

DMLex

1. DMLex Core

The DMLex Core is for monolingual lexical resources, where headwords, definitions, examples etc. are all in one and the same language.

1.1. LexicographicResource object type

A data set which can be viewed and used by humans as a dictionary and - simultaneously - ingested, processed and understood by software agents as a machine-readable database. Terminological note: *lexicographic* resource, not *lexical*.

Attributes:

- **Language** (IETF language code)

Children:

- **Entry** (one or more)

1.2. Entry object type

A part of a lexicographic resource which contains information related to exactly one headword.

Child of:

- **LexicographicResource**

Attributes:

- **headword** (non-empty string)
 - The headword can be a single word, a multi-word expression, or any expression in the source language which is being described by the entry in the lexicographic resource.
- **homographNumber** (number, optional)

Children:

- **PartOfSpeech** (zero or more)
- **Label** (zero or more)
- **Pronunciation** (zero or more)
- **InflectedForm** (zero or more)
- **Sense** (zero or more)

1.3. Sense object type

A part of an entry which groups together information relating to one of the (possibly multiple) meanings (or meaning potentials) of the entry's headword.

Child of:

- **Entry**

Attributes:

- **listingOrder**
 - Can be implicit from the serialization.
- **indicator** (optional, non-empty string)
 - A short statement that indicates the meaning of a sense and permits its differentiation from other senses in the entry.
- **definition** (optional, non-empty string)
 - A long statement that describes and or explains the meaning of a sense.

Children:

- **PartOfSpeech** (zero or more)
- **Label** (zero or more)
- **Pronunciation** (zero or more)
- **InflectedForm** (zero or more)
- **Example** (zero or more)

1.4. The difference between entries and senses

An entry is a container for those of the headword's properties which are meaning-independent. Those are usually the more "formal" properties such as orthography, morphology, syntax and pronunciation. An entry is also a container for pragmatic properties (such as register labels like 'vulgar' or 'archaic') if the lexicographer believes they are meaning-independent (= belong to all senses).

A sense is a container for those of the headword's properties which are meaning-dependent. Those are typically statements about semantics and pragmatics. A sense is also a container for formal properties (such as morphology, pronunciation) if these are meaning-dependent, such as when a headword has different plurals for different senses, or different pronunciations in different senses.

In a situation when an entry contains exactly one sense, it may be difficult to decide whether a piece of information (such as a pragmatic label) is to be assigned to the entry or to the sense. In such cases it should be assigned to the entry (ie. to the highest possible level).

When information of the same type appears on multiple levels in the entry, for example when one inflected form is given at the entry level and another inflected form at the level of an individual sense, the implication is that the forms are added together at the sense level: they both apply to the sense, neither overrides the other.

1.5. PartOfSpeech object type

Any of the word classes to which a lexical item may be assigned, e.g. noun, verb, adjective, etc.

Child of:

- **Entry**
- **Sense**

Attributes:

- **value** (non-empty string)
 - Can be constrained by the DMLex Controlled Vocabularies Module.

1.6. **Label** object type

An indication of some restriction on the use of the lexical item. The restriction can be pragmatic (time, region, register), semantic (domain, semantic type) or formal ('no plural').

Child of:

- **Entry**
- **Sense**
- **Pronunciation**
- **InflectedForm**

Attributes:

- **value** (non-empty string)
 - Can be constrained by the DMLex Controlled Vocabularies Module.

1.7. **Pronunciation** object type

Information about the pronunciation of its parent.

Child of:

- **Entry**
- **Sense**
- **InflectedForm**

Attributes (at least one):

- **transcription** (non-empty string)
- **recording** (string: name of URL of a sound file)

Children:

- **Label** (zero or more)

1.8. **InflectedForm** object type

An inflected headword is a form of the inflectional paradigm of its parent.

Child of:

- **Entry**
- **Sense**

Attributes:

- **label** (non-empty string) e.g. 'plural'
 - Can be constrained by the DMLex Controlled Vocabularies Module.
- **value** (non-empty string)

Children:

- **Label** (zero or more)
- **Pronunciation** (zero or more)

1.9. Example object type

An instance of a lexical item's usage in a specific sense.

Child of:

- **Sense**

Attributes:

- **text** (non-empty string)

2. DMLex Bilingual Module

Extends DMLex Core to support the encoding of bilingual lexicographic resources.

2.1. Extensions to **LexicographicResource** object type

Additional attributes:

- **targetLanguage** (IETF language code)

2.2. Extensions to **Sense** object type

Additional children:

- **HeadwordTranslation** (zero or more)

2.3. **HeadwordTranslation** object type

The translation equivalent of the headword in one of its senses.

Child of:

- **Sense**

Attributes:

- **text** (non-empty string)
 - Can be a single word, a multi-word expression, or indeed any expression in the target language.

Children:

- **TranslationPartOfSpeech** (zero or more)
- **TranslationLabel** (zero or more)
- **TranslationPronunciation** (zero or more)
- **TranslationInflectedForm** (zero or more)

2.4. TranslationPartOfSpeech object type

Any of the word classes to which the translation may be assigned, e.g. noun, verb, adjective, etc.

Child of:

- **HeadwordTranslation**

Attributes:

- **value** (non-empty string)
 - Can be constrained by the DMLex Controlled Vocabularies Module.

2.5. TranslationLabel object type

An indication of some restriction on the use of its parent. The restriction can be pragmatic (time, region, register), semantic (domain, semantic type) or formal ('no plural').

Child of:

- **HeadwordTranslation**
- **TranslationPronunciation**
- **TranslationInflectedForm**

Attributes:

- **value** (non-empty string)
 - Can be constrained by the DMLex Controlled Vocabularies Module.

2.6. TranslationPronunciation object type

Information about the pronunciation of its parent.

Child of:

- **HeadwordTranslation**
- **TranslationInflectedForm**

Attributes (at least one):

- **transcription** (non-empty string)
- **recording** (string: name or URL of a sound file)

Children:

- **TranslationLabel** (zero or more)

2.7. TranslationInflectedForm object type

A form of the inflectional paradigm of its parent.

Child of:

- **HeadwordTranslation**

Attributes:

- **label** (non-empty string) e.g. 'plural'
 - Can be constrained by the DMLex Controlled Vocabularies Module.
- **value** (non-empty string)

Children:

- **TranslationLabel** (zero or more)
- **TranslationPronunciation** (zero or more)

2.8. Extensions to Example object type

Additional children:

- **ExampleTranslation** (zero or more)

2.9. ExampleTranslation object type

The translation of an example.

Child of:

- **Example**

Attributes:

- **text** (non-empty string)

3. DMLex Entry Structuring Module

3.1. SenseGroup relation type

Represents the fact that a group of senses (all belonging to the same entry) should be grouped when

presented to a human user. Typically, when an entry has a large number of senses, it is a convenience to the human user to group them into a smaller number of groups by some broad criterion, such as by part of speech or by semantic similarity.

Participants:

- **Sense** (two or more)

Attributes:

- **indicator** (optional, non-empty string)
 - A short statement that indicates the broad meaning that unites the senses in this group and permits their differentiation from other senses in the entry.

Children:

- **PartOfSpeech** (zero or more)
- **Label** (zero or more)
- **Pronunciation** (zero or more)

3.2. Subsense relation type

Represents the fact that one sense (the subordinate sense) should be treated as a subsense of another sense (the subordinate). Both senses belong to the same entry.

Participants:

- the superordinate **Sense** (exactly one)
- the subordinate **Sense** (exactly one)

3.3. Subentry relation type

Represents the fact that one entry (= the subordinate entry) should be treated as a subentry inside the sense (= the superordinate sense) of another entry.

Participants:

- the superordinate **Sense** (exactly one)
- the subordinate **Entry** (exactly one)

Attributes:

- **listingOrder**
 - Can be implicit from the serialization.

4. DMLex Crossreferencing Module

4.1. Variant relation type

Represents the fact that two entries are understood by the lexicographer as variants (for example masculine and feminine counterparts, spelling variants).

Participants:

- **Entry** (two or more)

4.2. **Opposition relation type**

Represents the fact that two senses (typically - but not necessarily - belonging to two different entries) have opposite meanings. This includes antonyms, converses and so on.

Participants:

- **Sense** (exactly two)

4.3. **Similarity relation type**

Represents the fact that two or more senses (typically - but not necessarily - belonging to two different entries) have the same or similar meanings. This includes synonyms, near synonyms, immediate hypernyms/hyponyms and so on.

Participants:

- **Sense** (two or more)

4.4. **Pertainment relation type**

Represents the fact that two or more senses (typically - but not necessarily - belonging to two different entries) are related to each other, in ways other than opposition and similarity.

Participants:

- **Sense** (two or more)

5. **DMLex Inline Markup Module**

5.1. **Placeholder markup type**

Marks up a substring inside a headword (or inside a headword translation) which is not part of the expression itself but stands for things that can take its place, or constitutes some kind of meta-notation.

Examples:

- **beat [sb.] up**
- **continue [your] studies**

Markup of:

- **headword** attribute of **Entry**
- **text** attribute of **HeadwordTranslation**

5.2. Headword markup type

Marks up a substring inside an example (or inside an example translation) which corresponds to the headword (or to a translation of the headword).

Markup of:

- **text** attribute of **Example**
- **text** attribute of **ExampleTranslation**

6. DMLex Controlled Vocabularies Module

This module makes it possible to describe constraints on the values of some plain-text attributes of objects defined in DMLex Core and in DMLex Bilingual Module.

6.1. Extensions to **LexicographicResource** object type

Additional children:

- **PartOfSpeechOption** (zero or more)
- **TranslationPartOfSpeechOption** (zero or more)
- **LabelOption** (zero or more)
- **TranslationLabelOption** (zero or more)
- **InflectedFormOption** (zero or more)
- **TranslationInflectedFormOption** (zero or more)

6.2. **PartOfSpeechOption** object type

Represents one of several allowed values for the **value** attribute of **PartOfSpeech** objects.

Attributes:

- **value** (non-empty string)
- **displayValue** (optional)

Children:

- **OptionMapping** (zero or more)

6.3. **TranslationPartOfSpeechOption** object type

Represents one of several allowed values for the **value** attribute of **TranslationPartOfSpeech** objects.

Attributes and children same as above.

6.4. **LabelOption** object type

Represents one of several allowed values for the **value** attribute of **Label** objects. Attributes and children same as above.

6.5. **TranslationLabelOption** object type

Represents one of several allowed values for the **value** attribute of **TranslationLabel** objects. Attributes and children same as above.

6.6. **InflectedFormOption** object type

Represents one of several allowed values for the **label** attribute of **InflectedForm** objects. Attributes and children same as above.

6.7. **TranslationInflectedFormOption** object type

Represents one of several allowed values for the **label** attribute of **TranslationInflectedForm** objects. Attributes and children same as above.

6.8. **OptionMapping** object type

Represents the fact that an item in the controlled vocabulary is equivalent to item provided by an external authority.

Parents:

- **PartOfSpeechOption**
- **TranslationPartOfSpeechOption**
- **LabelOption**
- **TranslationLabelOption**
- **InflectedFormOption**
- **TranslationInflectedFormOption**

Attribute:

- **sameAs** (URI)