

2 Codification des principes rédactionnels

Marco Maggiore

2.1 Considérations sur la structure XML des articles du DÉRom

1 Introduction

À la différence de ses prédécesseurs, dont le plus éminent est le *Romanisches Etymologisches Wörterbuch* de Wilhelm Meyer-Lübke (REW₃ ; ³1930–1935 [1911–1920]), le *Dictionnaire Étymologique Roman* (DÉRom), lancé en janvier 2008, est élaboré entièrement par voie informatique : ses articles sont saisis directement en langage XML (dans le logiciel oXygen), ce qui permet, grâce au schéma et à la feuille de style élaborés par Gilles Souvay, l’informaticien du projet, leur publication à la fois sur le site internet de l’ATILF (<www.atilf.fr/DERom>) et sous forme papier (cf. DÉRom 1 et DÉRom 2). Dans ce chapitre, qui fait suite aux considérations de Gilles Souvay et Pascale Renders publiées dans le DÉRom 1 (Souvay/Renders 2014), nous fournissons une présentation synthétique du schéma XML du DÉRom et de son fonctionnement sur le site.

Nos développements, assez détaillés, s’appuient sur l’exemple concret de l’article */kresk-e-/ v.intr./tr. ‘croître ; accroître’ (cf. Maggiore 2011–2019 in DÉRom s.v., DÉRom 1, 492–495 et ici 363–367).¹ Le choix de cet article, que nous avons rédigé pendant notre post-doctorat à l’ATILF, est dû à des raisons purement pratiques : il va de soi que toutes les entrées du DÉRom ont la même structure informatique et reposent sur le même schéma XML sous-jacent. L’utilisation du langage XML permet de contrôler de manière rigoureuse le respect de la structure lexicographique de chaque article.

Nos considérations seront structurées en deux parties complémentaires : dans un premier temps, nous observerons la structure lexicographique intrinsèque d’un article du DÉRom (cf. aussi Celac 2016 et Kneib 2016, 196–214),

¹ L’article */kresk-e-/ bénéficie d’une republication dans le DÉRom 3 pour que le lecteur puisse s’y reporter aisément lors de la lecture du présent chapitre.

pour, dans un second temps, montrer comment cette dernière se reflète dans le schéma XML.

2 Structure lexicographique

2.1 Présentation

Un article du DÉRom peut se subdiviser en sept sections, dont six sont obligatoires : le lemme (ci-dessous 2.2), les matériaux (2.3), le commentaire (2.4), la bibliographie (2.5), les signatures (2.6) et la date de mise en ligne (2.7) ; une dernière section, facultative, est consacrée aux notes (2.8).

2.2 Lemme

La première section, dévolue au lemme, exprime la base étymologique reconstruite dans l'article. Elle se compose de trois parties : le signifiant, la catégorie grammaticale et le signifié.

Il est important de rappeler que le DÉRom présente une caractéristique particulière concernant la notation du signifiant des étymons : étant donné qu'il ne s'agit pas de mots du latin écrit, mais de protolèxèmes reconstruits sur la base de séries de cognats romans, leur signifiant est présenté en notation phonologique, à l'aide des symboles de l'alphabet phonétique international (IPA), placé entre barres obliques et précédé d'un astérisque, en l'occurrence */kresk-e-/. Comme nous le verrons plus tard, ce choix méthodologique entraîne des conséquences pour la structure XML des articles.

La catégorie grammaticale est exprimée par une abréviation, dans notre cas « v.intr./tr. », c'est-à-dire 'verbe intransitif et transitif'.

Enfin, le signifié de l'étymon est exprimé sous la forme d'une définition componentielle introduite entre guillemets : 'rendre progressivement de l'ampleur ; faire augmenter en extension ou en volume'.

Considéré de manière globale, le lemme synthétise ainsi le résultat de la reconstruction comparative réalisée par l'auteur de l'article, en énonçant les caractéristiques phonologiques, (micro-)syntaxiques et sémantiques de l'étymon.

2.3 Matériaux

Dans cette section de l'article, une ou plusieurs série(s) de cognats romans, c'est-à-dire un ou des ensemble(s) de données lexicales reconductibles à une ou plusieurs base(s) étymologique(s) commune(s), est ou sont enregistrée(s), chaque cognat occupant une place prédictible.

La section consacrée aux matériaux peut consister en un seul bloc de données ou bien être subdivisée en plusieurs paragraphes, comme c'est le cas de l'article */kresk-e-/, qui comporte deux subdivisions sémantico-valenciellement motivées, consacrées respectivement au verbe intransitif signifiant 'croître' (I.) et au verbe transitif signifiant 'accroître'. Chaque subdivision exhibe une liste ordonnée de cognats, et pour chacun de ces cognats, le lecteur est en mesure de repérer un certain nombre d'informations, exprimées à travers la même structure figée. Observons d'abord le premier cognat enregistré dans le premier paragraphe, le verbe intransitif sarde *krèskere* :

« **sard.** *krèskere* v.intr. 'prendre progressivement de l'ampleur, croître' (DES ; PittauDizionario 1) »

On pourrait subdiviser cet item en deux parties informationnelles : celle qui précède les parenthèses et celle qui apparaît à l'intérieur de ces dernières.

Quatre éléments peuvent apparaître avant les parenthèses : l'étiquette géolinguistique du cognat (« sard. »), son signifiant (« *krèskere* »), sa catégorie grammaticale (« v.intr. ») et son signifié (« 'prendre progressivement de l'ampleur, croître' »). Le premier de ces éléments, marqué en gras, est un glottonyme spécifiant l'idiome – qu'il s'agisse d'une langue nationale, d'un dialecte local ou de tout autre type de parler – dont relève le cognat. Puis le signifiant du cognat est indiqué, sous sa forme citationnelle, donc, dans le cas des verbes, en principe sous la forme de l'infinitif. Le signifiant, qui constitue pour ainsi dire le nom du cognat – car on peut faire référence à ce dernier à travers le premier –, est édité en italique s'il est de tradition écrite (« *krèskere* »), tandis que dans le cas de cognats enregistrés seulement sous leur forme orale, on recourt à la notation phonétique (entre crochets carrés). Le troisième élément précédant les parenthèses est la catégorie grammaticale, synthétisée dans ce cas par l'abréviation « v.intr. » ('verbe intransitif'). Le signifié, enfin, est présenté sous la forme d'une définition componentielle ('prendre progressivement de l'ampleur'), suivie d'une glose rapide ('croître'). Les deux premiers éléments, le glottonyme et le signifié, apparaissent de manière obligatoire pour tous les cognats, tandis que les deux derniers, la catégorie grammaticale et le signifié, sont absents de la structure de surface de l'article si leur valeur coïncide avec

celle du cognat précédent. Cela se vérifie, dans notre exemple concret, pour tous les cognats qui suivent le sarde, y compris le cognat français, qui figure en douzième position dans la série de cognats :

« fr. *croître* (dp. ca 1100 [*creistre*], RoLS: 142 = TLF ; Gdf ; FEW 2, 1323b ; TL ; AND₂ s.v. *crestre* ; DMF2012 ; ALF 362) »

Le fait que la catégorie grammaticale « v.intr. » et le signifié ‘prendre progressivement de l’ampleur, croître’ fournis explicitement pour le cognat sarde soient les seuls à apparaître dans ce paragraphe signifie implicitement que tous les cognats qui y sont listés partagent cette valence et cette valeur sémantique avec le cognat sarde placé au début de la liste.

Passons à présent à l’information présentée à l’intérieur des parenthèses. Si nous comparons à nouveau le cognat sarde avec le cognat français, nous remarquons encore une fois des différences. Cette fois, c’est le cognat français qui est pourvu de plus d’information que le sarde, puisqu’il comporte une date de première attestation (« dp. ca 1100 »), puis, entre crochets carrés, la forme de cette attestation (« [*creistre*] »). Il ne s’agit pas là d’éléments obligatoires : ils ne sont fournis que s’ils sont pertinents. Ainsi, la date de première attestation de certains cognats est tellement récente que sa mention ne ferait qu’alourdir le bloc des matériaux, sans apporter une plus-value informationnelle quelconque. C’est le cas du verbe sarde, dont nos sources ne fournissent pas d’attestations antérieures au 19^e siècle.

En revanche, des séries d’abréviations bibliographiques sont associées à chaque cognat. Ces abréviations renvoient à des sources, qu’elles soient de consultation obligatoire (cf. Morcov 2016) ou facultative, dépouillées durant le processus rédactionnel. Si une date de première attestation est indiquée, la source qui la contient est citée en premier ; pour le reste, la séquence des sources suit un ordre (chronologique) préétabli, de la plus ancienne à la plus récente.

En résumé, la section des matériaux fournit un ensemble d’informations assez complet : elle liste les idiomes qui attestent des lexèmes permettant d’établir une série de cognats donnée, elle précise le signifiant de ces cognats, leur catégorie grammaticale et leur signifié ; elle nous informe, le cas échéant, sur la date de leur première attestation, en précisant si nécessaire la forme de cette dernière ; enfin, elle présente une liste de sources consultables par qui voudrait en savoir plus.

Comme on peut s’en douter, l’ordre de citation des cognats dans chaque paragraphe de l’article n’est pas dû au hasard : il existe un ordre de citation spécifique déterminé par les normes rédactionnelles du dictionnaire (cf. Celac 2016, 291–295), qui établit que le premier cognat cité doit être, s’il existe, le sarde,

puis le dacoroumain, puis l'istroroumain etc., jusqu'au dernier, le portugais (ou, le plus souvent, le galégo-portugais). Des idiomes peuvent être absents de la liste, (s'ils n'attestent pas de lexème pouvant être rattaché à la série de cognats donnée), mais ils ne peuvent pas apparaître dans un ordre arbitraire : le sarde doit toujours précéder le dacoroumain, et le portugais ne peut pas être suivi par l'italien.

La série des cognats est précédée obligatoirement par le signifiant de l'étymon direct auquel elle est reconduite, dans le cas présent le mot-forme (infinitif) */'kresk-e-re/, une entité moins abstraite que celle portée dans le lemme (*/'kresk-e-/): à partir d'une série d'infinitifs romans, on ne peut que reconstruire, dans un premier temps, un infinitif protoroman.

Le seul élément qui peut précéder le signifiant de l'étymon direct, c'est le titre du paragraphe, qui n'apparaît que dans les cas où la section des matériaux contient au moins deux subdivisions. En l'occurrence, ce titre se lit ainsi : « I. Verbe intransitif : 'croître' ».

2.4 Commentaire

Troisième section d'un article du DÉRom, le commentaire étymologique se compose au minimum de deux alinéas de texte, dont le premier explicite la reconstruction de l'étymon protoroman porté dans le lemme, tandis que le second compare le résultat de la reconstruction comparative avec les données du latin écrit. En fonction de la complexité d'un article, d'autres paragraphes peuvent s'y ajouter, par exemple pour expliquer sa structure interne ou pour aborder une question de reconstruction (phonologique, sémantique, morphologique ou syntaxique) qui frappe la base protoromane faisant l'objet de l'analyse.

Le premier paragraphe du commentaire décrit de façon plus ou moins synthétique la ou les série(s) de cognats sur le(s)quel(s) se fonde la reconstruction et énonce de façon explicite le résultat de cette dernière :

« Tous les parlers romans sans exception présentent des cognats incitant à reconstruire protorom. */'kresk-e-/ v.intr. 'prendre progressivement de l'ampleur, croître', tr. 'faire augmenter en extension ou en volume, accroître'.

L'étymon y est mentionné sous la même forme que dans le lemme : signifiant, catégorie grammaticale, signifié. Toutefois, tandis que dans le lemme, chaque signifié est introduit par une définition componentielle, cette dernière est ici complétée, comme dans les matériaux, par une rapide glose, qui se réduit en général à un seul mot ('croître', 'accroître').

Un deuxième alinéa obligatoire, souvent placé à la fin du commentaire, aborde la question des rapports entre la base protoromane reconstruite et son corrélat en latin écrit :

« Le corrélat de I. en latin écrit, *crescere* v.intr. ‘croître’, est connu durant toute l’Antiquité (dp. Caton [* 234 – † 149], TLL 4, 1176), tandis que le latin écrit de l’Antiquité ne connaît pas de corrélat du type II. En revanche, le sens principal de *crescere*, ‘naître, venir au monde’, usuel durant toute l’Antiquité (dp. Ennius [* 239 – † 169], TLL 4, 1176), est étranger aux cognats romans (cf. aussi DOLR 1, 56-57) ».

Le commentaire de l’article */*kresk-e-*/ contient en outre un autre paragraphe, inséré entre ces deux paragraphes obligatoires, dont la finalité principale consiste à expliquer la structuration de l’article. D’une manière générale, il est important de préciser que bien que la structure interne du commentaire soit assez peu contrainte et qu’il n’existe pas de limite en termes de longueur, le rédacteur doit respecter un certain nombre de normes de rédaction et de mise en forme du texte (emploi des italiques, format de citation des références bibliographiques etc.).

2.5 Bibliographie

La quatrième section d’un article du DÉRom occupe la bibliographie. Il s’agit d’une liste de références bibliographiques, classées par ordre chronologique, qui portent sur des questions générales (« Romania en général ») liées à la base protoromane traitée, tandis que les références qui concernent les idiomes romans individuels sont enregistrées dans la section « Matériaux ».

À noter que certaines sources sont citées dans ce paragraphes d’une manière différente par rapport à la section des matériaux. C’est par exemple le cas du *Französisches Etymologisches Wörterbuch* (FEW), dont la référence est ici plus riche et plus explicite (« von Wartburg 1945 in FEW 2, 1323a-1330a, CRÉSCÈRE ») que dans les matériaux (« FEW 2, 1323a »).

2.6 Signatures

Introduite par le titre « Signatures » (en gras), la cinquième section rend compte des diverses contributions apportées par les chercheurs impliquées dans l’élaboration de l’article. Il s’agit souvent de vingt à vingt-cinq personnes, surtout des membres de l’équipe du DÉRom, mais aussi, exceptionnellement, des experts

extérieurs au projet sollicités par les rédacteurs : cette section témoigne du caractère éminemment collaboratif de l'entreprise.

Les contributeurs peuvent se rattacher à trois profils différents : la sous-section « Rédaction » explicite le nom de l'auteur (ou des auteurs) de l'article. Suit la sous-section « Réviseurs », qui liste toutes les personnalités intervenues dans la révision du travail, en apportant des observations sur des problèmes soulevés par des parlars romans en particulier ou sur l'analyse générale. La troisième sous-section, intitulée « Contributions ponctuelles », rend compte d'apports plus circonscrits à l'élaboration de l'article, comme la mise à disposition de photocopies ou de scans de matériaux bibliographiques ou des observations sur des aspects formels ou marginaux.

2.7 Dates de publication

Dans cette section, la date à laquelle l'article a été publié pour la première fois sur le site internet du DÉRom est indiquée (« Première version : 24/06/2011 »). Une deuxième date, qui peut être identique à la première ou en être différente, indique le moment où les dernières modifications ont été effectuées dans l'article (« Version actuelle : 20/12/2019 »). Les raisons d'une correction postérieure à la mise en ligne initiale d'un article peuvent être variées ; dans la plupart des cas, il s'agit de révisions internes ou de corrections suggérées dans un compte rendu.

2.8 Notes

La dernière section dans la structure d'un article du DÉRom, la seule optionnelle, est consacrée aux notes. Ces dernières apportent des informations sur des points de détail ou abordent des questions secondaires par rapport à l'objet du commentaire étymologique. Par exemple, la note 2 de l'article */kresk-e-/ explique pourquoi les deux séries de cognats assemblées dans l'article contiennent, à côté de cognats comme sard. *krèskere* ou fr. *croître*, dont le rattachement à */kresk-e-re/ ne pose aucun problème, des items comme esp. *crecer* (accentué sur la dernière syllabe), qui suggéreraient, en première approche, plutôt la reconstruction de **/kres'k-e-re/ :

« 2. En espagnol, asturien et galégo-portugais, les verbes remontant à la classe flexionnelle en */-e-/ sont passés à celle remontant à */-e-/ , exceptionnellement aussi à celle remontant à */-i-/ (cf. MeyerLübkeGLR 2, § 126 ; LausbergLinguistica 2, § 788) ».

Avec le commentaire, les notes constituent la seule section susceptible de contenir des paragraphes de texte. Elles peuvent se rattacher, à l'aide d'un appel de note, aux matériaux et au commentaire étymologique.

3 Reflet de la structure lexicographique dans le schéma XML

3.1 Présentation

Après avoir exposé la structure lexicographique d'un article du DÉRom, nous passons à présent au schéma XML, qui est le reflet informatique de cette structure. Nous effectuerons notre présentation du point de vue d'un linguiste qui utilise le système de rédaction XML sans prétendre être un expert en traitement automatique des langues.

La figure 1 ci-dessous, qui a été générée par copie du début de l'article */kresk-e-/ sous XML, donne une idée de la structure arborescente du schéma. Après des informations concernant l'encodage informatique du document, le texte enchaîne des paires de balises (comme <Article> ~ </Article>), qui peuvent être imbriquées entre elles (ainsi <Lemme> ~ </Lemme> dans <Article> ~ </Article>), et du contenu textuel (ainsi « prendre progressivement de l'ampleur », contenu de la paire de balises <Signifie> ~ </Signifie>).

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/css" href="derom.css"?>
<DERom xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="derom.xsd">
  <Article><Lemme>
    <Signifiant>'kresk-e-</Signifiant>
    <catgramm>v.intr./tr.</catgramm>
    <Signifie>prendre progressivement de l'ampleur</Signifie>
    <Signifie>faire augmenter en extension ou en volume</Signifie>
  </Lemme><Matériaux><subdiv>
    <titre>I. Verbe intransitif : « croître »</titre>
    <etym>'kresk-e-re</etym>
    <cognats>
      <premier.cognat>
        <idiome>sard.</idiome>
        <signifiant>krèskere</signifiant>
        <catgramm>v.intr.</catgramm>
        <signifie>prendre progressivement de l'ampleur, croître</signifie>
      <refbibl>
```



```

<unerefbibl>
  <reference>DES</reference>
  <reference>PittauDizionario 1</reference>
</unerefbibl></refbibl></premier.cognat>

```

Figure 1 : Début de la structure arborescente du schéma XML de l'article */'kresk-e-/'

Conformément à la métalangue du projet, les balises portent (en principe) des noms en français. Leur séquence définit la structure XML des articles, dont sept paires de balises fondamentales constituent en quelque sorte l'ossature :

```

<Lemme> ~ </Lemme>
<Materiaux> ~ </Materiaux>
<Commentaire> ~ </Commentaire>
<Bibliographie> ~ </Bibliographie>
<Signatures> ~ </Signatures>
<MiseEnLigne> ~ </MiseEnLigne>
<Notes> ~ </Notes>

```

À l'intérieur de ces balises fondamentales, d'autres types de balises peuvent être insérés. Leur agencement n'est pourtant pas libre : la plupart du temps, une paire de balises donnée ne peut apparaître que dans un contexte bien particulier. Par exemple, la paire de balises <cognats> ~ </cognats> ne peut se présenter qu'à l'intérieur de la paire de balises plus englobante <Materiaux> ~ </Materiaux>, tandis que les balises <reference> ~ </reference> peuvent apparaître à l'intérieur de <Materiaux> ~ </Materiaux>, <Commentaire> ~ </Commentaire>, <Bibliographie> ~ </Bibliographie> et <Notes> ~ </Notes>.

3.2 Lemme

Avant de rentrer dans le vif du sujet, il est important de rappeler que les balises XML, dont le rôle est principalement structurel, représentent en même temps, à travers la feuille de style en cascade (fichier CSS [Cascading Style Sheet]) associée au schéma, une balise de formatage : elles véhiculent donc aussi des informations quant à l'affichage des articles sur le site web du DÉRom et, par voie de conséquence, dans les volumes papier du dictionnaire. Par exemple, la balise <Lemme> est associée à un affichage de son contenu en gras et en taille plus grande que le reste de l'article.

Trois éléments obligatoires apparaissent dans le champ délimité par la paire de balises <Lemme> ~ </Lemme> : <Signifiant> ~ </Signifiant>, <catgramm> ~ </catgramm> et <Signifie> ~ </Signifie>.

La feuille de style prépare le plus possible le travail du rédacteur. La paire de balises `<Signifiant> ~ </Signifiant>` génère ainsi automatiquement un astérisque et une barre oblique avant et une barre oblique après le contenu qui y est inséré, ce qui évite des oublis lors de la rédaction et simplifie la saisie.

Les balises `<Signifiant>` et `<Signifie>` ne sont pas frappées de restrictions quant à leur contenu : elles acceptent du texte libre. En revanche, la balise `<catgramm>` ('catégorie grammaticale') requiert l'emploi d'une abréviation appropriée (en l'occurrence, « v.intr./tr. »), choisie dans une liste prédéterminée. L'usage d'une valeur inattendue, c'est-à-dire d'une abréviation de catégorie grammaticale qui ne serait pas prévue par les normes rédactionnelles, sera marqué comme une erreur (soulignée à l'aide de petits points rouges) par l'éditeur XML.

Une caractéristique distinctive de la balise `<Signifie>`, c'est qu'elle peut être répétée plusieurs fois, ce qui permet de caractériser un étymon comme (bi- ou) polysémique. C'est le cas de `*/'kresk-e-/`, qui est formé de deux unités lexico-sémantiques : un verbe intransitif signifiant 'croître' et un verbe transitif signifiant 'accroître' :

```
<Signifie>prendre progressivement de l'ampleur</Signifie>
<Signifie>faire augmenter en extension ou en volume</Signifie>
```

Grâce à la feuille de style, qui supplée les guillemets et le point-virgule séparant les deux sémèmes, la définition de l'étymon s'affichera donc ainsi : 'prendre progressivement de l'ampleur ; faire augmenter en extension ou en volume'.

3.3 Matériaux

La section dévolue aux matériaux constitue la partie structurellement la plus complexe de l'article. Elle contient de manière obligatoire au moins une occurrence de la paire de balises `<subdiv> ~ </subdiv>`. L'article `*/'kresk-e-/` en contient deux, la première entourant toutes les informations afférentes au verbe intransitif signifiant 'croître', la seconde délimitant toutes celles concernant le verbe transitif au sens 'accroître'.

Les deux subdivisions s'ouvrent chacune par un premier élément, délimité par les balises `<titre> ~ </titre>`, qui marque le titre du paragraphe en question (ainsi « I. Verbe intransitif 'croître' »). Il s'agit d'un élément optionnel, qui peut être supprimé sans affecter la validité du schéma XML, car aucun titre n'est prévu pour les articles dont les matériaux ne sont pas subdivisés en plusieurs sous-sections. Les rédacteurs sont conscients du fait que les articles dont les données sont agencées dans plusieurs subdivisions (séparant des types phonologiques,

morphologiques, sémantiques etc.), l'introduction de titres est obligatoire, afin de rendre explicite la structure des matériaux pour le lecteur.

Le premier élément obligatoire à l'intérieur de la paire de balises `<subdiv> ~ </subdiv>` – qui succède, dans le cas présent, à la balise dévolue au titre de la subdivision –, est la paire de balises `<etym> ~ </etym>`. Il s'agit d'une indication rapide de l'étymon, sous la forme de son signifiant. Cette balise a une fonction proche de celle de la balise `<Signifiant>` du lemme, mais concerne seulement le signifiant de l'ancêtre commun de la série de cognats réunie dans la subdivision en question. Dans la version finale de l'article, ce signifiant s'affiche entre barres obliques, précédé d'un astérisque et suivi du symbole 'supérieur à' (« > »), qui marque la descendance des cognats par rapport à l'étymon. Tous ces symboles sont générés de manière automatique par la feuille de style.

Après la balise dévolue à l'étymon direct intervient la paire de balises `<cognats> ~ </cognats>`, qui contient la série de cognats réunie dans la subdivision en question. À ce propos, il est intéressant de rappeler que le nom initialement attribué à cette balise, lors de l'élaboration du schéma XML en 2008, était `<issues>`. Or, une des caractéristiques majeures du projet DÉRom réside, en raison de son ancrage méthodologique dans la grammaire comparée, dans son orientation « bottom up » plutôt que « top down ». ² Pour cette raison, l'équipe a décidé il y a quelques années (vers 2015, si nos souvenirs sont bons) de remplacer `<issues>` par `<cognats>`. En effet, si les deux noms sont factuellement corrects, *cognat* semble plus approprié, dans la mesure où les lexèmes réunis dans une subdivision forment dans un premier temps une série de cognats permettant de reconstruire leur ancêtre commun, l'étymon, et que c'est seulement en vertu de cette reconstruction que, dans un second temps, les lexèmes cités dans cette subdivision peuvent être considérés comme des issues (ou des continueurs) de l'étymon.

Le champ délimité par la paire de balises `<cognats> ~ </cognats>` doit obligatoirement contenir l'élément `<premier.cognat> ~ </premier.cognat>`, qui peut être suivi d'un nombre illimité d'éléments `<cognat> ~ </cognat>`. Cela tient au fait que le premier cognat de la série de cognats – qui, très exceptionnellement, peut être le seul (cf. par exemple Greub 2014–2019 in DÉRom s.v. */'akuil-a/III.2.) –, est balisé d'une manière différente des autres. La structure des balises `<premier.cognat>` et `<cognat>` diffère en effet par le fait que pour la seconde, l'apparition des balises incluses `<catgramm>` et `<signifie>` est optionnelle,

2 Cf. Buchi/Schweickard (2009, 104) : « par cette pratique originale, le DÉRom se démarque clairement du REW : là où ce dernier répondait à la question du devenir du lexique latin (classique), le DÉRom se pose clairement celle de l'origine du lexique roman ».

l'information qu'elles véhiculent étant implicite si elle est identique à celle du *cognat* précédent (cf. ci-dessus 2.3). On a donc :

«**premier.cognat**» <idiome> </idiome> <signifiant> </signifiant> <catgramm> </catgramm> <signifie> </signifie> <refbibl> </refbibl> </**premier.cognat**>

Mais :

«**cognat**» <idiome> </idiome> <signifiant> </signifiant> [<catgramm> </catgramm>] [<signifie> </signifie>] <refbibl> </refbibl> </**cognat**>

Appliquée à sard. *krèskere* et fr. *croître* (cf. ci-dessus 2.3), la structure (simplifiée pour ce qui est des références bibliographiques) se présente ainsi :

«**premier.cognat**» <idiome>sard.</idiome> <signifiant>krèskere</signifiant> <catgramm> v.intr.</catgramm> <signifie>prendre progressivement de l'ampleur, croître</signifie> <refbibl>DES [...]</refbibl> </**premier.cognat**>

«**cognat**» <idiome>fr.</idiome> <signifiant>croître</signifiant> <refbibl>dp. [...]</refbibl> </**cognat**>

Pour ce qui est de la paire de balises <refbibl> ~ </refbibl>, elle peut accueillir un ou plusieurs élément(s) <unerefbibl>³ subdivisé(s) en un nombre illimité de balises <reference>. Ces dernières forment le cœur du dispositif bibliographique du DÉRom, puisqu'elles assurent le lien, sur le site web du projet, entre les sigles (abréviations) et les références bibliographiques qui leur sont associées (cf. Costa/Baudinot/Perignon ici 477–580). En cliquant sur un sigle, le lecteur de l'article en ligne est ainsi redirigé sur une page présentant l'ensemble des informations bibliographiques pertinentes. Il s'agit certainement là de l'un des plus grands avantages de l'adoption du langage XML pour l'encodage des articles.

Par ailleurs, il faut souligner encore une fois que les balises XML sont aussi, en raison de l'association du schéma avec la feuille de style, des balises de formatage. Par exemple, la balise <idiome> génère automatiquement un formatage en gras, tandis que la balise <signifiant> (à ne pas confondre avec <Signifiant>, à l'initiale majuscule, qui peut apparaître seulement dans le lemme), sélectionne en principe l'italique. En outre, chaque sigle bibliographique est

³ La juxtaposition de deux paires de balises <unerefbibl> ~ </unerefbibl> est extrêmement rare : elle ne concerne que des cas où il est indispensable de présenter les références bibliographiques dans plusieurs paquets.

associé à un formatage spécifique. Par exemple, la séquence « AND2 s.v. crestre » dans le cognat français sera transformée automatiquement en « AND₂ s.v. *crestre* », avec le nombre 2 en indice et la deuxième partie de la référence en italique.

Assez régulièrement, les rédacteurs sont amenés à apporter une précision à la mention de la première attestation d'un cognat donné. Pour ce faire, ils ont recours à la paire de balises `<precision>` ~ `</precision>`, qui peut s'insérer entre la mention de la date de la première attestation (`</date>`) et celle de sa source (`<reference>`). L'appel de cette balise génère une zone de texte délimitée par des crochets carrés, où le rédacteur peut introduire les précisions souhaitées. Nous avons dû procéder ainsi pour fr. *croître*, car la première attestation de ce cognat se présente sous une forme différente de la graphie moderne (*creistre*), de sorte que la précision était obligatoire (cf. Celac 2016, 286). Pour que *creistre* apparaisse en italique, nous lui avons appliqué la balise de formatage `<hi>` (« highlighting »). Il s'agit de la seule balise de formatage prévue par le schéma, et elle nécessite une précision, que ce soit « bold » ('gras'), « italic » ('italique'), « smallcaps » ('petites capitales'), « subscript » ('en indice'), « superscript » ('en exposant') ou « underscore » ('souligné', option non prévue pour les parties des articles du DÉRom gérées par les rédacteurs). L'absence de précision sera interprétée comme une erreur par le système de contrôle XML.

En outre, le rédacteur a la possibilité d'ajouter d'autres précisions, par exemple des informations diatopiques, après n'importe quelle référence bibliographique, en utilisant la balise optionnelle `<precisionreference>`.

3.4 Commentaire

La structure XML du commentaire est beaucoup plus simple que celle de la section consacrée aux matériaux. Elle peut être résumée comme suit :

```
<Commentaire> <p> <p> <p> </p> [...] </Commentaire>.
```

Le commentaire peut accueillir un nombre illimité de paragraphes de texte, délimités chacun par une paire de balises `<p>` ~ `</p>`. Cette section fonctionne ainsi comme un espace de texte très libre, où le rédacteur peut développer les résultats de son analyse, poser les bonnes questions et, si possible, y répondre. Trois paires de balises, `<etymsignifiant>` ~ `</etymsignifiant>`, `<etymsignifie>` ~ `</etymsignifie>` et `<correlatlatin>` ~ `</correlatlatin>`, méritent d'être mentionnées ici.

La balise `<etymsignifiant>`, qui apparaît au moins une fois, dans le premier paragraphe du commentaire, contient le même contenu que la balise `<Signifiant>`

rattachée au lemme : il s'agit de la forme de l'étymon protoroman reconstruit. En revanche, la présentation matérielle est différente de celle du lemme : si le signifiant est, ici aussi, marqué par un astérisque et des barres obliques, il n'apparaît ni en gras ni dans une taille de police plus grande.

La balise <etymsignifie> apparaît elle aussi au moins une fois, de même dans le premier paragraphe du commentaire. À la différence de la balise <Signifie> du lemme, qui ne contient qu'un élément, à savoir une définition componentielle, la balise <etymsignifie> se subdivise en deux sous-balises : <analytique> pour la même définition componentielle et <glose> pour une glose rapide. Cette dernière balise permet de faire des recherches sémantiques sur le site du dictionnaire, à travers un moteur de recherche qui indexe les éléments balisés <glose>.

La balise <correlatlatin>, enfin, marque les corrélats en latin écrit de l'Antiquité des bases protoromanes reconstruites. Cette balise est associée, elle aussi, au moteur de recherche du site web. Elle permet ainsi de consulter le DÉRom à travers le latin (classique), ce qui est très utile pour les romanistes de formation traditionnelle, comme le souligne Harald Völker :

« Durante el congreso de Innsbruck, la presentación de la elaboración del nuevo diccionario etimológico románico DÉRom por Éva Buchi y Wolfgang Schweickard desencadenó una discusión animada provocada por la decisión de no indicar como entrada de los artículos un étimo del latín clásico, sino un étimo protorrománico reconstruido. Es cierto que los editores alegaron buenos argumentos lingüísticos para esta decisión – pero sigue existiendo un contra-argumento de tipo práctico : Una entrada tendría que servir – entre otras finalidades – de interfaz entre los conocimientos del usuario y las informaciones del diccionario. Y los conocimientos aun del lingüista regular normalmente no implican los étimos protorrománicos reconstruidos.

La hipertextualización del DÉRom ofrece una salida a este problema porque posibilita más de una categoría de entrada : Además del étimo protorrománico puede establecerse un segundo tipo de entrada <mot de référence en latin classique> que contenga el lexema del latín clásico que formalmente precedía el lexema románico en cuestión. Así es el usuario quien selecciona su entrada preferida » (Völker 2010, 393).

Une balise optionnelle utilisée couramment dans le commentaire (ainsi que dans les notes) est <RenvoiInterne>, qui gère les références croisées, à travers un lien informatique, entre articles du dictionnaire.

Les rédacteurs peuvent aussi renvoyer dans le commentaire à tout type de référence bibliographique, en utilisant la balise <reference>. La balise <hi> est disponible pour formater des parties de texte, ainsi, par exemple, pour mettre en italique les abréviations latines *cf.* et *s.v.* Pour sa part, la balise <table> permet d'introduire, dans le cas d'analyses plus complexes, de l'information sous forme de tables (cf. Buchi/Delabarre/Fister/Huguet/Hütsch/Juroszek/Moisson/Pausé 2016–2019 in DÉRom s.v. */'βad-e-/).

La balise `<appelnote>` (qui peut être appelée aussi dans la section consacrée aux matériaux) insère un appel de note sous forme de nombre en exposant, auquel se rattache une note (cf. ci-dessous 3.8). Le rédacteur doit toutefois surveiller la correspondance entre l'ordre des appels de note et la séquence des notes elles-mêmes : les deux entités sont insérées manuellement, sans connexion structurelle prédéterminée entre elles, ce qui semble constituer une limite du schéma. Mais, comme nous serons amené à le préciser dans ce qui suit, le site web du DÉRom met à disposition des auteurs des articles des outils performants qui permettent de détecter les erreurs possibles (cf. ci-dessous 4).

3.5 Bibliographie

Le champ délimité par la paire de balises `<Bibliographie>` ~ `</Bibliographie>` consiste en une liste de sigles bibliographiques, étiquetés à l'aide de la balise déjà commentée précédemment `<reference>` (cf. ci-dessus 3.3). La structuration informatique de cette section est donc très simple.

À noter que le contenu de la balise `<reference>` consacrée au *Romanisches Etymologisches Wörterbuch* (REW₃) est exploité par le moteur de recherche du site web du DÉRom, qui permet ainsi de faire des recherches sur les entrées du REW₃.

3.6 Signatures

La section délimitée par la paire de balises `<Signatures>` ~ `</Signatures>` contient obligatoirement les balises suivantes : `<signaturesRedaction>` (pour le[s] auteur[s] de l'article), `<signaturesRevision>` (pour les experts ayant révisé l'article du point de vue d'un domaine linguistique en particulier ; cette balise prévoit des sous-balises par domaines géographiques), `<signaturesRevisionFinale>` (pour les réviseurs ayant révisé l'article dans son ensemble) et `<signaturesContributions>` (pour les chercheurs ayant apporté une contribution ponctuelle à l'article).

Chacune de ces catégories de signatures peut recevoir plusieurs occurrences des balises indissociables `<prenom>` et `<nom>` ; ces dernières sont exploitées par le moteur de recherche en ligne, qui permet ainsi de faire des recherches par contributeur dans l'ensemble des articles.

3.7 Dates de publication

La paire de balises `<MiseEnLigne>` ~ `</MiseEnLigne>` gère la mention des dates de publication des différentes versions de l'article. Elle inclut obligatoirement deux balises : `<Version1>` et `<VersionActuelle>`, qui sont toutes les deux exploitées par le moteur de recherche.

3.8 Notes

Enfin, la paire de balises `<Notes>` ~ `</Notes>` délimite le champ consacré aux notes, qui est facultatif, et ne peut recevoir qu'une succession de balises `<uneNote>`, qui sont numérotées automatiquement. La structure de ces dernières est très simple :

```
<uneNote> <p> </p> [...] </uneNote>
```

À l'intérieur de la paire de balises `<p>` ~ `</p>`, qui délimite des paragraphes, du texte libre peut être inséré. Ce dernier peut accueillir toutes les balises optionnelles utilisables aussi dans le commentaire. Ainsi, pour reprendre l'exemple de la note 2 de l'article `*/kresk-e-/` (cf. ci-dessus 2.8), les balises `<hi>` (« highlighting »), `<reference>` (renvoi à une source) et `<sequenceprotoromane>` (suite de phonèmes, morphème ou lexème protoroman) y trouvent leur emploi :

```
<uneNote><p>En espagnol, asturien et galégo-portugais, les verbes remontant à la classe flexionnelle en <sequenceprotoromane>-e-</sequenceprotoromane> sont passés à celle remontant à<sequenceprotoromane>-e-</sequenceprotoromane>, exceptionnellement aussi à celle remontant à <sequenceprotoromane>-i-</sequenceprotoromane> ( <hi rend="italic">cf</hi>. <reference>MeyerLübkeGLR 2, § 126</reference> ; <reference>LausbergLinguistica 2, § 788</reference>).</p></une Note>
```

4 Conclusion

Le schéma XML du DÉRom, dont nous venons de décrire la structure fondamentale, s'avère extrêmement utile dans l'élaboration des articles du dictionnaire : il ne garantit non seulement l'homogénéité du travail rédactionnel, qui repose sur les épaules de nombreux chercheurs dispersés à travers toute l'Europe (et au-delà), mais prévient aussi bien des erreurs de surface susceptibles de défigurer inutilement le dictionnaire, voire guide l'analyse du rédacteur, qui est invité à se poser obligatoirement un certain nombre de questions.

En outre, avant d'envoyer leur article en révision finale, les rédacteurs sont tenus de vérifier sa validité formelle en utilisant un outil mis à disposition par Gilles Souvay sur le site du DÉRom. En effet, après avoir cliqué sur « Passer en mode rédaction » et s'être identifiés, ils y ont accès à un onglet intitulé « Articles XML », qui leur permet, à travers la fonctionnalité « Contrôler un article », de soumettre le fichier XML de leur article à un contrôle informatique qui vérifie la validité de la structure XML et le respect de toutes les règles formelles, en générant, le cas échéant, une liste des erreurs à corriger.

En conclusion, nous dirons qu'un article comme */kresk-e-/ tire, certes, sa valeur des analyses linguistiques qu'il contient, mais que l'apport pratique de la structure XML a une réelle utilité dans le processus rédactionnel.

5 Bibliographie

- Buchi, Éva/Schweickard, Wolfgang, *Romanistique et étymologie du fonds lexical héréditaire : du REW au DÉRom* (« Dictionnaire Étymologique Roman »), in : Alén Garabato, Carmen/Arnavielle, Teddy/Camps, Christian (edd.), *La romanistique dans tous ses états*, Paris, L'Harmattan, 2009, 97–110.
- Celac, Victor, *Normes rédactionnelles*, in : Buchi, Éva/Schweickard, Wolfgang (edd.), *Dictionnaire Étymologique Roman (DÉRom) 2. Pratique lexicographique et réflexions théoriques*, Berlin/Boston, De Gruyter, 2016, 257–327.
- Costa, Victoria/Baudinot, Pascale/Perignon, Jessika, *Bibliographie*, in : Buchi, Éva/Schweickard, Wolfgang (edd.), *Dictionnaire Étymologique Roman (DÉRom) 3. Entre idioroman et protoroman*, Berlin/Boston, De Gruyter, 2020, 477–580.
- DÉRom = Buchi, Éva/Schweickard, Wolfgang (dir.), *Dictionnaire Étymologique Roman (DÉRom)*, Nancy, ATILF, <<http://www.atilf.fr/DERom>>, 2008–.
- DÉRom 1 = Buchi, Éva/Schweickard, Wolfgang (edd.), *Dictionnaire Étymologique Roman (DÉRom). Genèse, méthodes et résultats*, Berlin/Munich/Boston, De Gruyter, 2014.
- DÉRom 2 = Buchi, Éva/Schweickard, Wolfgang (edd.), *Dictionnaire Étymologique Roman (DÉRom) 2. Pratique lexicographique et réflexions théoriques*, Berlin/Boston, De Gruyter, 2016.
- FEW = Wartburg, Walther von, et al., *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*, 25 vol., Bonn/Heidelberg/Leipzig-Berlin/Bâle, Klopp/Winter/Teubner/Zbinden, 1922–2002.
- Kneib, Marie-Thérèse, *Le DÉRom expliqué aux lecteurs non spécialistes, mais dotés d'une saine curiosité*, in : Buchi, Éva/Schweickard, Wolfgang (edd.), *Dictionnaire Étymologique Roman (DÉRom) 2. Pratique lexicographique et réflexions théoriques*, Berlin/Boston, De Gruyter, 2016, 163–218.
- Morcov, Mihaela-Mariana, *Bibliographie de consultation et de citation obligatoires*, in : Buchi, Éva/Schweickard, Wolfgang (edd.), *Dictionnaire Étymologique Roman (DÉRom) 2. Pratique lexicographique et réflexions théoriques*, Berlin/Boston, De Gruyter, 2016, 329–359.

- REW₃ = Meyer-Lübke, Wilhelm, *Romanisches Etymologisches Wörterbuch*, Heidelberg, Winter, 1930–1935 [¹1911–1920].
- Souvay, Gilles/Renders, Pascale, *Traitement informatique*, in : Buchi, Éva/Schweickard, Wolfgang (edd.), *Dictionnaire Étymologique Roman (DÉRom). Genèse, méthodes et résultats*, Berlin/Munich/Boston, De Gruyter, 2014, 247–257.
- Völker, Harald, *Edición de textos, hipertextos y lexicografía*, in : Iliescu, Maria/Siller-Runggaldier, Heidi/Danler, Paul (edd.), *Actes du XXV^e congrès international de linguistique et de philologie romanes (Innsbruck 2007)*, vol. 6, Berlin/New York, De Gruyter, 2010, 383–395.