Exploitation de l'Ontology Web Language pour la rédaction des notices de l'*Atlas Linguistique de la Wallonie*

Nicolas Mazziotta

Linguistique du français et dialectologie wallonne – B-4000 Liège – Belgique

Abstract

In order to build a digital edition of the upcoming volumes of the *Atlas Linguistique de la Wallonie*, we have decided to make use of the *Ontology Web Language* as described by the World Wide Web Consortium. We aim to develop a software meant to help the writing of onomasiological articles. It should allow us to describe located dialectal forms and progressively define groups of them, according to the way they *resemble* each other (e. g., from an etymological, morphological or phonetic point of view). The software-assisted process would then conform to methodological requirements of traditional scientific work, as it has thus far. OWL provides an easy way of defining form lists and grouping them afterwards.

Résumé

Pour l'informatisation future des prochains volumes de l'*Atlas Linguistique de la Wallonie*, nous avons décidé de tirer parti du potentiel de l'*Ontology Web Language* du World Wide Web Consortium. L'objectif que nous poursuivons est la création d'un logiciel d'aide à la rédaction de notices onomasiologiques, qui permette de partir d'un ensemble de formes dialectales localisées et de les organiser progressivement en groupes de formes partageant une *ressemblance* d'un point de vue étymologique, morphologique, phonétique, etc. De cette manière, la démarche assistée de classement et d'analyse pourra se conformer aux exigences méthodologiques du travail scientifique de rédaction tel qu'il a été mené traditionnellement jusqu'à présent. Le langage employé permet de construire aisément une liste de formes et de définir des ensembles *a posteriori*.

Mots-clés : géographie linguistique, wallon, XML, Ontology Web Language (OWL), ontologie.

1. Introduction

Dans cette contribution, nous présenterons la manière dont nous comptons mobiliser les techniques issues du Web sémantique (*Ontology Web Language*, soit OWL) pour travailler à la conception d'un logiciel d'aide à la rédaction de l'*Atlas linguistique de la Wallonie*. Nous avons montré précédemment (Baiwir et Mazziotta, à paraître) comment il est possible d'exprimer la structure des notices existantes à l'aide d'OWL et nous aimerions montrer comment ce langage peut être mobilisé pour tenir compte de la *dynamique rédactionnelle* d'une notice. C'est l'optique de la rédaction et non celle de l'utilisation que nous développerons donc.

L'ALW est une vaste entreprise d'édition de matériaux linguistiques. Une présentation générale de l'ouvrage ayant déjà été faite ailleurs (Counet, 1990 : 21-28 ; Boutier, 1996), nous nous contenterons ici d'un exposé succinct (0.1). Par ailleurs, Marie-Guy Boutier a effectué un travail de définition des structures de l'ALW dans son article à paraître : « Cinq relations de base pour définir la structure de l'Atlas linguistique de la Wallonie », un article qui a servi

de base à notre réflexion (0.2). Nous terminerons cette introduction par le plan de la suite de l'exposé (0.3).

1.1. L'ALW en deux mots

L'ALW est un atlas linguistique : il met à la disposition du lecteur un ensemble de formes linguistiques localisées géographiquement. L'ouvrage décrit les dialectes du nord du domaine d'oïl, en Belgique romane. Le père de l'ALW est Jean Haust, qui a entrepris dès 1924 l'enquête destinée à rassembler les données linguistiques. Ces données, essentiellement lexicales, ont été recueillies par le biais de questionnaires demandant aux locuteurs de *traduire* différents mots et énoncés français dans leur dialecte[1]. Une fois traitées, elles sont présentées sous la forme de notices comportant généralement un classement des mots qui servent à exprimer différentes notions dans ces dialectes (démarche onomasiologique)[2]. Chaque notice correspond à la notion qui en constitue le titre (transcrit par convention en petites capitales ; p. ex. : FRÈRE, ARBRE,...).

Ce qui fait l'originalité et l'intérêt de l'ALW est que cet ouvrage fournit des données synchroniques *classées* et *interprétées*, dans une perspective qui fait surgir la dimension historique. Loin de ne livrer qu'une simple liste de formes localisées (comme c'est le cas dans l'*Atlas linguistique de la France*), le rédacteur de chaque notice[3] analyse les formes phonétiques localisées et cherche les points communs étymologiques, sémantiques, lexicaux, morphologique et phonétiques, etc.

1.2. Structure des notices de l'ALW

D'après Marie-Guy Boutier (à paraître ; les références entre parenthèses de cette section se rapportent toutes à ce texte), il est possible de définir la structure idéale des notices de l'ALW à l'aide de cinq types de relations, que nous allons brièvement exposer ci-dessous.

Chacune des formes dialectales traitées est définie par un ensemble de relations de *déterminations* (§1) qui peuvent être considérées comme des données préalables à l'analyse : 1/ un sens ; 2/ une catégorisation morphosyntaxique ; 3/ une localisation géographique (point d'enquête) ; 4/ d'autres déterminations facultatives et accessoires. L'ALW étant onomasiologique, le sens de toutes les formes classées dans chaque notice est censé être stable[4]. La deuxième relation est celle de *hiérarchie* et d'*équivalence*, selon laquelle les formes sont groupées et hiérarchisées en fonction de ressemblances étymologiques, puis morpholexicales ou morphologiques et enfin phoniques (§2). Les formes sont unies par une relation d'*ordination* (§3), qui complète la précédente : elles sont exprimées en séquence, de manière à former des groupes et à définir des priorités. La relation de *contraste* (§4) unit deux variantes relevées dans un même dialecte (en un même point d'enquête). Enfin, la relation d'*intégration* (§5), rend compte de la manière dont les matériaux de l'enquête peuvent être coordonnés au reste de la documentation (autres enquêtes et études lexicales), à une échelle belgo-romane, gallo-romane, voire romane.

---

[1] Ce qui constitue un type d'échange linguistique particulier ; voir Boutier, à paraître : §1.1, *in fine*.

[2] Les volumes 1 et 2 sont grammaticaux, le premier couvrant la phonétique et le deuxième couvrant la morphologie (voir ALW 1 : 69).

[3] Un rédacteur travaille toujours sur un ensemble de notices liées par une cohérence thématique ou des similitudes morpholexicales. Cet ensemble peut constituer un volume.

[4] Il est fréquent que cela ne soit pas le cas (v. Boutier, à paraître : §2.3, *in fine*).

1.3. Plan de l'exposé

Le développement d'un outil informatique d'aide à la rédaction des notices de l'ALW se doit non seulement de prendre en compte sa dimension interprétative, mais également de respecter les principes qui gouvernent cette interprétation.

Notre exposé se focalisera essentiellement sur les trois premières des cinq relations décrites par Marie-Guy Boutier : *détermination*, *hiérarchie-équivalence* et *ordination*.

Après avoir brièvement expliqué en quoi consistait OWL à l'aide d'exemples applicables à l'ALW (2), nous présenterons concrètement comment les formes d'une notice peuvent être décrites d'un point de vue phonétique (3), puis comment cette description sert à les grouper (4). Nous montrerons alors comment les groupes obtenus peuvent être hiérarchisés et ordonnés (5) avant de conclure (6).

2. Présentation de l'Ontology Web Language

Nous décrirons brièvement la philosophie et les principaux concepts à la base d'OWL (2.1) et nous donnerons immédiatement un exemple de son utilisation dans le cas qui nous intéresse (2.2).

2.1. Philosophie et principaux concepts

L'*Ontology Web Language* (voir OWL-REF) est un langage dérivé de RDF (*Ressource Description Framework*). Ce dernier est un projet du W3C pour décrire les objets et les relations sémantiques d'Internet sous la forme de nœuds XML et de ressources identifiées (URI). L'objectif est de rendre les données interprétables par les machines.

Au delà de cet aspect ayant trait à l'automatisation, nous disposons d'un langage métaconceptuel formalisé pour définir les *classes* (nom donné aux concepts dans la terminologie OWL) et les *propriétés* (nom donné aux relations entre les concepts en OWL). Les classes sont actualisées en *individus*. Un ensemble conceptuel de ce type est appelé *ontologie*. Ces classes, propriétés et individus sont *déclarés* dans un langage conforme à la norme XML. La déclaration des éléments se fait de manière séquentielle et OWL prévoit un mécanisme d'extension qui permet aux déclarations ultérieures de modifier ou d'annuler les précédentes.

Cette technologie permet ainsi de définir un ensemble de classes communes à toutes les notices de l'ALW (comme celles présentées sous 2.2), ainsi que de nouvelles classes spécifiques à certaines notices et créées au fur et à mesure de la rédaction (voir sous 3).

2.2. Exemple

Nous avons exposé dans Baiwir et Mazziotta (à paraître) un ensemble de classes, de propriétés et d'individus communs à l'ensemble de l'ALW. Nous ne reprendrons pas ici l'ensemble de ces définitions, mais nous présenterons les principales d'entre elles de manière à illustrer les conventions de déclaration en OWL. Ces conventions exposées, nous ne les expliquerons plus par la suite. Nous ferons également l'économie de la définition exacte de certaines propriétés ; il nous semble que la compréhension de l'exposé n'en pâtira pas.

Dans l'ALW, les formes dialectales sont localisées, c'est-à-dire qu'une relation de *détermination* unit la forme à sa localisation. Il faut donc définir une classe des *formes* et une classe des *localisations*, ainsi qu'une propriété unissant ces deux classes. Les formes relevées

pour chaque notice sont autant d'individus actualisant la classe des *formes* et entrant en relation avec d'autres individus actualisant la classe des *localisations*. Les déclarations de classes sont représentées comme suit :

```
<owl:Class rdf:ID="Forme" />
<owl:Class rdf:ID="Localisation" />
```

Quant à celles des propriétés, elles suivent le schéma :

```
<owl:ObjectProperty rdf:ID="localisation">
 <rdfs:domain rdf:resource="#Forme"/>
 <rdfs:range rdf:resource="#Localisation"/>
</owl:ObjectProperty>
```

Soit : « La classe Forme et la classe Localisation existent dans cette ontologie ; une propriété localisation peut les unir »[5]. Schématiquement :

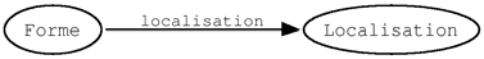


Fig. 1 : Propriété localisation, unissant la classe Forme à la classe Localisation

Les formes dialectales sont alors représentées par des éléments XML Forme, des *individus*. Nous verrons un exemple de ce type dans la section suivante (voir sous 3.1).

3. Description des formes de la notice nèfle

Nous prendrons comme exemple ALW 6 : not. 129 NÈFLE, que nous reproduisons intégralement en annexe. Cette notice traite les formes qui traduisent le français *nèfle* dans les différents dialectes belgo-romans[6]. Nous commencerons par présenter les formes du tableau et la manière de les enregistrer en OWL (3.1) avant de montrer comment il est possible de les catégoriser en fonction de faits phonétiques (3.2).

3.1. Relevé

Pour le rédacteur, l'analyse commence par un *relevé de formes*, « qui consiste [...] à dresser l'inventaire de toutes les formes et de leurs localisation. » (Boutier, à paraître : §1.3). Ces données sont enregistrées sur des fiches de papier par Jean Haust et les continuateurs de l'enquête. À ce stade, le travail consiste déjà en une harmonisation et en une simplification des transcriptions impressionnistes des enquêteurs[7].

En suivant les conventions présentées sous 2.2, nous pouvons décrire la première forme du tableau de la notice « nèfle » (*mèsp*, localisée à Ch ´54, c'est-à-dire dans la localité de Farciennes) sous la forme d'un enregistrement XML identifié par le contenu d'un attribut rdf:ID :

```
<Forme rdf:ID="f1">
 <formePhonique rdf:resource="#FP_mèsp" />
 <sens rdf:resource="#S_nèfle"/>
```

[5] En réalité, la relation localisation est orientée et définie comme une sous-propriété d'une autre propriété ayant pour nom détermination (Baiwir et Mazziotta, à paraître: §3.1).

[6] Notre analyse est donc l'étude d'une notice achevée dans le but d'en comprendre le processus de rédaction. Marie-Guy Boutier, qui a rédigé la notice,nous a aidé dans cette tâche.

[7] Sur l'impressionnisme dans la transcription phonétique, voir l'introduction de Louis Remacle au premier volume de l'ALW (en partic. p. 18 et 19).

```
 <catégorisation rdf:resource="&alw;#C_nf"/>
 <localisation rdf:resource="&alw;#L_Ch54"/>
</Forme>
```

Les balises Forme rassemblent un ensemble de propriétés dont les noms sont explicites. À l'exception de formePhonique[8], ces propriétés correspondent aux *déterminations* définies dans Boutier, à paraître : §1 (voir sous 0.2 ci-dessus). Ces propriétés mettent la forme dialectale en relation avec d'autres individus, appartenant aux classes FormePhonique, Sens, etc., qui sont définis ailleurs dans l'ontologie (Baiwir et Mazziotta, à paraître : §3.1) et mentionnés comme valeurs dans l'attribut rdf:resource. Schématiquement :

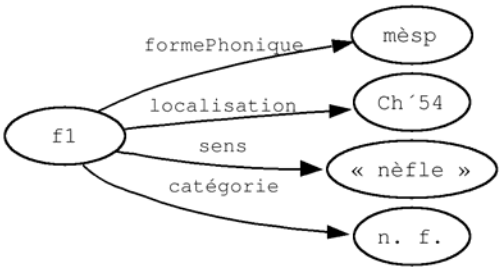


Fig. 2 : Description des formes dialectales

Idéalement, les déterminations de sens et, dans une moindre mesure, de catégorie, sont communes à toutes les formes d'une notice.

3.2. Description phonétique

Du fait de la combinaison des sons, chaque forme est le témoin d'un ensemble d'aboutissements phonétiques. La description de ces faits n'est généralement pas automatisable : elle ne peut être déduite des différentes formes que par une étude historique. Ainsi, en observant le relevé-, le rédacteur remarque en effet qu'ils diffèrent phonétiquement par quatre phénomènes. Premièrement, la consonne initiale est soit *m-*, soit *n-*. Deuxièmement, la voyelle accentuée est le plus souvent *-è-* ou *-ɛ-* (plus rarement *-e-*, *-é-* ou *-ẽ-*). Troisièmement, le groupe consonantique *-spl-* (chute de la post-tonique) évolue en *-sp*, *-spl*, *-ps*, *-s*, *-š*, *-p*, *-pl*, *-t*, *-l*, *-fl* et *-f*. Grâce à la phonétique historique, il est possible de construire la progression *-spl* > *-sp* > *-s* (variante *-š*) ou *-p* (variante *-t*), parallèle à **-sl* > *-l* (champenois). Le rédacteur voit également que *-fl-* > *-f-*. Enfin, certaines formes rares se terminent par une voyelle.

Tous ces phénomènes phonétiques n'ont pas le même statut. Dès l'étape du relevé, le rédacteur peut décider que certaines oppositions ne sont *pas intéressantes*. Ainsi, dans la notice que nous étudions, la variation du timbre de la voyelle accentuée et la présence ou non d'une voyelle à la suite du groupe consonantique sont relevées, mais ne conditionnent pas la structure de la notice.

L'ontologie de l'ALW définit une classe Phonétique (sous-classe de la classe Forme), qui correspond à un ensemble de formes ayant une particularité phonétique commune. Il est possible, en fonction des besoins rédactionnels propres à chaque notice, de définir des sous-classes qui lui sont spécifiques ; par exemple, la classe des formes commençant par une bilabiale :

[8] Cette propriété est opératoire et n'est ajoutée dans l'ontologie que pour des raisons techniques.

```
<owl:Class rdf:ID="P_m-">
 <rdfs:comment>Classe des formes qui commencent par une
bilabiale</rdfs:comment>
 <rdfs:subClassOf rdf:resource="&alw;Phonétique"/>
</owl:Class>
```

Pour que la traduction en OWL soit complète, il faut ajouter que chacune de ces classes est strictement disjointe des autres classes avec lesquelles elle est en distribution complémentaire sur l'axe diatopique ; par exemple, une forme ne peut commencer simultanément par *n-* et par *m-*[9]. Le rédacteur peut alors créer dynamiquement une nouvelle classe P_n- (classe des mots commençant par *n-*) et compléter la description de la classe P_m- en y faisant directement référence (attribut rdf:about) :

```
<owl:Class rdf:about="#P_m-"> <owl:disjointWith rdf:resource="#P_n-"/>
</owl:Class>
```

Une fois les différentes sous-classes de FormePhonique décrites, chaque individu Forme peut être spécifié par la relation type, qui lie la forme décrite à sa description phonétique (le second type rassemble les formes dont le groupe consonantique complexe est *-sp*). À nouveau, OWL emploie un attribut rdf:about :

```
<Forme rdf:about="#f1">
 <rdf:type rdf:resource="#P_m-"/>
 <rdf:type rdf:resource="#P_-sp"/>
</Forme>
```

Ainsi, la forme identifiée par f1, soit *mèsp*, localisée à Ch ´54, commence par une bilabiale *m-* et comprend un groupe consonantique complexe *-sp*. Le typage par l'attribut rdf:type explicite l'analyse phonétique. D'un point de vue technique, la précision progressive de l'individu à l'aide d'un système de références et de pointeurs est confortable et permet de séparer les données de leur première catégorisation.

Toutes les formes relevées peuvent être décrites suivant les mêmes conventions d'encodage. Cela fait, nous avons sous la main un catalogue de formes dialectales localisées et décrites du point de vue de leur forme phonique. Il nous reste à organiser les informations que ce catalogue représente de manière à l'expliquer ou à le rendre comparable à d'autres notices[10].

4. Groupement des formes

L'intérêt des faits phonétiques ne se limite pas à témoigner de la variation diatopique : ils peuvent de plus être la marque de différences d'ordre étymologique (4.1). D'autre part, pour qu'ils soient appréhendables, les faits phonétiques doivent être hiérarchisés (4.2).

4.1. Classement étymologique

Pour les mots signifiant « nèfle », il est possible de montrer par analyse linguistique que les formes dont le groupe consonantique a abouti à *-fl ou -f* sont empruntées au français et qu'il y a lieu de distinguer deux traditions: 1/ les formes héréditaires issues de lat. *mĕspĭla* ; 2/ les formes empruntées, issues de fr. *nèfle*. On peut donc déclarer deux classes correspondantes,

[9] Il peut arriver que plusieurs *variantes* coexistent en un même point (Boutier, à paraître : §4). Néanmoins, cela ne change rien au fait qu'une même forme ne peut cumuler deux réalisations phonétiques différentes.

[10] Par exemple, la notice ALW 6 : 130, rassemblant les dénominations du néflier, est organisée de la même manière que la notice que nous étudions ; cette homogénéité permet d'évaluer la cohérence des matériaux (et, en l'occurrence, de voir que les deux notices sont concordantes).

qui sont des sous-classes disjointes de la classe Étymon (soit la classe des formes rassemblées par une même tradition étymologique) :

```
<owl:Class rdf:ID="E_mespila">
 <rdfs:label xml:lang="lat">mĕspĭla</rdfs:label>
 <rdfs:comment>Formes issues héréditairement de lat.
<i>mĕspĭla</i></rdfs:comment>
 <rdfs:subClassOf rdf:resource="&alw;Étymon"/>
 <owl:disjointWith rdf:resource="#E_nèfle"/>
</owl:Class>
```

et

```
<owl:Class rdf:ID="E_nèfle">
 <rdfs:label xml:lang="fr">nèfle</rdfs:label>
 <rdfs:comment>Formes issues de fr. <i>nèfle</i></rdfs:comment>
 <rdfs:subClassOf rdf:resource="&alw;Étymon"/>
 <owl:disjointWith rdf:resource="#E_mespila"/>
</owl:Class>
```

Ces déclarations traduisent une relation d'*équivalence* (Boutier, à paraître : §2.1 et 2.2) par laquelle les formes similaires peuvent être groupées, ainsi qu'une relation de *hiérarchie* (*ibid.*), puisque les ensembles contiennent des formes distinctes, mais qui se ressemblent à un niveau supérieur.

L'analyse de f1 peut ainsi être complétée :

```
<Forme rfd:about="#f1"> <rdf:type rdf:resource="#E_mespila"/> </Forme>
```

ou bien cette relation peut être exprimée à un niveau supérieur, celui des classes de formes phoniques (et donc appliquée par déduction aux individus membres de ces classes) :

```
<Class rdf:about="#P_-fl"> <rdfs:subClassOf rdf:resource="#E_mespila"/></Class>
```

et parallèlement pour les autres formes dialectales.

Cette première séparation, qui différencie des traditions, est d'un ordre supérieur à toutes les autres subdivisions basées sur des critères phonétiques (Boutier, à paraître : §2.3). Le rédacteur la détecte à partir des formes et elle lui permet de faire deux groupes.

Si nous observons le texte fourni en annexe (v. 7.1), nous constatons que cette différence de tradition n'est pas explicitée dans le tableau des formes. Néanmoins, les traitements ont été ordonnés de manière à faire apparaître les formes empruntées en dernier lieu ; l'opposition est en outre décrite dans l'introduction. En d'autres termes, même si elle est en quelque sorte cachée, la partition étymologique est sous-jacente à l'élaboration du tableau.

Par ailleurs, il faut déterminer une hiérarchie dans ce qui reste.

4.2. Hiérarchie des critères de classement phonétique

En décrivant isolément les différentes particularités phonétiques retenues, on obtient un potentiel théorique de 2 (traitement de l'initiale) x 9 (traitement du groupe consonantique) = 18 formes. En pratique, seules 12 de ces formes sont attestées[11] : *mèsp*, *mèspl*, *mèps*, *nèsp*, *mès*, *nès*, *mèš*, *mèp*, *nèp*, *nèpl*, *nèt*, *mèl*.

[11] Nous avons noté *è*, quel que soit le timbre de la voyelle accentuée.

La description phonétique des formes mène immédiatement à un classement grossier de celles-ci. Cependant, si l'on tente d'évaluer la structure du classement obtenu, on constate que ce dernier est un *graphe* et non un arbre, certains critères se combinant avec d'autres[12].

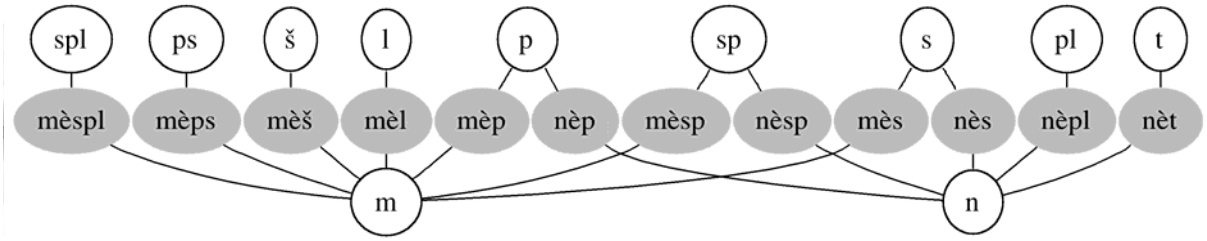


Fig. 3 : Relations entre les formes et les traitements de l'initiale et du groupe -spl-

En effet, le graphe obtenu montre, d'une part, que les formes (en grisé) peuvent être classées selon le traitement de la consonne initiale ou celui du groupe consonantique et, d'autre part, que quelle que soit la première clef choisie pour ce classement, les critères se recoupent.

4.2.1. Choix du sens de lecture

Le rédacteur doit alors choisir la configuration sous forme d'arbre qui explique le mieux les données. Il *privilégie* ainsi une lecture du graphe au détriment d'une autre. Ce passage d'une organisation sous forme de graphe à une organisation sous forme d'arbre est essentiel : il s'agit de livrer au lecteur une interprétation compréhensible des matériaux et non de le perdre dans la richesse de ces derniers. En ce sens, l'ALW est une véritable entreprise d'*édition critique* des matériaux lexicaux. Ce principe de *focalisation* (Boutier, à paraître : §2.4) laisse une marge de manœuvre appréciable au rédacteur, dont le rôle est d'exposer le problème particulier de la notice. La difficulté consiste à choisir correctement la focalisation.

En l'occurrence, il est évident qu'une bipartition est plus facile à appréhender qu'une partition en neuf éléments. D'autre part, on peut juger que le traitement de *-spl-* est complexe et donc digne d'être mis en évidence. Structurer la notice à partir de cette variation permet en outre de la comparer facilement avec d'autres notices comportant le groupe rare *-sp-*, comme ALW 1 : not. 48 GUÊPE. Ces arguments justifient le choix suivant : la hiérarchie de la notice sera fondée en premier lieu sur les aboutissements du groupe consonantique. Cette importance relative des classements, *décidée par le rédacteur* en fonction des matériaux propres à chaque notice, peut être exprimée à l'aide de la relation précéder, qui unit deux classes de formes :

```
<owl:TransitiveProperty rdf:ID="précéder">
 <rdfs:domain rdf:resource="#Forme"/>
 <rdfs:range rdf:resource="#Forme"/>
</owl:TransitiveProperty>
```

Ici, Marie-Guy Boutier a décidé de dire que la classe P_sp- précède la classe P_m- :

```
<owl:Class about="#P_sp-">
 <rdfs:subClassOf>
  <owl:Restriction>
   <owl:onProperty rdf:resource="#précéder" />
   <owl:someValuesFrom rdf:resource="#P_m-" />
 </owl:Restriction>
 </rdfs:subClassOf>
</owl:Class>
```

[12] Les nœuds de ce graphe sont agencés de manière à rendre les connections lisibles.

Traditionnellement, un autre classement peut être suppléé de manière implicite en observant la carte (également fournie en annexe, v. 7.2)[13]. Dans le cas qui nous occupe, ce moyen de visualisation offre une méthode simple et efficace de représenter le traitement de l'initiale.

4.2.2. Typisation

Tous les aboutissements hérités ne sont pas à mettre sur le même pied : certains d'entre eux, nous l'avons vu (v. sous 3.2.), sont liés par une parenté génétique, d'autres sont simplement *ressemblantes*. Ainsi, le rédacteur a regroupé *-spl*, *-sp* et *-ps*, *-pl* et *-p*. À nouveau, on aurait pu procéder autrement et rapprocher *-p* de *-t*, qui en est une variante. Ayant posé ses choix, le rédacteur déclare que les variantes qu'il regroupe sont membres d'ensembles de formes, soit, par exemple[14] :

```
<owl:Class rdf:ID="#P_groupe_-p_-pl">
  <owl:unionOf rdf:parseType="Collection">
  <owl:Class rdf:about="#P_-pl"/>
  <owl:Class rdf:about="#P_-p"/>
 </owl:unionOf>
</owl:Class>
```

Ces classes, identifiées (rdf:ID), mais pas nécessairement nommées (Boutier, à paraître : §2.2) peuvent alors être ordonnées de la même manière que les autres.

5. Ordination des formes et des ensembles

Nous avons fait remarquer *supra* (voir sous 4.1) que bien que le rédacteur n'ait pas hiérarchisé explicitement les formes sur la base de la tradition linguistique dont elles étaient issues, il a placé les formes empruntées à la fin de la liste qu'il a dressée. Cette relation d'ordination (Boutier, à paraître : §3, en partic. §3.1) est structurante[15].

Pour exprimer cette ordination, nous employons également la propriété précéder, sachant que la sémantique de cette dernière est différente quand elle lie deux classes disjointes (une forme ne pouvant être à la fois héréditaire et empruntée[16]) :

```
<owl:Class about="#E_mespila">
 <rdfs:subClassOf>
  <owl:Restriction>
   <owl:onProperty rdf:resource="#précéder" />
   <owl:someValuesFrom rdf:resource="#E_nèfle" />
 </owl:Restriction>
 </rdfs:subClassOf>
</owl:Class>
```

[13] L'emploi d'une carte n'est pas systématique : quand la matière d'une notice est trop complexe ou trop pauvre, la représentation graphique qu'on peut en tirer n'est pas exploitable (ALW 1 : 17).

[14] Ces classes basées sur une typisation peuvent recevoir le label correspondant à la graphie correspondante, reportée en gras dans le tableau de formes. Cette graphie, en orthographe dite *Feller* (du nom de son inventeur), « [g]râce à son caractère approximatif, [...] convient à merveille pour la typisation : elle permet d'inclure sous une graphie "moyenne" des variantes légèrement différentes. » (ALW 1 : 21).

[15] Contrairement à la liste des localisations des différentes formes, qui est ordonnée, mais ne correspond pas à une structuration particulière des données (Boutier, à paraître : §3.3).

[16] Dans le cas contraire, nous aurions créé une classe spécifique pour grouper les cas de « contaminations » entre les formes héréditaires et les formes empruntées.

6. Conclusion

Au bout de son travail, le rédacteur a donc: 1/ décrit chaque forme phonétiquement ; 2/ vérifié si les différences phonétiques n'étaient pas l'indice de différences d'un autre ordre ; 3/ ordonné et hiérarchisé les formes en fonction de ces différences. Chacune de ces étapes est représentée ci-dessus par un ensemble de classes et de propriétés OWL, qui traduisent exactement le cheminement du classement.

Nous obtenons ainsi une présentation *arborescente* des relations entre les formes dialectales, exprimée de manière *linéaire* par le tableau. On voit dans le texte en annexe que les points I à VII opposent des traitements différents du groupe -*spl*- et que certains points peuvent être séparés en *a* et *b* en fonction du traitement de l'initiale[17]. Schématiquement, nous pouvons représenter cette hiérarchie par la figure 4, ci-dessous. La structure (assez simple) de la notice y est explicitée par les ellipses blanches reliées de traits noirs. L'ordination des éléments de même niveau est exprimée par une succession de gauche à droite. Les ellipses grises (reliées aux autres par des traits de la même couleur) représentent les éléments de structuration qui sont : 1/ soit présents dans la notice, mais non explicités par l'appareil typographique (*A* vs *A´*) ; 2/ soit présents à un niveau inférieur dans la hiérarchie choisie et auraient pu être présents à un niveau supérieur :

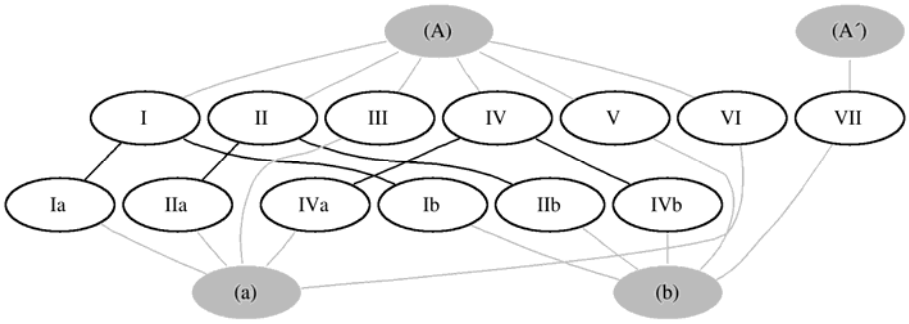


Fig. 4 : représentation de la notice achevée

Nous espérons avoir pu montrer qu'il était possible d'employer OWL pour exprimer la structure des notices de l'ALW. Plus important encore, nous avons voulu présenter comment le langage que nous avons choisi permet de tenir compte de la *démarche* de classement et de la développer *progressivement*. En d'autres termes, il paraît que la technique choisie ne dénature pas le travail des rédacteurs tel qu'il est actuellement pratiqué.

7. Annexe : la notice « nèfle »

7.1. Texte introductif et tableau des formes

Les appels de note renvoient aux notes en fin de notice et non en bas de la page.

129. NÈFLE (carte 59)

Q.G. 92 « *nèfle*, nèflier »

Le classement des continuateurs de *MĚSPĬLA1 (de MĚSPĬLUM) se fonde en premier lieu sur le traitement différencié du groupe complexe -*spl*-. La plupart des issues s'expliquent par une réduction initiale à *sp* (I) ; de là, soit *s* (II), avec variante *š* (III), soit *p* (IV), avec variante *t* (V) ; c'est une réduction initiale à **sl* qui explique le produit champenois *l* (VI). Les formes en *f* (VII) sont empruntées.

[17] D'autres subdivisions minimes sont présentes.

Le tableau oppose en second lieu les formes qui conservent *m* initial (a) à celles où il s'est dissimilé en *n* comme en français (b) ; v. BOURCIEZ 192, I ; FOUCHÉ 551.

La comparaison entre 'nèfle' et 'guêpe' dévoile par ailleurs un jeu d'influences réciproques ; v. ALW 1, c. 48 (introduction) et ci-dessous, n. 4.

◆ ALF 902 ; ALCB 644 ; BRUN., *Enq.* 1056 ; FEW 6/2, 44b MĔSPĬLUM (formes en *m*-), 45b (formes en *n*-).

I. a. +**mèspe**, +**mèsple**...: *mèsp* Ch ´54; Ni 11, 39, ´97, 107; Na 59-99, 107-116, ´118, 127, 130; Ph 6 (ou *nèsp*), 16, 33, 53, 61, 81, 84; Ar 1 (v. *mès*), 2; D 7, 38, 68-81, 94, 96, 110; Ne ´8, 9 | *mɛsp* Na 135; Ph 79, D 36 | *mèspl*2 D ´100; My 1 (?, ou *mès*) || +**mèpse**:3 *mèps* Na 1; D 40, 113; Ve 24 || b. +**nèspe**...: *nèsp* A 2, 7; Mo 58; Ch 28, 72; Th ´51, 53, 62, 72, 82; Ph 6 (ou *mèsp*), ´13, 37, 42, 69, 86 | *nɛsp* Ph 15, 45, 54, 79 (ou *mɛsp*).

II. a. +**mèsse**...: *mès* Ch 33; Ni 2, ´5, 6-´9 (ou *nès*, dans +*poume dɛ̊* ~4), 26 (+*pome dɛ̊* ~[4]), 28, 38, 45 (id.[4]), 61, 73, 98, ´102; Na 6-49, 101, ´120, 129; Ar 1 (Ch. G.; v. *mèsp*); D 15-34, 46, 58, ´74, 84; W ´39, ´45, 63, 66 (peu connu); H 1, 2, 21-49, 53, 67, 69, ´70; L 1, 7 (ou *mèš* ?), 43, 45, 66, 85, 94-113, 116; Ve 8, 31 (ou *mèš*), 32-35, 37 (id.), 38-44; My 1 (ou *mèspl* ?), 3, 6; Ma 1-24, 35-39, 40 (rare), 46; B 9 | *mɛs* Ch 61; Ni 19, 20; Ma 42 (fruit inc. jadis) || b. +**nèsse**...: *nès* A 37; Mo ´5, 9 (ou *nèfl*), 41, 42 (ou *nèp*); S 1-19, 31, 37; Ch 4, 16, ´19, 27, 43, 63, ´64; Th 5, ´9, 24-46, 54, 64; Ni 1, 6-´9 (ou *mès*, dans +*poume dɛ̊* ~[4]), 33, ´35, 36, 72, 90, 112; D ´91 | *nɛs* S 29, 36; Ch 26; Th ´2.5

III. +**mèche**:**6** *mèš* H 50, 68; L 7 (ou *mès* ?), 29, ´32, 35, 61, ´75, 87, ´90, ´99, 114; Ve 1, 6, 26, 31 (ou *mès*), 37 (id.).

IV. a. +**mèpe**: *mèp* To 7 (dans *pɛ̃ d* ~, *pɛ̃° d* ~[4]) || b. +**nèpe**, **-ple**...: *nèp* To 1, 27, 37 (ou +*pin d'* ~[4]), 39, 58, ´71, 73 (enq. compl., ou *nèfl*), 94; A 1, 12, ´13, ´18, ´20, 28, ´40, 44-60; Mo ´29, 37, 42 (ou *nès*), 44, 64, ´75, ´77, 79; Th 73; Vi 6, 8, ´21 | *nep* To 78 | *nép* To 99 | *nèpl* To 48; A ´10 | *-plœ* To 28 |*-plœ̀* To 13 | «*nèple*» To 43 | *nɛplœ* To 24 | « *néple* » To ´36 | *nɛ̃plœ* To ´57.

V. +**nète**:7 *nèt* Mo 1, 23.

VI. +**mèle**:8 *mèl* D ´130, 136, ´141; Ne 51, 65, 69; Vi 2.

VII. +nèfe, -fle...: *nèf* No 1, 3; Mo 9 (ou *nès*), 17; S 10 (J.H., ou *nès*); Th 14; Ph ´11; D 101, 123; L 101 (ou *mès* selon enq.); Ma 43; Ne 60; Vi 18, 37 («fruit inc. ici; on dirait ~»); Vi 43-46 (enq. compl., mais inc. J.H.) | *nèfl* To 73 (J.H., ou *nèp*); Mo 20;[7] D ´130 (néol., v. *mèl*), ´141 (id.); Ne 49.9

Notes

1. Le type manque dans une grande partie de la province de Luxembourg et de l'arrondissement de Waremme.

2. Forme confirmée pour My 1 par VILL. +*mèsple* (à comparer avec SCIUS +*mèpse*). Ajouter PIRS. +*mèsple*, ou +*mèsse*, ou +*mèpse* (comparer EH +*mèspe*). – La conservation de *spl* à la finale s'explique-t-elle par l'influence du dérivé +*mèsplî* etc. 'néflier' (not. 130)? La notation est-elle déficiente, comme le pense REMACLE, *Étym. et phon.* 198, n. 17 ?

3. Forme résultant de la métathèse des consonnes *s* et *p*; comparer +*wèpse*, +*wapse*, variantes de +*wèspe* et de +*waspe* (ALW 1, c. 48 GUÊPE). Pour les attestations des glossaires dialectaux, v. n. 2.

4. Littéralement 'pomme de nèfle'. Aux formes de l'EH, ajouter MAES +*(pin d') mèpe* ou +*nèpe*, BOURG. +*pôre* ['poire'] *de mèpe*. En ces deux pts (To 6, 7), +*mèpe* est aussi la forme locale de ˹guêpe˺ (résultant de l'influence de ˹nèfle˺ sur ˹guêpe˺), et il est possible que les syntagmes soient réinterprétés d'après ce second sens. – V. aussi ˹pomme de néflier˺, ci-dessous, n. 9.

5. Ajouter, d'après DEPR.-NOP., le redoublement enfantin +*nènèsse*, synon. de +*nèsse*.

6. Comparer +*wèche* 'guêpe', variante de +*wèsse* (ALW 1, c. 48 sous III b).

7. +*nète* est absent des glossaires montois; v. DELM. +*nèpe*, +*nèple*, SIG. +*nèpe*. – À Mo 20, +*nèpe* 'nèfle' (q. 135-136); v. q. 152 PRUNELLE, n. 13.

8. Ajouter MASS. +*mèle*.

9. Réponses peu sûres (fournies avec point d'interrogation). 1° +*wèspe* Ph ´47: corruption de +*nèspe* ou de +*mèspe* sous l'influence de +*wèspe* 'guêpe' (v. introduction et n. 4, 5)?; – 2° +*blète* Vi 25: altération de +*nèpe* sous l'influence de {blette} (la nèfle se consomme blette)? – Périphrases: {pomme de néflier} Ni 85; W 63 (ou +*mèsse*); ˹poire --˺ My 4. – Réponse approximative: +*tchitche* Ni 17 (+*lɛ̊ mèsplî pwate dès ~s*; on les met sécher au four, mais on les appelle déjà ~*s* avant); proprement 'fruit tapé', v. ALW 4, 325 et vol. ultérieur. – Sens dérivé de ˹nèfle˺: +*nèsse* 'petite pomme avortée' Mo ´5; v. d'autres mentions à la not. 96 AVORTONS DE FRUITS, *T.*

7.2. Carte

Voir à la suite des références.

Références

ALW 1 = Remacle L. (1953). *Atlas linguistique de la Wallonie. Tome 1. Introduction générale. Aspects phonétiques*. Vaillant-Carmanne.

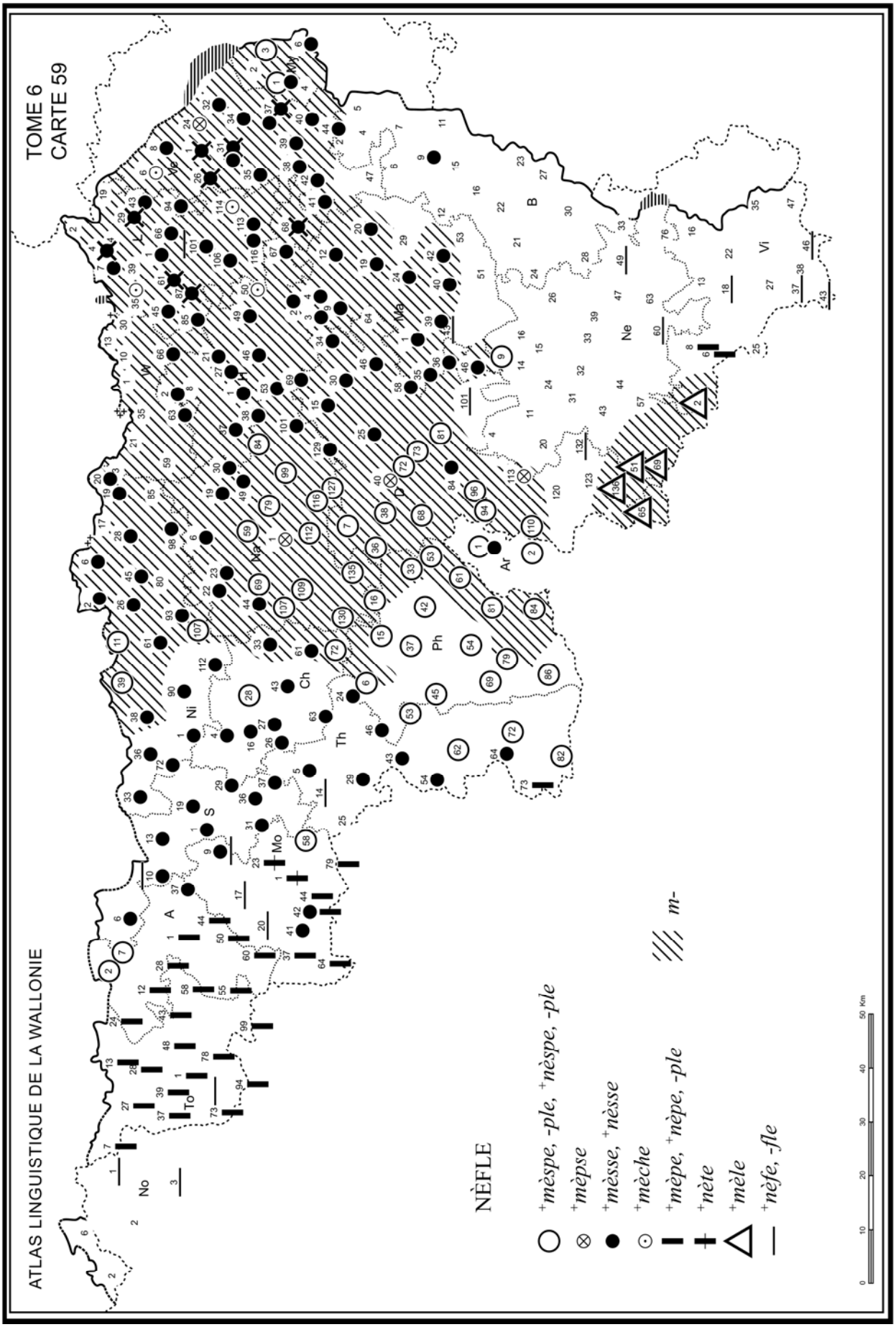
ALW 6 = Boutier M.-G., Counet M.-T., Lechanteur J. (2006). *Atlas linguistique de la Wallonie. Tome 6. La terre, les plantes et les animaux (1re partie)*.Université de Liège.

Baiwir E. and Mazziotta N. (À paraître). Modèle OWL des notices de l'Atlas Linguistique de la Wallonie. Préalables à l'informatisation de l'ouvrage. Proposé au *Thirteenth International Conference on Methods in Dialectology, 4-8 August 2008* [disponible sur demande aux auteurs].

Boutier M.-G. (À paraître). Cinq relations de base pour définir la structure de l'*Atlas linguistique de la Wallonie*, *Estudis romànics*.

Counet M.-T. (1990). *Mots et choses de Wallonie. Aspects du lexique dialectal de nos régions*. Société de Langue et Littérature Wallonnes.

OWL-REF = Bechhofer S., van Harmelen F., Hendler J., Horrocks I., McGuinness D. L., Patel-Schneider P. F. et Stein L. A. *OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004*. [= http://www.w3.org/TR/owl-ref/].


ATLAS LINGUISTIQUE DE LA WALLONIE
TOME 6
CARTE 59
NÈFLE
+mèspe, -ple, +nèspe, -ple
+mèpse
+mèsse, +nèsse
+mèche
+mèpe, +nèpe, -ple
+nète
+mèle
+nèfe, -fle
m-
0 10 20 30 40 50 Km