

PROJET
FINAL

NORME
INTERNATIONALE

ISO/FDIS
24613-3

ISO/TC 37/SC 4

Secrétariat: KATS

Début de vote:
2020-12-24

Vote clos le:
2021-02-18

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

Partie 3: Extension étymologique

iTeh **STANDARD PREVIEW**
(standards.itih.ai)
*Language resource management — Lexical markup framework
(LMF) —
Part 3: Etymological extension*

[ISO/FDIS 24613-3](#)

<https://standards.itih.ai/catalog/standards/sist/86a73b0d-88c5-425c-b359-ea1d68ba26c9/iso-fdis-24613-3>

LES DESTINATAIRES DU PRÉSENT PROJET SONT INVITÉS À PRÉSENTER, AVEC LEURS OBSERVATIONS, NOTIFICATION DES DROITS DE PROPRIÉTÉ DONT ILS AURAIENT ÉVENTUELLEMENT CONNAISSANCE ET À FOURNIR UNE DOCUMENTATION EXPLICATIVE.

OUTRE LE FAIT D'ÊTRE EXAMINÉS POUR ÉTABLIR S'ILS SONT ACCEPTABLES À DES FINS INDUSTRIELLES, TECHNOLOGIQUES ET COMMERCIALES, AINSI QUE DU POINT DE VUE DES UTILISATEURS, LES PROJETS DE NORMES INTERNATIONALES DOIVENT PARFOIS ÊTRE CONSIDÉRÉS DU POINT DE VUE DE LEUR POSSIBILITÉ DE DEVENIR DES NORMES POUVANT SERVIR DE RÉFÉRENCE DANS LA RÉGLEMENTATION NATIONALE.



Numéro de référence
ISO/FDIS 24613-3:2020(F)

© ISO 2020

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/FDIS 24613-3
[https://standards.iteh.ai/catalog/standards/sist/86a73b0d-88c5-425c-b359-
ea1d68ba26c9/iso-fdis-24613-3](https://standards.iteh.ai/catalog/standards/sist/86a73b0d-88c5-425c-b359-ea1d68ba26c9/iso-fdis-24613-3)



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO 2020

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos	iv
1 Domaine d'application	1
2 Références normatives	1
3 Termes et définitions	1
4 Extension étymologique de LMF	2
4.1 Classes Cognate et Etymon.....	2
4.2 Classe Etymologizable.....	2
4.3 Classes Etymology et EtyLink.....	3
4.4 Classe CognateSet.....	4
4.5 Classe Date.....	4
4.6 Classe Gloss.....	4
Annexe A (informative) Exemples de typologies étymologiques possibles	6
Annexe B (normative) Catégories de données pour la description étymologique	15
Bibliographie	22

iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO/FDIS 24613-3

<https://standards.iteh.ai/catalog/standards/sist/86a73b0d-88c5-425c-b359-ea1d68ba26c9/iso-fdis-24613-3>

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (IEC) en ce qui concerne la normalisation électrotechnique.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives).

L'attention est attirée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la nature volontaire des normes, la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir le lien suivant: www.iso.org/iso/fr/avant-propos.

Le présent document a été élaboré par le comité technique ISO/TC 37, *Langage et terminologie*, sous-comité SC 4, *Gestion des ressources linguistiques*.

Cette première édition de l'ISO 24613-3, conjointement avec l'ISO 24613-1, l'ISO 24613-2, l'ISO 24613-4 et l'ISO 24613-5, annule et remplace l'ISO 24613:2008, qui a fait l'objet d'une révision technique.

Les principales modifications par rapport à l'édition précédente sont les suivantes:

- révision complète du contenu et de sa subdivision en plusieurs parties.

Une liste de toutes les parties de la série ISO 24613 se trouve sur le site web de l'ISO.

Il convient que l'utilisateur adresse tout retour d'information ou toute question concernant le présent document à l'organisme national de normalisation de son pays. Une liste exhaustive desdits organismes se trouve à l'adresse www.iso.org/fr/members.html.

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

Partie 3: Extension étymologique

1 Domaine d'application

Le présent document décrit une extension de l'ISO 24613-1 et de l'ISO 24613-2 et facilite l'élaboration des descriptions détaillées de phénomènes étymologiques courants et/ou d'informations diachroniques par rapport aux entrées lexicales des lexiques numériques et/ou rétronumérisés. Il fournit à la fois un métamodèle pour une extension de ce type ainsi que les catégories de données pertinentes.

2 Références normatives

Les documents suivants sont cités dans le texte de sorte qu'ils constituent, pour tout ou partie de leur contenu, des exigences du présent document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document de référence s'applique (y compris les éventuels amendements).

ISO 8601-1, *Date et heure — Représentations pour l'échange d'information — Partie 1: Règles de base*

ISO 8601-2, *Date et heure — Représentations pour l'échange d'information — Partie 2: Extensions*

ISO 24613-1, *Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) — Partie 1: Modèle de base*

ISO 24613-2, *Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) — Partie 2: Modèle de dictionnaire lisible par ordinateur (MRD)*

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions de l'ISO 24613-1 ainsi que les suivants, s'appliquent.

L'ISO et l'IEC tiennent à jour des bases de données terminologiques destinées à être utilisées en normalisation, consultables aux adresses suivantes:

- ISO Online browsing platform: disponible à l'adresse <https://www.iso.org/obp>;
- IEC Electropedia: disponible à l'adresse <http://www.electropedia.org/>.

3.1

cognat

dans une langue apparentée, forme qui partage une origine étymologique commune avec une forme dans la langue du lexique

3.2

étymologisable

qui remplit les conditions requises pour avoir une *étymologie* (3.3)

Note 1 à l'article: Le terme «étymologisable» se rapporte à une catégorie d'éléments et d'usages lexicaux (englobant par exemple des entrées lexicales, des sens et des mots-formes).

**3.3
étymologie**

origine et développement historique de tout aspect d'un élément lexical donné

**3.4
étymon**

entrée lexicale dont découle une autre entrée lexicale

Note 1 à l'article: Un étymon peut également être une phase antérieure d'un élément lexical.

**3.5
onomasiologie**

étude sémantique des mots qui, en partant d'un concept donné, examine les différents éléments lexicaux utilisés dans une ou plusieurs langues pour se référer à ce concept

4 Extension étymologique de LMF

NOTE Voir l'Annexe A pour obtenir des exemples de typologies étymologiques possibles.

4.1 Classes Cognate et Etymon

Cognate et Etymon sont définies comme des sous-classes de la classe LexicalEntry du module de base LMF (voir Figure 1)¹⁾. Les deux classes définissent des entrées lexicales qui ont été ajoutées à une ressource lexicale dans le but de décrire les étymologies d'une ou plusieurs autres entrées lexicales. Aux instances Etymon ou Cognate peut être assignée une langue qui est différente de celle du lexique dans son ensemble (cette langue est spécifiée dans la classe LexiconInformation décrite dans l'ISO 24613-1).

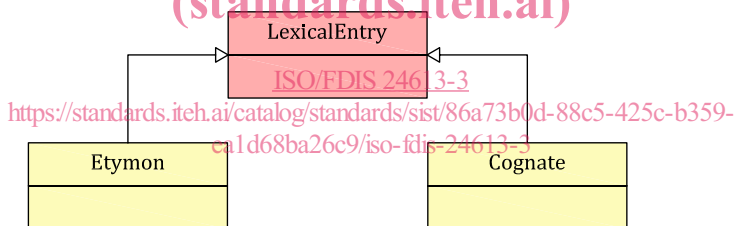


Figure 1 — Sous-classes Cognate et Etymon de la classe LexicalEntry

Les individus des sous-classes Etymon et Cognate doivent être en relation d'agrégation avec au moins un individu de type EtyLink (voir 4.3). Lors de la description des étymologies, il est nécessaire de traiter, dans certains cas, des instances LexicalEntry (et, de ce fait, également des instances de relation de sous-classe d'Etymon et de Cognate) qui sont des radicaux, et en particulier des radicaux reconstruits. Dans ces cas, le fait d'être un radical et le type du radical en question doivent être spécifiés en utilisant l'attribut rootType. Dans le cas de radicaux reconstruits ou d'autres mots-formes, l'attribut status sert à associer l'élément à une description écrite de sa probabilité d'avoir été utilisé (voir l'exemple de A.8). Voir le Tableau 1 pour obtenir une liste d'attributs à utiliser avec ces deux sous-classes.

4.2 Classe Etymologizable

La classe Etymologizable permet de désigner l'ensemble des éléments linguistiques qui peuvent avoir des étymologies. En définissant une seule classe englobant tous les éléments «étymologisables» de ce type, les classes d'éléments qui peuvent avoir des étymologies pourront être facilement étendues en fonction des besoins. Les classes suivantes sont des sous-types de la classe Etymologizable (voir Figure 2): LexicalEntry, Sense, Form et CognateSet (voir 4.4).

1) Les schémas du présent document utilisent les codes de couleurs suivants : les classes en jaune sont introduites dans le présent document, alors que les classes en rose l'avaient été précédemment dans l'ISO 24613-1 et l'ISO 24613-2.

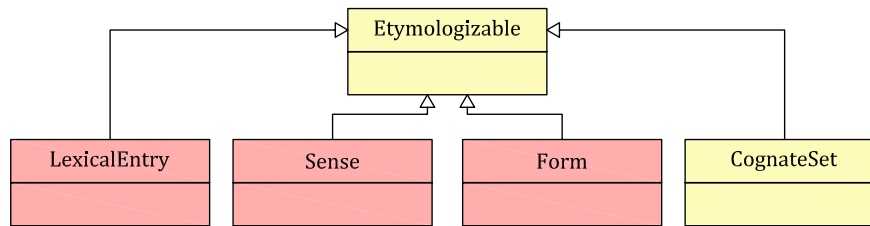


Figure 2 — Classe Etymologizable et ses sous-classes

4.3 Classes Etymology et EtyLink

La classe Etymology permet de décrire l'étymologie d'un élément linguistique. Plus spécifiquement, elle permet de décrire les éléments linguistiques qui sont des sous-classes de la classe Etymologizable. Le type ou les types de processus étymologique impliqués dans une étymologie donnée peuvent être spécifiés en utilisant l'attribut type, et aussi potentiellement l'attribut subtype (si le type de l'étymologie peut également être précisé). Les valeurs possibles pour les attributs type et subtype peuvent varier en fonction de l'approche théorique adoptée par le compilateur d'une ressource et/ou de l'orientation linguistique ou éditoriale de la ressource en question. L'imbrication des instances Etymology permet de combiner les processus étymologiques à décrire. Les processus étymologiques qui doivent être utilisés comme valeurs pour les attributs type/subtype peuvent par exemple être: *emprunt, héritage; formation des mots: mots composés, dérivation; glissements de sens: restriction, élargissement, amélioration, péjoration, métaphore, métonymie; processus phonétiques/phonologiques: assimilation, dissimilation, épenthèse, métathèse, durcissement, affaiblissement, etc.* La liste des catégories de données fournie à l'[Annexe B](#) doit être utilisée en complément des classes appropriées. Un type peut également être spécifié pour des liens individuels entre deux éléments au sein d'une étymologie (voir la description d'EtyLink ci-dessous). Comme une instance Etymology peut être extraite d'une source externe, elle peut être associée à une instance Bibliography, qui doit être définie conformément à l'ISO 24613-2 (voir [Figure 3](#)).

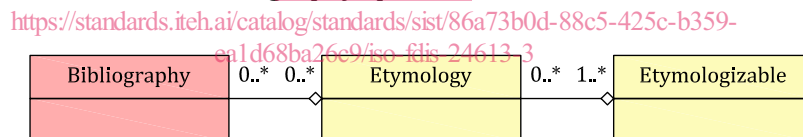


Figure 3 — Classe Etymology

Les instances Etymology sont associées à une ou plusieurs instances EtyLink, chacune d'elles représentant une seule phase ou étape dans l'étymologie d'un élément lexical donné (voir [Figure 4](#)). EtyLink sert à rattacher des individus appartenant aux sous-classes Etymologizable. EtyLink est une sous-classe de la classe CrossREF définie dans l'ISO 24613-1. L'utilisation de CrossREF impose d'affecter des attributs id aux objets cibles représentant le contenu lexical spécifique. L'utilisation de l'attribut id, en tant que cible, sur un individu de la classe Etymologizable permet de modéliser un tri temporel séquentiel générique d'éléments multiples, avec les attributs prev et next. Des instances de la classe EtyLink peuvent également spécifier des relations temporelles supplémentaires en utilisant divers attributs temporels associés à la source et à la cible de chaque instance EtyLink.

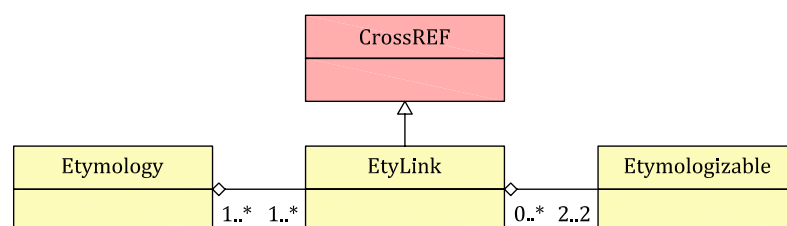


Figure 4 — EtyLink

Les individus des classes Etymon et Cognate (sous-types de LexicalEntry) doivent être associés à au moins un individu EtyLink. Voir le [Tableau 1](#) pour obtenir une liste d'attributs à utiliser avec les classes Etymology et EtyLink.

4.4 Classe CognateSet

La classe CognateSet (voir [Figure 5](#)) est un conteneur d'ensembles d'un ou plusieurs éléments Cognate et de zéro ou plusieurs éléments Bibliography (voir l'ISO 24613-2). La classe CognateSet est une construction liée à l'onomasiologie. Elle contient des éléments issus de langages apparentés à celui d'une classe LexicalEntry donnée (et donc liés par la relation de sous-classe de tout Etymon ou Cognate donné) et qui ont été collectés dans le but de démontrer des similarités ou des dissimilarités linguistiques de type essentiel. L'utilisation d'une classe CognateSet implique que les instances LexicalEntry (et donc Etymon et Cognate) qu'elle contient partagent une source étymologique.

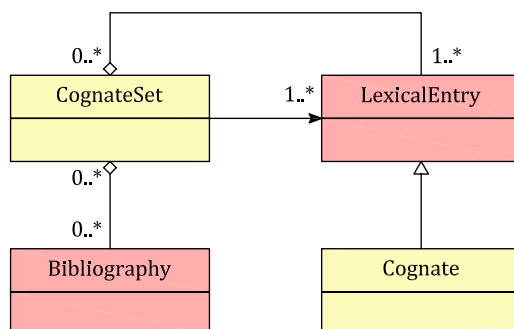


Figure 5 — CognateSet
(standards.iteh.ai)

4.5 Classe Date

ISO/FDIS 24613-3

<https://standards.iteh.ai/catalog/standards/sist/86a73b0d-88c5-425c-b359-1168022c5132-41355>

Les composants d'une classe LexicalEntry et de ses sous-classes doivent être associés à une date spécifique en utilisant la classe Date. De plus, cette classe Date permet de spécifier un nombre de degrés de précision. Une année précise, et potentiellement un mois et un jour donnés, doivent être spécifiés en utilisant l'attribut date et une date approximative avec l'attribut circa. Sur un laps de temps donné avec différents niveaux de spécificité, il est possible d'utiliser un ou plusieurs attributs de date. Lorsque le laps de temps est connu (ou évalué), ses bornes inférieure et supérieure peuvent être précisées en utilisant respectivement notBefore et notAfter. Pour les formats de date et d'heure, l'ISO 8601-1 et l'ISO 8601-2 doivent être utilisées.

4.6 Classe Gloss

La classe Gloss (voir [Figure 6](#)) représente une description textuelle de la signification d'un mot ou d'un syntagme qui est destiné à être compris par l'homme. Les individus de cette classe peuvent représenter des paraphrases ou des synonymes qui peuvent être rédigés dans la langue de l'entrée ou dans une autre langue. Voir le [Tableau 1](#) pour obtenir une liste d'attributs à utiliser avec cette classe.

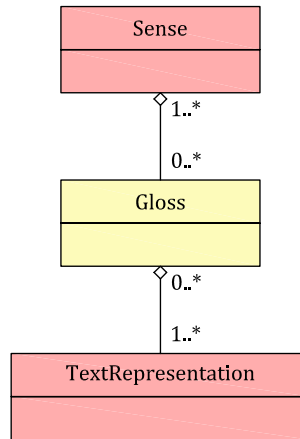


Figure 6 — Gloss

Tableau 1 — Exemple d’affectation de classes

Nom de la classe	Exemple d’attributs
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 ISO/FDIS 24613-3

<https://standards.iteh.ai/catalog/standards/sis/86a73b0d-88c5-425c-b359-ea1d68ba26c9/iso-fdis-24613-3>

Annexe A (informative)

Exemples de typologies étymologiques possibles

A.1 Exemple d'héritage simple

L'exemple de la [Figure A.1](#) décrit l'héritage d'une entrée lexicale provenant d'une langue parente, dans ce cas l'adverbe sarde *sempe*, qui provient du mot latin *semper*. Le mot sarde est lié à une seule instance Etymology qui est associée au type *inheritance* (voir l'[Annexe B](#) pour une définition de ce type). L'instance Etymology est ensuite associée à un individu de type EtymLink qui représente le processus d'évolution de l'étymon latin jusqu'à l'entrée lexicale sarde.

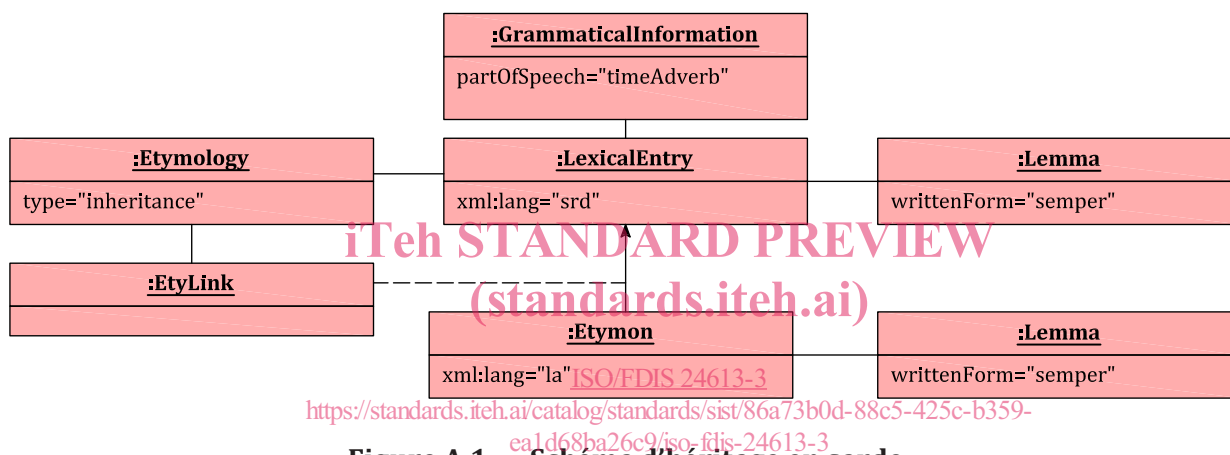


Figure A.1 — Schéma d'héritage en sarde

A.2 Exemple de processus étymologique diachronique (héritage avec changement phonologique)

Dans l'exemple suivant, l'évolution du mot signifiant «vin» (ipa: [veŋ]) dans la variété de vins italiens Emiliano est suivie en utilisant une série d'étymons qui sont triés et mutuellement liés en utilisant des instances EtymLink. Ce processus mène du latin vulgaire *vinu* jusqu'au prédécesseur immédiat du mot sous sa forme actuelle. Ces liens individuels sont accessibles via un individu Etymology (avec le type *inheritance*) qui représente l'historique de LexicalEntry. Le tri des instances EtymLink est effectué au moyen des deux attributs prev et next. Ces attributs ne sont pas représentés sur la [Figure A.2](#) pour des raisons de lisibilité, mais leur utilisation est indiquée sur la [Figure A.6](#), pour l'exemple de [A.6](#), et sur la [Figure A.10](#), pour l'exemple de [A.8](#).

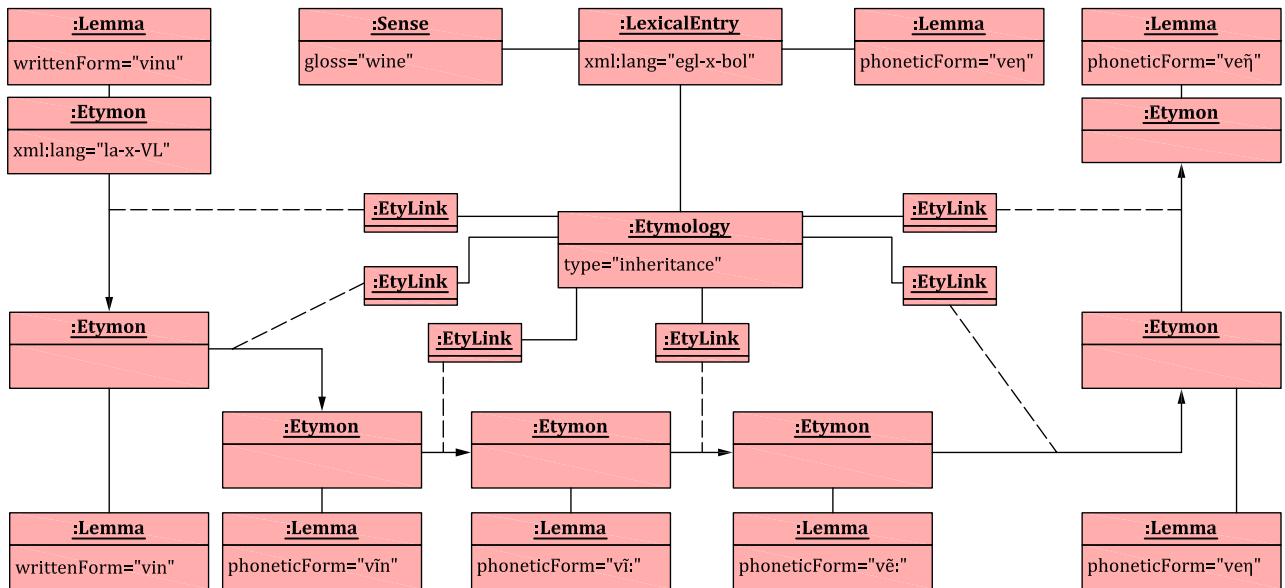


Figure A.2 — Schéma d'un héritage en plusieurs étapes et d'un changement phonologique en bolonais

A.3 Exemple d'héritage de mot-forme

La dérivation des formes singulière et plurielle du nom portugais *nação* 'nation', *nação* (sg) et *nações* (pl), respectivement, dérivé de deux formes d'un nom en latin vulgaire (LV), *nātiōnem* (sg, acc), *nātiōnes* (pl, acc) est décrite sur la Figure A.3. Dans ce cas, comme les formes portugaises concernées sont toutes deux au pluriel et au singulier, l'étymon a deux instances WordForm correspondant chacune à un nombre grammatical. Il faut notamment remarquer l'association des attributs `grammaticalCase` et `inflectionType` à la sous-classe WordForm d'Etymon via une instance GrammaticalInformation. Dans un lexique de portugais détaillé qui contiendrait de telles informations étymologiques pour un nombre suffisant d'entrées lexicales, il serait possible, en opposant le contenu de l'instance WordForm dans LexicalEntry à la classe Etymon, de percevoir les phénomènes langagiers suivants: 1) perte du cas grammatical portugais; 2) la grande majorité de ses noms découlent du cas accusatif en LV; 3) lorsque la terminaison singulière (accusatif) en LV était *-tiōnem*, la forme portugaise s'écrit «-ção» et se prononce [sɛ̃w]; lorsque la terminaison pluriel (accusatif) en LV était *-tiōnes*, la forme portugaise s'écrit «-ções» et se prononce [sõ̃s].

NOTE Cette étymologie peut être davantage articulée en ajoutant les types de processus phonologiques pour chaque phase de la diachronie. Cela peut être effectué dans le modèle en ajoutant la catégorie de données appropriée définie dans l'Annexe B, à la valeur de type d'EtyLink pour les phases indiquées.