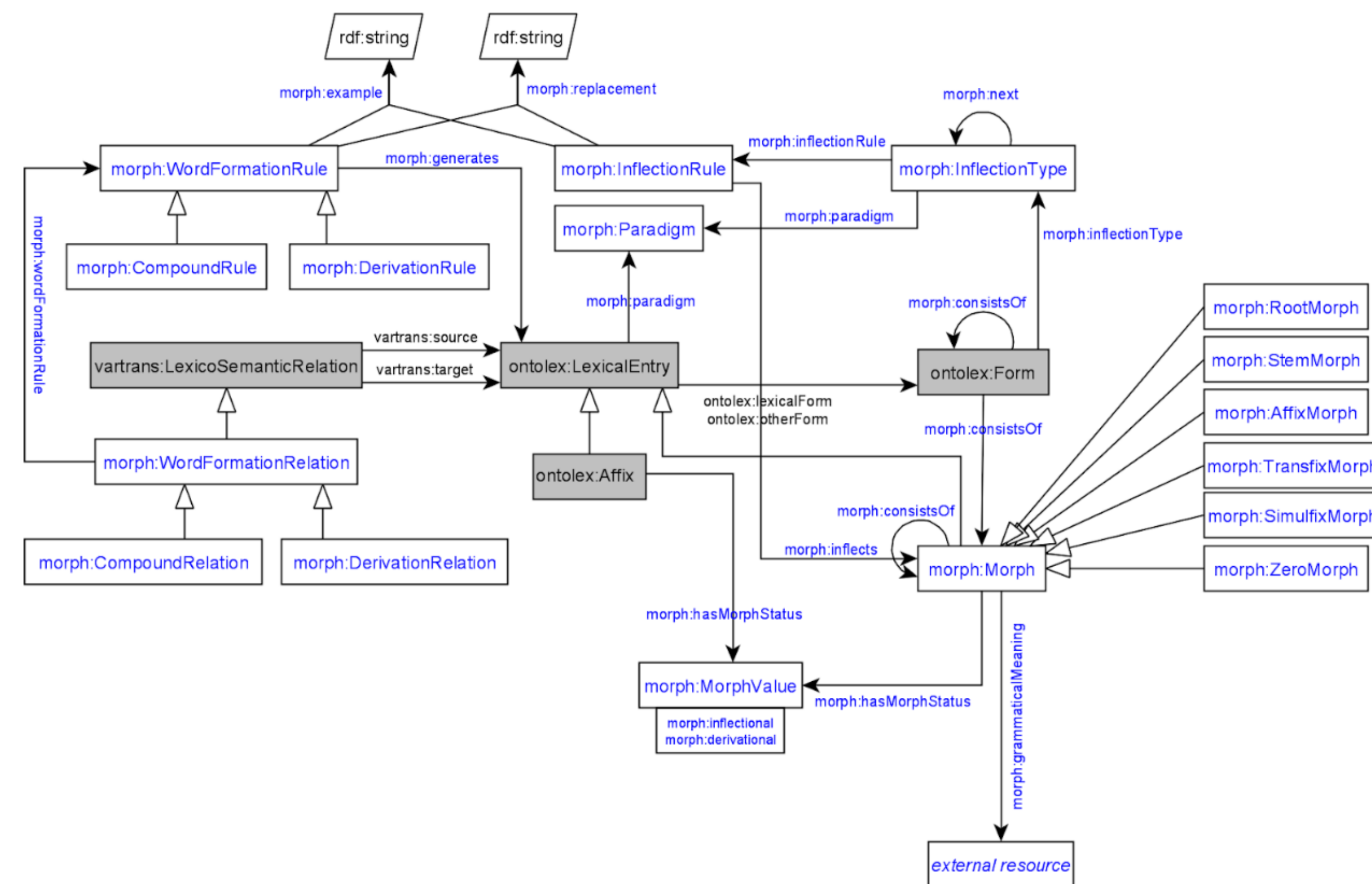


TL;DR

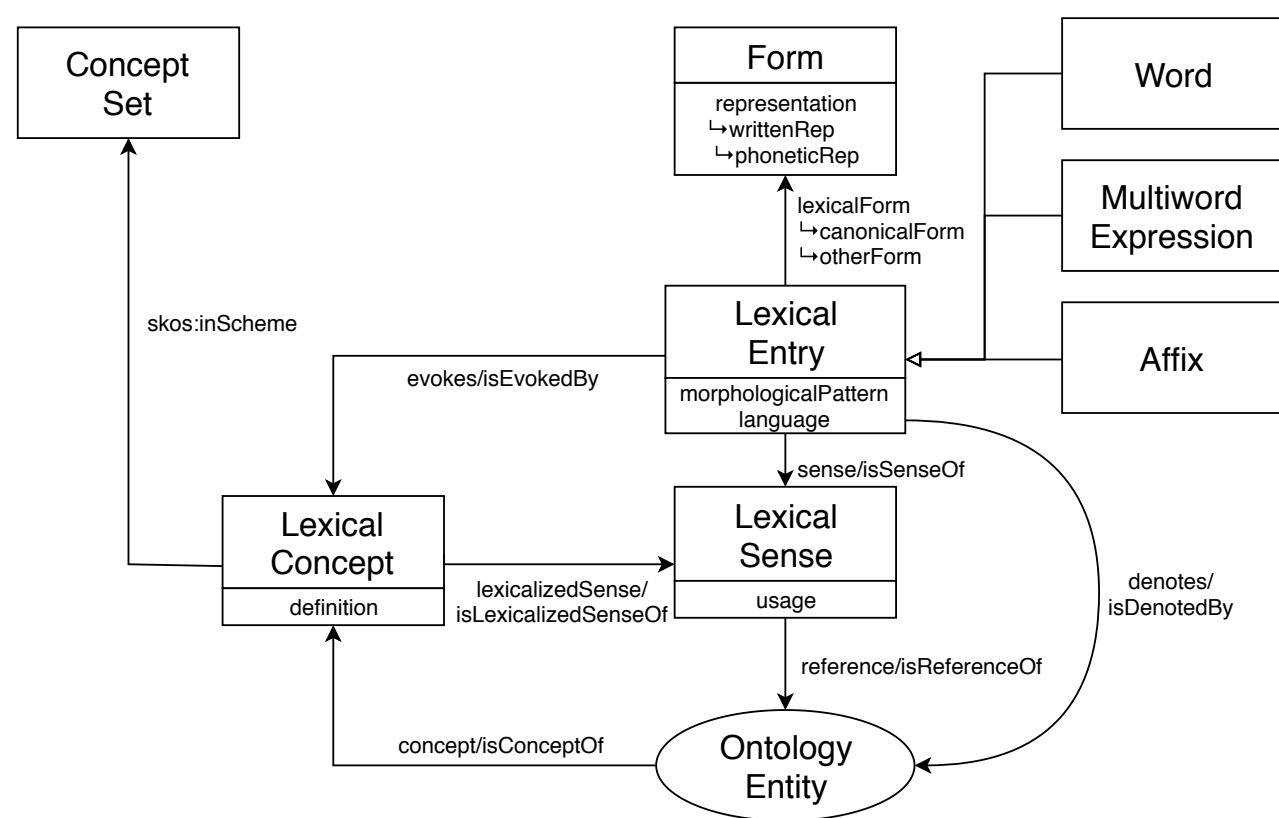
- **OntoLex-lemon** is a common way for **publishing lexical resources** in a **machine-readable way**
- **Lexical resources** might contain a considerable amount of **morphological information** (inflection, derivation, machine-readable ways of generating word-forms)
- ...**BUT** OntoLex core module is not expressive enough to represent it
- **OntoLex-morph** is being developed as an extension of OntoLex to **bridge this gap**
- **We converted 4 different types of morphological resources** to OntoLex-morph to **validate its current state**
- Overall, **the model in its current state can be successfully applied** to represent a wide array of morphological data
- Some problems and suggestions that we found are being addressed and will improve the model before its final publication

Morphology module: Current state



Background: OntoLex-lemon

- Used for modelling lexical information in RDF
- *Core model* extended with additional modules for special types of data: *synsem*, *lexicog*, *decomp*
- Morphology in the core model is limited to providing:
 - Morphosyntactic categories
 - Different forms for a lexical entry



OntoLex-lemon model is widely used for representing different lexical data:

- Dictionary data
- Lexical databases (WordNet)
- Etymological resources, etc.

Inflection: Unimorph

- Provides inflectional data for more than 100 languages
- Tables with three rows: lemma, (inflected) form and (morphosyntactic) features

Zufall Zufalls N;GEN;SG

- We extract morphological patterns by comparing canonical form and generated from, creating `morph:InflectionRule` with a regular expression to transform one to the other
- Conversion with TARQL: a single SPARQL query for TSV → RDF

```
form:Zufalls_N a ontolex:Form ;
morph:consistsOf <morph#s_N.GEN.SG> ;
morph:inflectionType type:N.GEN.SG .

type:N.GEN.SG a morph:InflectionType ;
morph:inflectionRule rule:s_N.GEN.SG ;
rule:s_N.GEN.SG a morph:InflectionRule ;
morph:inflects <morph#s_N.GEN.SG> ;
morph:replacement "s/$/".

<morph#s_N.GEN.SG> a morph:Morph ;
ontolex:lexicalForm "-s" ;
morph:grammaticalMeaning feats:N.GEN.SG .
```

Composition: GermaNet

- GermaNet — one of the resources for morphological composition in German
- Data is in TSV, as with Unimorph:

Zufallszahl Zufall Zahl

- Note the "linker" morpheme that is present in the compound but not in its parts
- Conversion with TARQL using string diff operations

```
entry:Zufallszahl a ontolex:LexicalEntry ;
ontolex:canonicalForm form:Zufallszahl .
[ a morph:CompoundRelation ]
vartrans:source entry:Zahl ;
vartrans:target entry:Zufallszahl ;
morph:contains <morph#s> .

entry:Zufallszahl decomp:subterm entry:Zahl ;
decomp:subterm entry:Zufall .

rule:s a morph:CompoundRule ;
morph:generates entry:Zufallszahl .
<morph#s> a morph:Morph , ontolex:Affix ;
ontolex:lexicalForm form:s_ .
```

Background: OntoLex-Morph

- **Central element** of the module — **morph:Morph**
 - Represents a morph, not a morpheme
 - Subclass of a `LexicalEntry`
- **Descriptive component**
 - Inflection
 - * `LexicalEntry` → `Form`
 - * Forms consist of `morph:Morphs`
 - * Representation of morph ordering is still not standardised
 - Derivation
 - * `LexicalEntry` → `LexicalEntry`
 - * Two lexical entries are connected with `morph:WordFormationRelations`
 - * Compounds can be represented with a subclass `morph:CompoundRelation` or using `decomp` module
- **Dynamic component**
 - Rules to describe how to create new derived entries or inflected forms
 - `morph:WordFormationRule` and `morph:InflectionRule`
 - Regular expression with source and replacement
 - Derivation: One rule per derived form
 - Inflection: One rule per `morph:InflectionType`
 - * only one for fusional languages (e.g. German, Latin)
 - * one for each grammatical category for agglutinative languages (e.g. Finnish, Turkic) connected with the `morph:next` property

Derivation: UDer

- Similar to Unimorph, provides derivational information for multiple languages.
- TSV data:

Zufall_Nm zufällig_A dNA05

- Conversion with TARQL by splitting the forms and their morphological information

```
entry:Zufall_Nm a ontolex:LexicalEntry ;
ontolex:canonicalForm form:Zufall_Nm .

entry:zufällig_A a ontolex:LexicalEntry ;
ontolex:canonicalForm form:zufällig_A .

form:zufällig_A a ontolex:Form ;
morph:consistsOf <morph#dVA03%3E> .

rule:dNA05%3E a morph:DerivationRule ;
morph:generates entry:zufällig_A ;
morph:replacement [] .

[ a morph:DerivationRelation ]
vartrans:source entry:Zufall_Nm ;
vartrans:target entry:zufällig_A ;
morph:contains <morph#dVA03%3E> .
```

Rules: SMOR/Morphisto

- Morphisto is a morphological analyser and generator based on SFST
- Generation rules are in explicit form
- Representing these rules in OntoLex-morph provide interoperability with other FST systems (XFST, Foma, etc.)
- The vocabulary was designed with linguistic perspective in mind, but can be adapted to FSTs:
 - `morph:InflectionType` → states
 - `morph:next` → transitions
 - `morph:InflectionRule` → replacement operation
- Recursive iteration over states applying corresponding transformations as long as there is a `morph:next` property will generate forms.
- This basic approach does not take into account morphophonological transformations (e.g. assimilation) and derivation but shows the potential of using OntoLex-morph for FST grammar representation bringing portability and interoperability.

Results, challenges and future solutions

- Overall, the model worked for all the types of morphological data it to which it was applied.
- Ordering of morphs is still not standardised;
- `morph:InflectionType` is highly debated and is subject to change.