

---

---

**Language resource management —  
Lexical markup framework (LMF) —**

**Part 3:  
Etymological extension**

*Gestion des ressources linguistiques — Cadre de balisage lexical  
(LMF) —*

*Partie 3: Extension étymologique*

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-3:2021



STANDARDSISO.COM : Click to view the full PDF of ISO 24613-3:2021



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2021

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
CP 401 • Ch. de Blandonnet 8  
CH-1214 Vernier, Geneva  
Phone: +41 22 749 01 11  
Email: [copyright@iso.org](mailto:copyright@iso.org)  
Website: [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

	Page
<b>Foreword</b> .....	<b>iv</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Normative references</b> .....	<b>1</b>
<b>3 Terms and definitions</b> .....	<b>1</b>
<b>4 The LMF etymology extension</b> .....	<b>2</b>
4.1 The Cognate class and the Etymon class.....	2
4.2 The Etymologizable class.....	2
4.3 The Etymology class and the EtyLink class.....	3
4.4 The CognateSet class.....	4
4.5 The Date class.....	4
4.6 The Gloss class.....	4
<b>Annex A (informative) Examples of possible etymological typologies</b> .....	<b>6</b>
<b>Annex B (normative) Data categories for etymology description</b> .....	<b>15</b>
<b>Bibliography</b> .....	<b>22</b>

STANDARDSISO.COM : Click to view the full PDF of ISO 24613-3:2021

## Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see [www.iso.org/directives](http://www.iso.org/directives)).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see [www.iso.org/patents](http://www.iso.org/patents)).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see [www.iso.org/iso/foreword.html](http://www.iso.org/iso/foreword.html).

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 4, *Language resource management*.

This first edition of ISO 24613-3, together with ISO 24613-1:2019, ISO 24613-2:2020, ISO 24613-4:2021 and ISO 24613-5:—<sup>1)</sup>, cancels and replaces ISO 24613:2008, which has been technically revised.

The main changes compared to the previous edition are as follows:

- entire revision of the content and its subdivision into several parts.

A list of all parts in the ISO 24613 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at [www.iso.org/members.html](http://www.iso.org/members.html).

---

1) Under preparation. Stage at the time of publication: ISO/DIS 24613-5:2020.

# Language resource management — Lexical markup framework (LMF) —

## Part 3: Etymological extension

### 1 Scope

This document describes an extension to ISO 24613-1 and ISO 24613-2 to support the development of detailed descriptions of common etymological phenomena and/or diachronic information with respect to lexical entries in born-digital and/or retro-digitized lexicons. It provides both a meta-model for such an extension as well as the relevant data categories.

### 2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8601-1, *Date and time — Representations for information interchange — Part 1: Basic rules*

ISO 8601-2, *Date and time — Representations for information interchange — Part 2: Extensions*

ISO 24613-1, *Language resource management — Lexical markup framework (LMF) — Part 1: Core model*

ISO 24613-2, *Language resource management — Lexical markup framework (LMF) — Part 2: Machine-readable dictionary (MRD) model*

### 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 24613-1 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <http://www.electropedia.org/>

#### 3.1

##### **cognate**

form in a related language which shares a common etymological origin as a form in the language of the lexicon

#### 3.2

##### **etymologizable**

meeting the conditions for having an *etymology* (3.3)

Note 1 to entry: "Etymologizable" is a category of lexical elements and usages (encompassing for instance lexical entries, senses, word forms).

#### 3.3

##### **etymology**

origin and historical development of any aspect of a given lexical item

**3.4 etymon**

lexical entry from which another lexical entry is derived

Note 1 to entry: An etymon can also be simply an earlier stage of a lexical item.

**3.5 onomasiology**

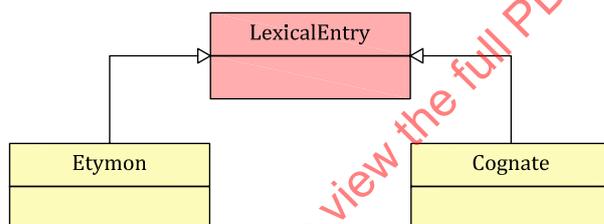
approach to the investigation of word meaning which takes a given concept as a starting point and studies the different lexical items in a language or languages that are used to refer to it

**4 The LMF etymology extension**

NOTE See [Annex A](#) for examples of possible etymological typologies.

**4.1 The Cognate class and the Etymon class**

Cognate and Etymon are defined as subclasses of the LexicalEntry class from the LMF core module (see [Figure 1](#))<sup>2)</sup>. Both classes define lexical entries which have been added to a lexical resource with the purpose of describing the etymologies of one or more other lexical entries. Instances of either Etymon or Cognate can be assigned a language which is different from the language of the lexicon as a whole (this is specified in the LexiconInformation class as described in ISO 24613-1).



**Figure 1 — Cognate and Etymon as subclasses of LexicalEntry**

Individuals of both the Etymon and the Cognate classes shall be in an aggregation relationship with at least one individual of type EtyLink (see [4.3](#)). When describing etymologies, there are cases in which it is necessary to deal with instances of LexicalEntry (and hence also by the subclass relation instances of Etymon and Cognate) which are roots, and in particular reconstructed roots. In these cases, the fact of being a root and the type of the root in question shall be specified using the attribute rootType. In the case of reconstructed roots or other word forms, the attribute status serves to associate the element with a written description of the likelihood of its having been in use (see the example in [A.8](#)). See [Table 1](#) for a list of attributes to be used with these two classes.

**4.2 The Etymologizable class**

The Etymologizable class provides a means of referring to the set of linguistic elements that can have etymologies. By defining a single class encompassing all such ‘etymologizable’ elements, the classes of elements which can have etymologies can be easily extended in the future wherever the necessity arises. The following classes are subtypes of the Etymologizable class (see [Figure 2](#)): LexicalEntry, Sense, Form and CognateSet (see [4.4](#)).

2) In this document, the following colour scheme is used in diagrams: classes in yellow are introduced in this document, and classes in pink have been previously introduced in ISO 24613-1 and ISO 24613-2.

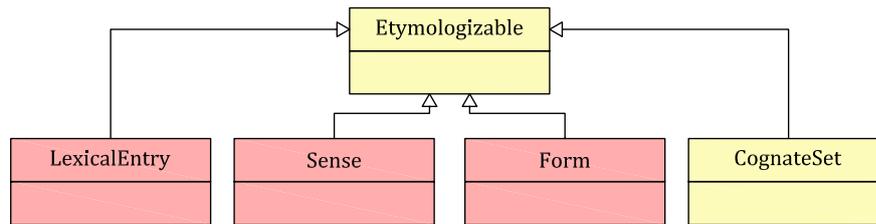


Figure 2 — The Etymologizable class and its subclasses

### 4.3 The Etymology class and the EtyLink class

The Etymology class allows for the description of the etymology of a linguistic element. More specifically, it allows for the description of those linguistic elements that are subclasses of the Etymologizable class. The type or types of etymological process involved in a given etymology can be specified using the type attribute, and also potentially the subtype attribute (in the case when the type of the etymology can be further specified). Possible values for type and subtype can vary according to the theoretical approach adopted by the compiler of a resource and/or the linguistic or editorial focus of the resource. The use of nested Etymology instances allows a combination of etymological processes to be described. Examples of etymological processes that shall be used as values for type/subtype: *borrowing, inheritance; word formation: compounding, derivation; sense shifts: narrowing, widening, amelioration, pejoration, metaphor, metonymy; phonetic/phonological processes: place assimilation, dissimilation, epenthesis, metathesis, hardening, weakening, etc.* The list of data categories provided in [Annex B](#) shall be used in complement to the appropriate classes. Individual links between two elements in an etymology can also be given a type, see the description of EtyLink below. Given that an Etymology instance can be taken from an external source, it can be associated with a Bibliography instance, which shall be defined as per ISO 24613-2 (see [Figure 3](#)).

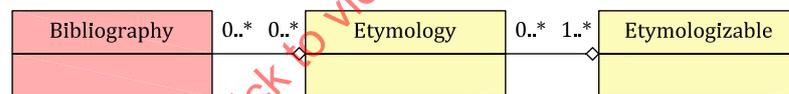


Figure 3 — The Etymology class

Instances of Etymology are associated with one or more EtyLink instances, each of which represents a single stage or step in the etymology of a given lexical item (see [Figure 4](#)). EtyLink serves to link together individuals belonging to the subclasses of Etymologizable. EtyLink is a subclass of the CrossREF class as defined in ISO 24613-1. The use of CrossREF requires that the target objects representing the given lexical content be given id attributes. The use of the id attribute on an individual of the Etymologizable class as a target allows for the modelling of a generic sequential temporal ordering of multiple elements, using the attributes prev and next. Instances of the EtyLink class can further specify additional temporal relationships using various temporal attributes associated with the source and target of each EtyLink instance.

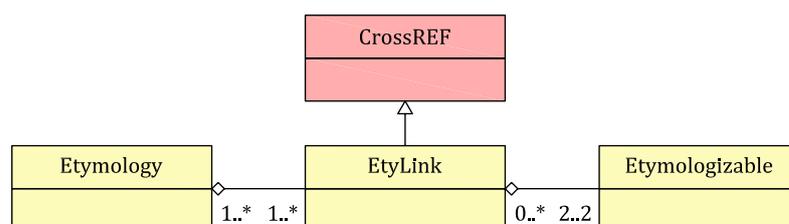


Figure 4 — EtyLink

Individuals of the Etymon and Cognate classes (subtypes of LexicalEntry) shall be associated with at least one individual of EtyLink. See [Table 1](#) for a list of attributes to be used with the Etymology and EtyLink classes.

#### 4.4 The CognateSet class

The CognateSet class (see [Figure 5](#)) is a container for sets of one or more Cognate items and zero or more Bibliography items (see ISO 24613-2). The CognateSet is a construct related to onomasiology. Its contents are items from languages related to that of a given LexicalEntry (and therefore by the subclass relation of any given Etymon or Cognate) and which have been gathered together with the purpose of demonstrating linguistic similarities or dissimilarities of salient kinds. The use of CognateSet implies that the LexicalEntry (and therefore Etymon and Cognate) instances which it contains share an etymological source.

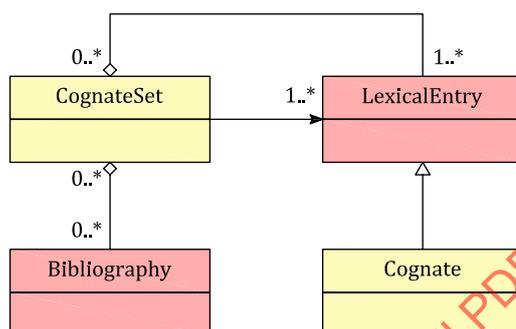


Figure 5 — CognateSet

#### 4.5 The Date class

The components of a LexicalEntry and its subclasses shall be associated with a specific date by making use of the Date class. Furthermore, Date allows the specification of a number of degrees of precision. A precise year, and potentially month and day, shall be stated using the date attribute and a rough date with the attribute circa. Within a span of time with different levels of specificity, there is the possibility of using one or more dating attributes. Where a span of time is known (or asserted), the lower and upper ends of the span can be specified using notBefore, notAfter respectively. For date and time formats, ISO 8601-1 and ISO 8601-2 shall be used.

#### 4.6 The Gloss class

The Gloss class (see [Figure 6](#)) represents a textual description of the meaning of a word or a phrase that is intended for human consumption. Individuals of the class can either represent paraphrases or synonyms and these may be in the language of the entry or in another language. See [Table 1](#) for a list of attributes to be used with this class.

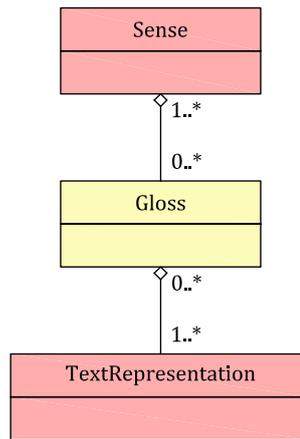


Figure 6 — Gloss

Table 1 — Example of class adornment

Class name	Example of attributes
Etymon	xml:lang, gloss, rootType, status
Etymology	type, subtype
EtyLink	type, prev, next
CognateSet	
Cognate	xml:lang, gloss, rootType, status
Date	notBefore, notAfter, circa, date
Gloss	xml:lang

## Annex A (informative)

### Examples of possible etymological typologies

#### A.1 Example of simple inheritance

The example in [Figure A.1](#) describes the inheritance of a lexical entry from a parent language, in this case the adverb *semper* in Sardinian which comes from the Latin word *semper*. The Sardinian word is linked to a single Etymology instance which is associated with the type *inheritance* (see [Annex B](#) for a definition of this type). The Etymology is then associated with an individual of type Etymology which represents the process of change from the Latin etymon to the Sardinian lexical entry.

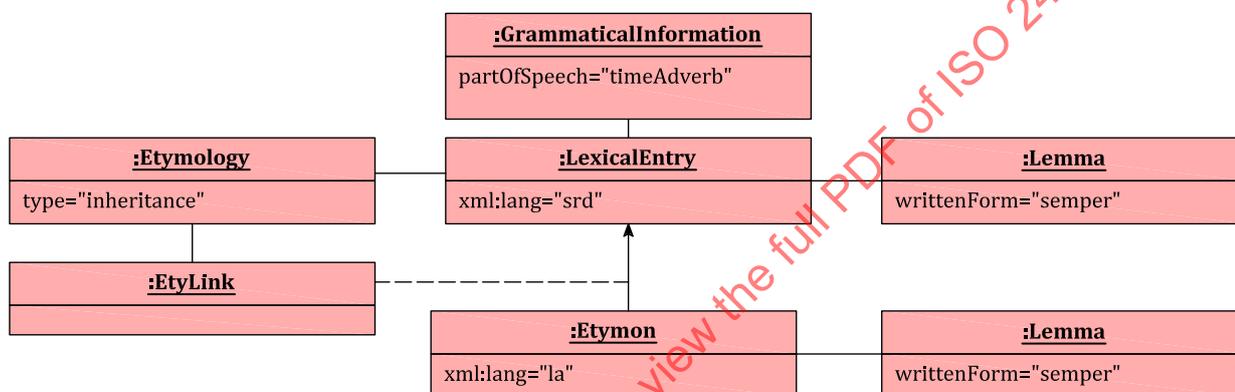


Figure A.1 — Diagram of inheritance in Sardinian

#### A.2 Example of a diachronic etymological process (inheritance with phonological change)

In the following example, the development of the word meaning 'wine' (ipa: [veŋ]) in the Emiliano variety of Italian is traced back using a series of Etymons which are ordered and linked together using Etymology instances, from the Vulgar Latin *vinu* through to the immediate predecessor of the word in its current manifestation. These individual links can be accessed through an Etymology individual (with the type *inheritance*) which represents the history of the LexicalEntry. The ordering of Etymons is implemented by means of the two attributes `prev` and `next`. These attributes are not displayed in [Figure A.2](#) for reasons of space, but their use is shown in [Figure A.6](#), for the example given in [A.6](#), and in [Figure A.10](#), for the example given in [A.8](#).

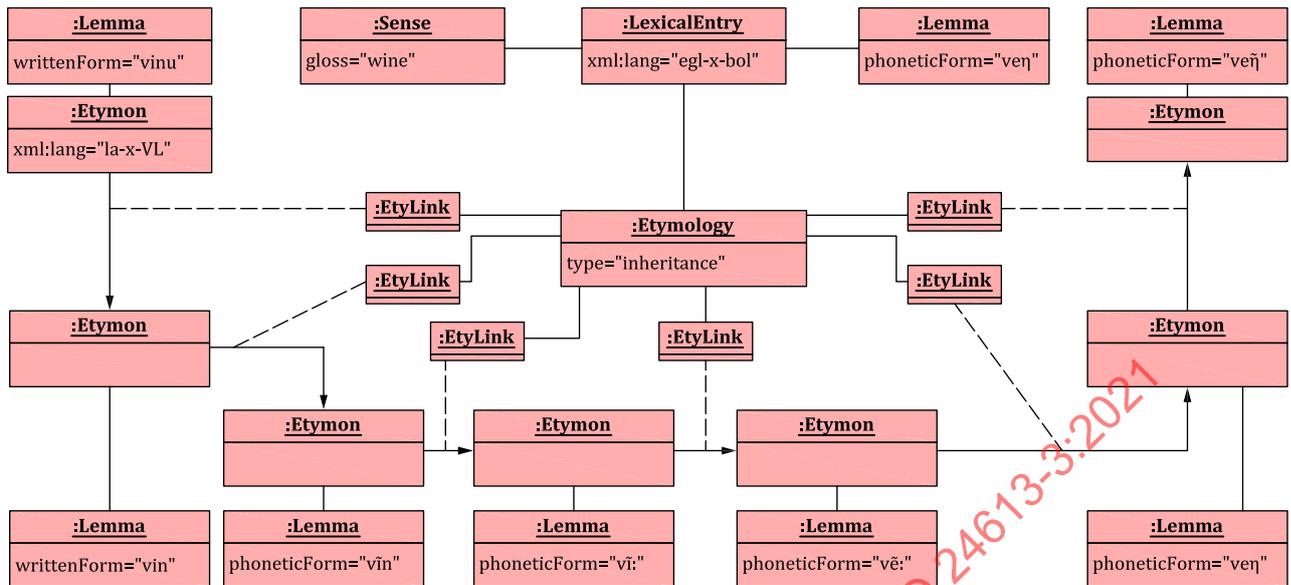


Figure A.2 — Diagram of multi-stage inheritance and phonological change in Bolognese

### A.3 Example of word form inheritance

The derivation of the singular and plural forms of the Portuguese noun *nação* ‘nation’, *nação* (sg) and *nações* (pl), respectively, derived from two forms of a Vulgar Latin (VL) noun, *nātiōnem* (sg, acc), *nātiōnes* (pl, acc) is described in Figure A.3. Herein, since the Portuguese forms concerned are both the plural and singular, the Etymon has two WordForm instances, one for each grammatical number. Note in particular the association of grammaticalCase and inflectionType attributes with the WordForm of the Etymon via a GrammaticalInformation instance. In a comprehensive lexicon of Portuguese that contained such etymological information for a sufficient number of lexical entries, it would be possible, by contrasting the contents of the WordForm in the LexicalEntry with the Etymon, to appreciate the following language-wide phenomena: 1) Portuguese lost grammatical case; 2) the vast majority of its nouns come from the VL accusative case; 3) where the VL singular (accusative) ending is *-tiōnem*, the Portuguese form is written “-ção” and pronounced [sẽw]; where the VL plural (accusative) ending is *-tiōnes*, the Portuguese form is written “-ções” and pronounced [sõj̃s].

NOTE This etymology can be further articulated by adding the phonological process types for each stage of the diachrony. This can be done in the model by adding the appropriate data category defined in Annex B to the value of type on the EtymLink for the given stages.

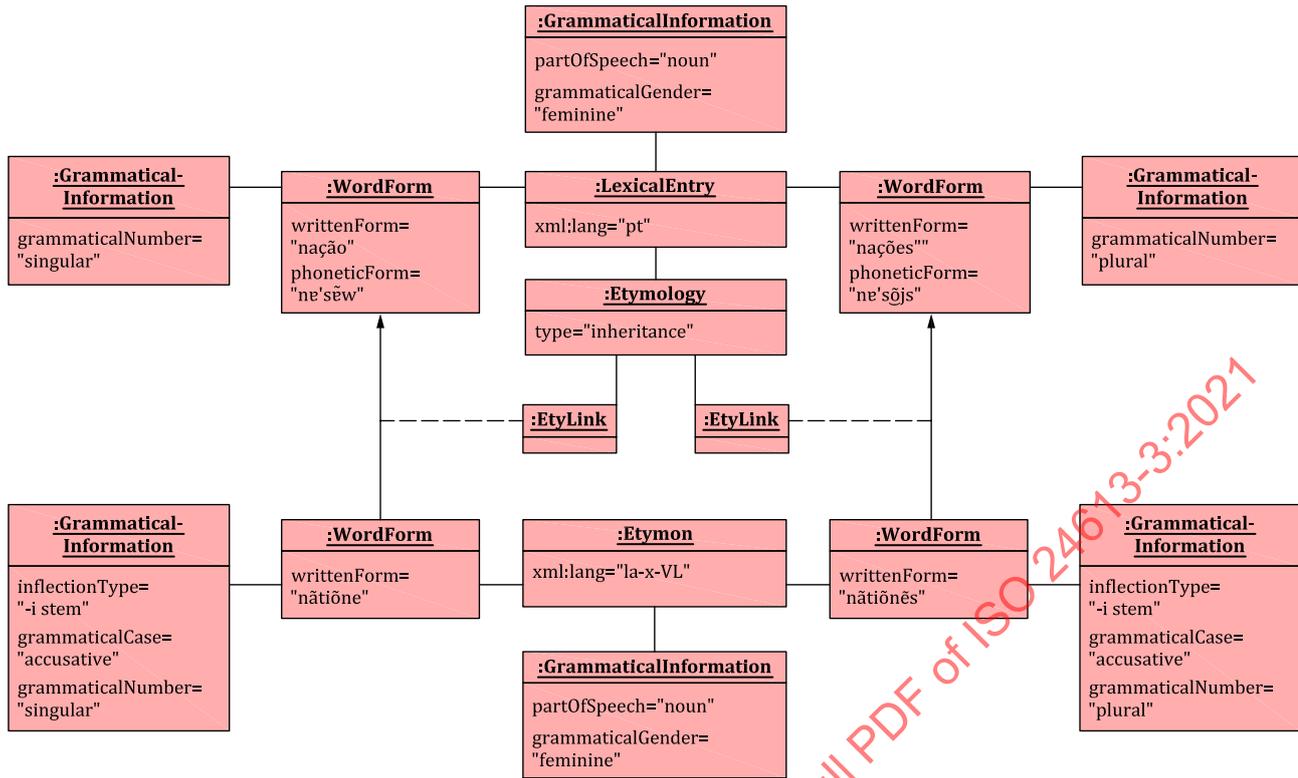


Figure A.3 — Diagram of inheritance of Portuguese inflections

#### A.4 Example of metaphorical semantic shift

The example in Figure A.4 from Mixtepec-Mixtec shows the derivation of one sense from another via metaphor. In this case, the word translating to ‘kidney’ was derived from a metaphorical extension of the word meaning ‘bean’ *ntuchi*. The Etymology is attached to the kidney sense of the word *ntuchi*. The EtymLink instance contains the attributes source and target with the id values of the respective components specifying the directionality of the process. As a defining feature of the process of metaphor is that there has to be a change in the semantic domain between the source and target sense, each sense has a domain specified in the subject field. The CrossREF class defined in ISO 24613-2 is used to specify a URI corresponding to the dbpedia entry for each sense.

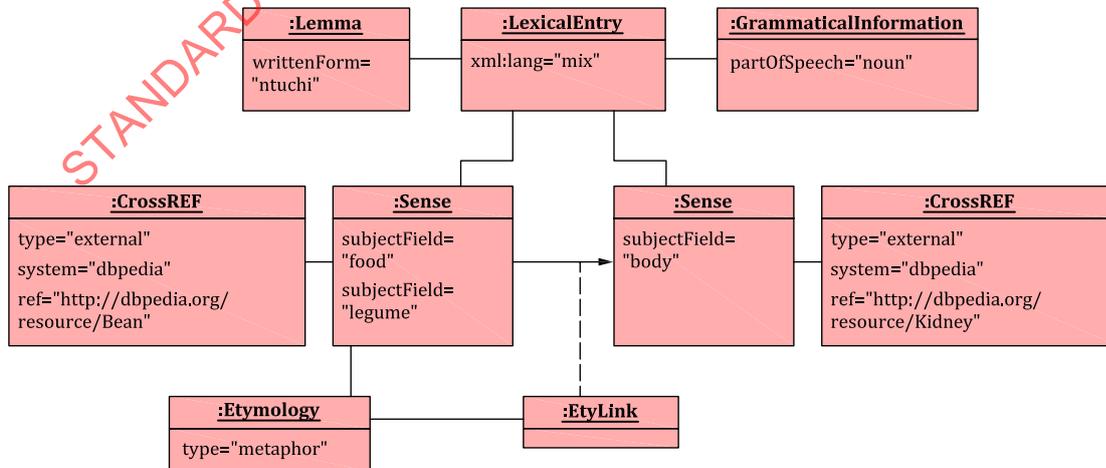


Figure A.4 — Diagram of metaphorical sense shift in Mixtepec-Mixtec

## A.5 Example of borrowing and compounding

In [Figure A.5](#), an example of a complex etymological borrowing is given, the borrowing of *pamplemousse* ‘grapefruit’ by French from Dutch (see References [4] and [18]). In addition to the borrowing, the etymology shows the diachrony of the word formation process that occurred within Dutch in which the compound *pompelmousse* was formed from the etymons *pompel* and *limoes*, with the former subsequently being borrowed into French. A CrossREF (see ISO 24613-2) instance is used to specify that the borrowed Etymon is a compound in Dutch and that its components are ordered. The components of the compound are represented as Etymons containing the salient forms as Lemma.

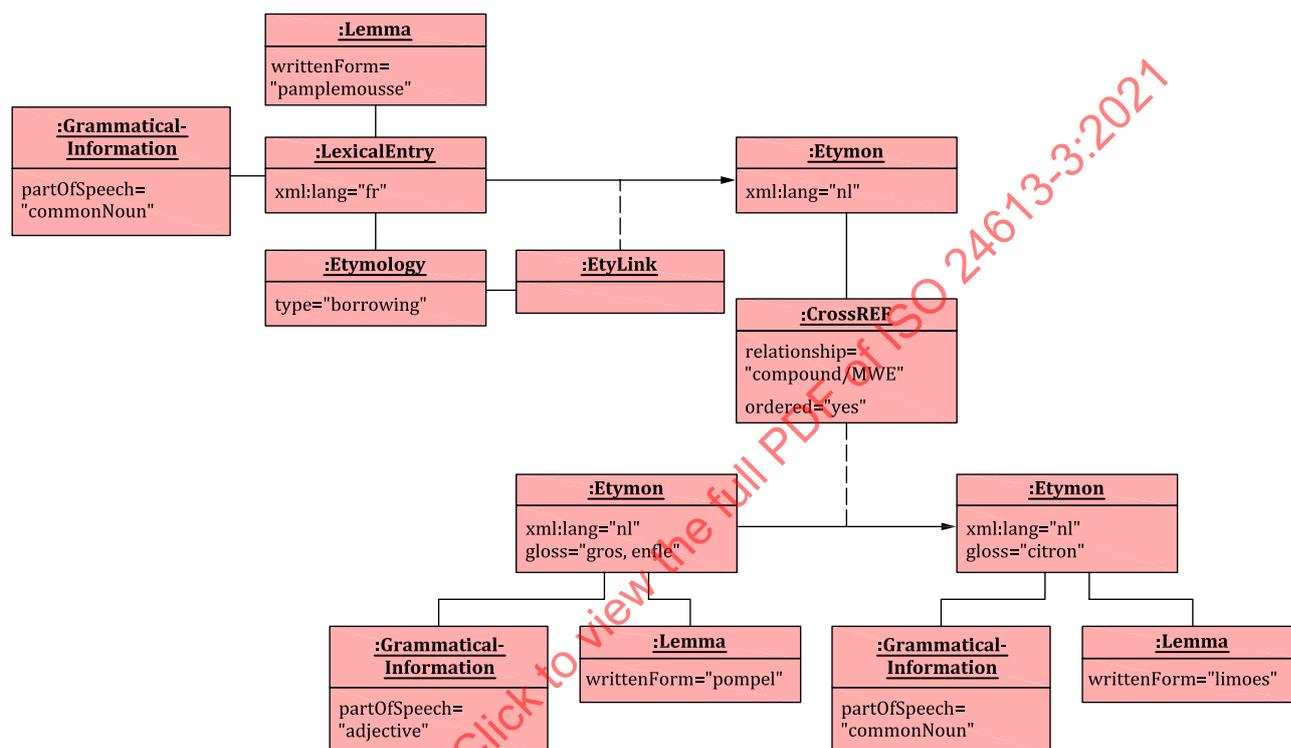


Figure A.5 — Diagram of borrowing and compounding of French *pamplemousse*

## A.6 Example of use of temporal information

The example in [Figure A.6](#) shows the encoding of the etymology of the French lexical item *chef* with a focus on the diachronic stages and time frame of its phonetic development (see Reference [13]). Herein at the level of Etymology, the process specifying the fact that *chef* was inherited from Latin is encoded in the attribute type which is given the value “inheritance”. Within the Etymology, there is a series of Etymon items containing the temporal attributes notBefore and notAfter which are used to specify the spans of time, and the sequential ordering asserted is that a given form of the item would have been in use. Note also the ordering of Etymon items via the use of prev and next attributes on the appropriate EtyLink individuals.

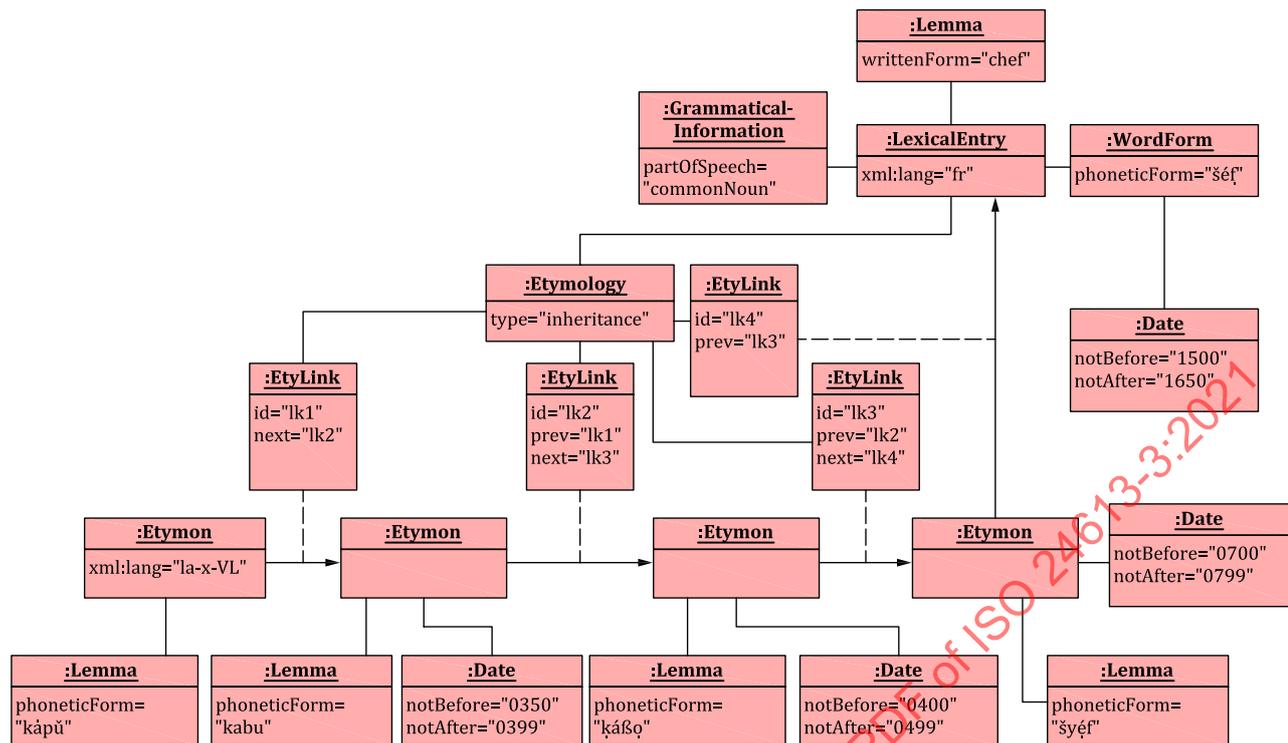


Figure A.6 — Diagram of multi-stage phonological changes of French *chef*

### A.7 CognateSet with Bibliography

Figure A.7 shows an example of a cognate set as taken from an actual etymological dictionary source (see Reference [19]). A number of cognate forms in related languages are given to compare to the lexical item *girl*. Although the LexicalEntry is inherently part of the CognateSet, this relationship is assumed and thus it is not represented as Cognate in the encoding.

A major point of interest in this entry is the fact that the lexical entry *girl*, whose contemporary English sense is given in the Definition for the main entry, was used early on to denote a young person regardless of gender, and therefore became subject to a semantic shift over its long period of use. Each Cognate has an associated Sense and a Definition instance which together describe related lexical entries in Germanic languages, all of which are hypothesised as having been inherited from a common source.

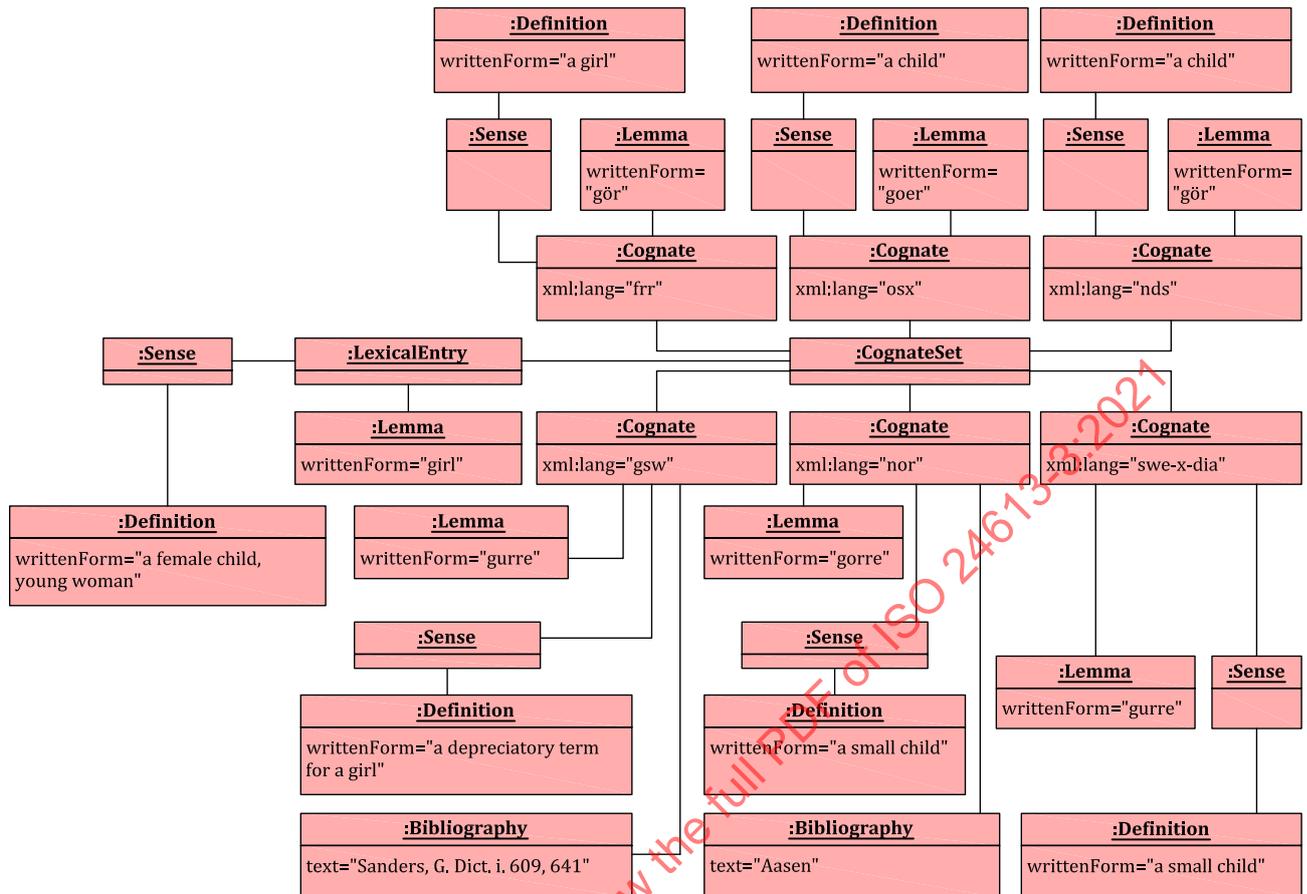


Figure A.7 — Diagram of cognate set for English entry *girl*

### A.8 Comprehensive etymological example

In order to demonstrate some of the expressivity of the current model and its potential for dealing with realistic examples of etymological data, the etymological information of part of a single entry, namely the entry for the Latin word *amārus* ‘bitter’ in [Figure A.8](#)<sup>[6]</sup>, is modelled in [Figures A.9](#) and [A.10](#).

The entry consists of a listing of two of the word’s derivatives along with a series of its etymons and cognates. The entry also contains a paragraph of textual commentary on the etymology of *amārus* (that is the paragraph starting with ‘The suffix –arus’...); this commentary is not modelled as structured data in the example. The original text of the entry is shown in [Figure A.8](#).

**amārus** 'bitter' [adj. *o/ā*] (Pl.+)

Derivatives: *amāritūdō* 'bitterness' (Varro+), *amāror* [m.] 'bitter taste' (Lucr.+).

Plt. *\*o/am-?*

PIE *\*h<sub>2</sub>h<sub>3</sub>m-ro-?* IE cognates: Skt. *amlá-* 'sour, acid', Olc. *apr.* 'sharp, cold', OE *ampre* 'sour one', MDu. *amper* 'bitter, sour' < PGm. *\*am(p)ra-* 'sour'; ? OIr. *om* 'raw', W. *of* possibly < *\*h<sub>2</sub>h<sub>3</sub>-emo-*, Skt. *āmá-* [adj.] 'raw, uncooked', Gr. *ώμός* 'raw', Arm. *howm* < *\*h<sub>2</sub>eh<sub>3</sub>mo-*.

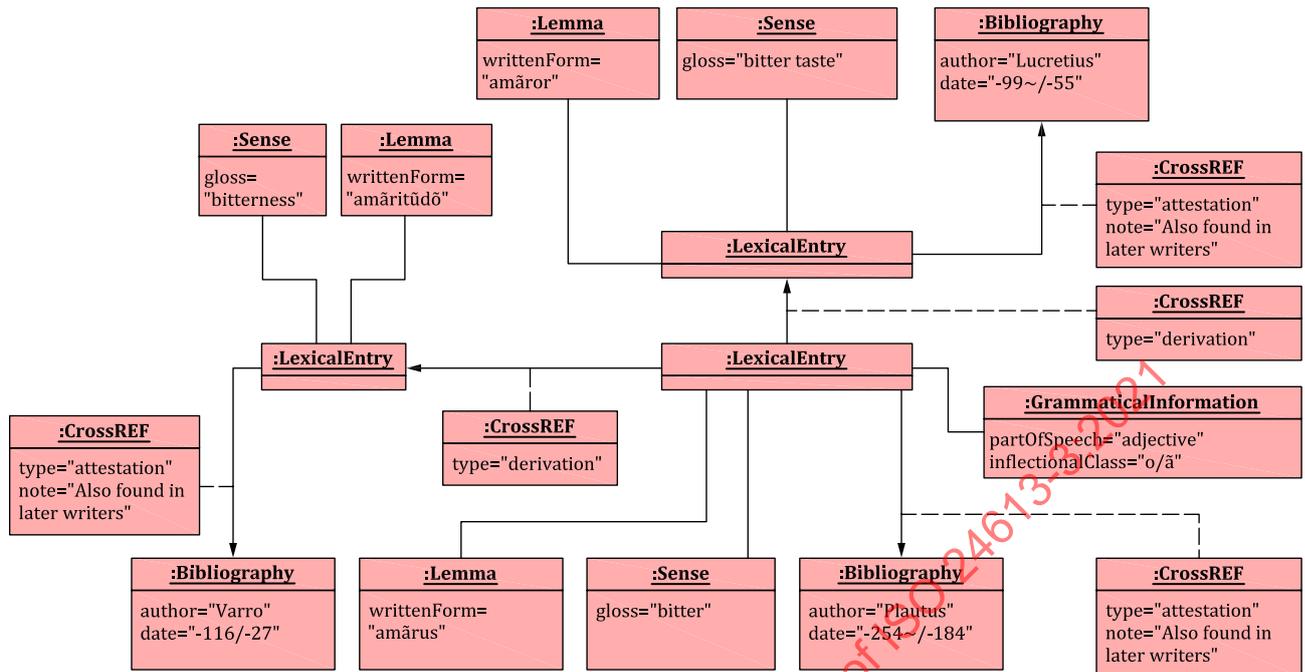
The suffix *-ārus* is only attested as a primary derivative in *avārus* 'greedy' to *aveo*, but there is no verb to which *amārus* might belong. Kortlandt 1980a connects it with Greek and Armenian words for 'raw' and reconstructs *\*HHm-*, but Schrijver 1991: 77 argues that the semantics are not convincing. The connection with Skt. and Gm. 'sour' is semantically more convincing; these might go back to PIE *\*h<sub>2</sub>em-ro-*. In that case, Latin must have replaced the suffix *\*-ro-* by *\*-āro-*. Thus, some details remain unclear. It is also possible to reconstruct earlier *\*om-* 'raw' for Italo-Celtic, and posit unrounding of *\*#omV-* > *amV-* as per Rasmussen 1993:181.

Bibl.: WH I: 35, EM 25, IEW 777f., Kortlandt 1980a, Schrijver 1991: 43, Heidermanns 1993: 99f.

SOURCE De Vaan (2008)<sup>[6]</sup>, reproduced with permission.

### Figure A.8 — Sample entry of Latin *amārus*

The modelling of this entry is presented in two parts. First is a diagram (Figure A.9) which details the LMF modelling of the first two lines of the entry (shown in Figure A.8): the lines containing the lemma, meaning and part of speech of the lexical entry as well as its derivatives. In the LMF modelling of the entry, the lexical entry *amārus* is linked to its derivatives using the CrossREF element (defined in ISO 24613-2). CrossREF is used as a means of modelling the attestation of a given derivation or sense of the entry. The CrossREF individuals in question are typed as attestations using the attribute type, and with any additional information (in this case the presence of the plus sign after the bibliographic citation in Reference [6] means that the linguistic usage in question can also be found in later sources) added as text using the note attribute. In this instance, the bibliographic element refers to a single author (and therefore to their corpus of works) and the date attribute refers to the whole period encompassing their lifespan.



SOURCE De Vaan (2008)<sup>[6]</sup>, reproduced with permission.

**Figure A.9 — Part 1 of the entry *amārus* — Lemma, derivatives and bibliographies**

The second part of the example (Figure A.10) details the modelling of the etymological information itself (for ease of reading the elements of each of the two cognate sets are coloured using two different colours, cyan and purple). The Cognates are each shown as belonging to one of two CognateSet elements and the entire etymology is attributed to the source by associating the Bibliography with the Etymology element for the LexicalEntry. Two other Etymology elements are used to describe etymologies for Cognates included within the CognateSet. The attribute rootType is used to indicate that the form of the etymon is reconstructed, and the attribute status to indicate a level of uncertainty.



## Annex B (normative)

### Data categories for etymology description

This annex contains a structured list of categories that shall be used as possible values for describing etymological processes, including sound changes as well as relevant lexical subfeatures. It describes, for each category, its name, a camel case representation, a conceptual domain (for complex data categories), a definition (with its source(s)), one or several examples (with their source(s) when applicable), as well as some usage notes where needed.

Where applicable, the camel case form should be used in utilizing the features as a value for the attribute type in declaring the process of an Etymology or EtyLink.

**NOTE** These features are not exhaustive and do not account for all the possible theoretical super- and subtypes and variant definitions or descriptions.

#### **Borrowing**

**Camel case code:** borrowing

**Definition:** The process in which a lexical item, phrase or other linguistic feature from a foreign language or dialect is introduced into a given language or dialect.

**Source:** Reference [4].

**Note:** The lexical items which result from borrowing are loanwords.

**Example:** Russian *картошка* is a loanword from High German *Kartoffel*.

**Alternate term:** loaning, importing, transferring, copying

#### **Calque**

**Camel case code:** calque

**Definition:** A loanword in which only the meaning is borrowed with the term being literally translated into the borrowing language.

**Source:** Reference [5].

**Example:** Spanish *rascacielos* (rasca 'scratch', 'scrape' + cielos 'sky') from English *skyscraper*

**Alternate term:** loan translation, semantic loan

#### **Inheritance**

**Camel case code:** inheritance

**Definition:** Inheritance, although often not regarded as an etymological process in itself, is used to identify situations where a lexical item is known to, or is presumed to, have been inherited from a predecessor or “parent” language(s). Items inherited from a predecessor language often undergo any number of different separate etymological processes over the course of their diachrony.

**Source:** Reference [4].

**Example:** Sardinian *semper* was inherited from Latin *semper*.

## **Compounding**

**Camel case code:** compounding

**Definition:** Compounding is the process of creating a new lexical item from multiple pre-existing ones. While compounding is often described in terms of synchronic linguistics, particularly when a compound has undergone a process of lexicalization, compounds often are created via one or more cognitive semantic processes such as metaphor and metonymy and thus can be classified and analysed etymologically according to the particular mechanisms underlying the compound's creation on the semantic level.

**Source:** Reference [4].

**Example:** Chinese: 父母 'parents' is a compound of: 父 'father' and 母 'mother'.

## **Derivation**

**Camel case code:** derivation

**Definition:** Derivation is the process by which new lexical items or inflectable stems are formed by the addition of an affix. The derived lexical items are often of a different class than the original, in such cases they will take the inflectional affixes of the new word class.

**Example:** Silacayoapan Mixtec uses the nominalizing prefix "ñā-" to create new nouns from other parts of speech: *tū hún* 'to lie' > *ñā-tū hún* 'a lie'.

## **Back formation**

**Camel case code:** backFormation

**Definition:** Back formation is generally a type of folk etymology in which either a word is modified morphologically based on mistaken assumptions of its morphological composition, or in which a new word of one class is created based on root components of another.

**Source:** Adapted from Reference [5].

**Example:** English *cherry*, a loanword from Old French *cheris* (Modern French *cerise*) was assumed to be plural because of the *-s*, thus the singular form in English was adapted to *cherry* and the plural *cherries*. An example of the second type of back formation can be found in the English verb *to conversate* from the noun *conversation*.

**Alternative term:** retrograde formation

## **Portmanteau**

**Camel case code:** portmanteau

**Definition:** Generally the difference between compounding and portmanteau is that compounds do not involve the contraction of their component forms. In this process, like in compounding, portmanteau items are created via one or more cognitive semantic processes such as metaphor and metonymy and thus can be classified and analysed etymologically according to the particular mechanisms underlying the compound's creation on the semantic level.

**Example:** The word *Brexit* which is the noun referring to the exit of Great Britain from the European Union is a portmanteau of *Britain* + *exit*.

**Alternative term:** blending, blend, fusion

**Note:** It should be noted that while there is theoretical overlap between the ideas, the term "blending" with regards to this particular word formation process should not be attributed to the conceptual blending theory developed in References [8] and [9].

**Narrowing**

**Camel case code:** narrowing

**Definition:** The process of semantic change where over time a term is used to refer to a more restricted aspect of its original meaning.

**Example:** In Old English the ancestor to the modern English word *meat* originally referred to food in general but then its sense narrowed to refer only to 'flesh used as food'.

**Alternate term:** hyponymic shift

**Widening**

**Camel case code:** widening

**Definition:** The process of semantic change where over time a term is used to refer to a more general aspect of its original meaning.

**Example:** The English word *camp* (n) originally pertained to military specific location with open, level space but has generalized to non-military senses by the 1580's.

**Alternate term:** broadening, extension, generalization, hypernymic shift

**Amelioration**

**Camel case code:** amelioration

**Definition:** The process by which the sense of a word which formerly has some negative connotation changes to a positive or at least a reduced degree of negativity.

**Example:** The English word *knight* originally meant a boy who was a servant or attendant (c.f., the German word *knecht*) but has become a word to designate a rank of nobility.

**Pejoration**

**Camel case code:** pejoration

**Definition:** The process by which the sense of a word which formerly has a neutral or positive connotation changes to a negative or at least a reduced degree of positivity.

**Example:** The word *puttana*, which in its original Latin form just meant 'young woman', in Italian means 'prostitute' or 'slut'.

**Alternate term:** derogation, worsening

**Metaphor**

**Camel case code:** metaphor

**Definition:** The process of metaphor as a vehicle of lexical change is based in cognitive processes by which one concept from domain X is understood in terms of another concept from domain Y. This is mirrored lexically, where one concept in a given domain is described or referred to using lexical items (and/or vocabulary) from another. The driving forces for this process are ontological correspondences between the attributes or internal structures of entities in the source domain with those of the target domain.

**Source:** Adapted from References [14] and [15].

**Note:** Metaphor in the context of etymology is differentiated from other spontaneous uses in language by the need for the term to be adopted (lexicalized) within the lexicon.

**Note:** Metaphor (where occurring not in the context of compounding or portmanteau) gives rise to lexical polysemy.

**Example:** Mixtepec-Mixtec

*sa'a* 'foot (person or animal)' > 'bottom part (of an object)'

### **Metonymy**

**Camel case code:** metonymy

**Definition:** Like metaphor, metonymy is another process based in human cognition which is reflected in language and leads to synchronic lexical polysemy. In metonymy, one vehicle entity provides mental access to a target entity within a single domain by highlighting different aspects of part-whole relationships (meronymy).

**Source:** Adapted from Reference [12].

**Example:** Mixtepec-Mixtec

- I. *nuu* 'face (of a person or animal)' > 'the front (of a person or animal)'
- II. *nuu* 'the front (of a person or animal)' > 'in front (of a person or animal)'

### **Grammaticalization**

**Camel case code:** grammaticalization

**Definition:** Diachronically, grammaticalization is a unidirectional process through which lexical items over time become more grammatical in nature. Central to grammaticalization is a continuum of bondedness in which on one end are lexical items which are semantically richer and which occur in relatively free syntactic constructions and on the other are items which are less semantically rich and which may occur in much more restricted (morpho-)syntactic contexts. At the extreme of the latter end of the continuum items may disappear altogether.

**Source:** Adapted from References [16] and [20].

**Note:** Complex category which may involve various combinations of discourse, morpho-syntactic, semantic and possibly phonological changes to a lexical item.

**Example:** Old English to Modern English

*a:n* 'one' > *a(n)* indefinite article

### **Folk Etymology**

**Camel case code:** folkEtymology

**Definition:** The process by which some component of a lexical item or phrase is mentally reconfigured (rebracketed) due to an incorrect interpretation of its original form and content by speakers. The result of this mental reconfiguration may or may not result in changes to the item's form. Often these incorrect or popular re-interpretations (or re-analyses) are motivated by the phonotactic similarity of the item (or parts of it) to a more prominent or culturally familiar concept.

**Source:** Adapted from Reference [17].

**Example:** Norfolk place-name *Crosthwick* changed from the historical compound of *cross* + *thwaite* (derived from Old Norse *kross* and *pveit*). The change is attributed to being triggered by the similarity to the name of a neighbouring village called *Postwick*.

**Alternate term:** popular etymology, pseudo-etymology, analogical reformation