



Proceedings of the XVI EURALEX International Congress:

The User in Focus

15-19 July 2014, Bolzano/Bozen

Informatiser le *Französisches etymologisches Wörterbuch*: la nécessaire prise en compte de l'utilisateur

Pascale Renders, Esther Baiwir
F.R.S-FNRS/Université de Liège
pascale.renders@ulg.ac.be, ebaiwir@ulg.ac.be

Résumé

« Know your user » (Atkins & Rundell 2008 : 5). Ce conseil ne s'applique pas uniquement à la conception et la rédaction d'un nouveau dictionnaire. La transformation d'un dictionnaire imprimé en dictionnaire électronique peut également bénéficier d'une étude réévaluant les parcours de consultation suivis par les utilisateurs et les difficultés qu'ils rencontrent. Dans une étude préalable à l'informatisation du *Französisches Etymologisches Wörterbuch* de Walther von Wartburg, la prise en compte du point de vue des utilisateurs a mis en évidence deux facettes de l'ouvrage, vu tantôt comme un recueil de monographies, tantôt comme un thesaurus. Après un résumé de l'avancement du projet (qui est maintenant dans sa phase de production), cette communication expose la façon dont les deux visions ont influencé la modélisation informatique du discours lexicographique, permettant ainsi de résoudre une grande partie des problèmes rencontrés par les utilisateurs et ouvrant la voie à une mise à jour de l'ouvrage.

Keywords: FEW; digitalization; user perspective

1 Introduction

Il est aujourd'hui admis que le concepteur d'un dictionnaire doit d'abord définir précisément le public auquel il s'adresse:

[...] the most important single piece of advice we can give to anyone embarking on a dictionary project is : know your user. [...] This doesn't imply a superficial concern with 'user-friendliness', but arises from our conviction that the content and design of every aspect of a dictionary must, centrally, take account of who the users will be and what they will use the dictionary for. (Atkins & Rundell 2008 : 5)

Cette remarque s'applique-t-elle aussi pour un dictionnaire existant, déjà publié, qu'on voudrait transformer en dictionnaire électronique ? Si l'objectif de l'informatisation est d'augmenter les potentialités de consultation et de résoudre des problèmes d'utilisation, ne faut-il pas avoir une idée précise de l'identité des utilisateurs du dictionnaire, de ce qu'ils y cherchent, de la façon dont ils le consultent, des difficultés qu'ils rencontrent et des fonctionnalités qu'ils voudraient y trouver ? Plus générale-

ment, l'utilisation effective de l'ouvrage diverge-t-elle de ce qui était prévu lors de la conception du projet initial et peut-elle être améliorée par le changement de medium ?

Ces questions ont été le point de départ d'une étude qui s'interrogeait sur les possibilités d'informatisation du *Französisches etymologisches Wörterbuch* de Walther von Wartburg (ou FEW), dictionnaire de référence en linguistique historique française et romane. Cette étude a été achevée en 2011 ; l'informatisation du FEW est en cours depuis octobre 2012. Après un résumé du projet et de son avancement, nous proposons d'exposer comment le point de vue de l'utilisateur a modifié l'analyse des structures du dictionnaire et comment ce changement de perspective a été pris en compte dans l'informatisation proprement dite.

2 Le projet d'informatisation du FEW

L'idée d'informatiser le FEW vient de ses utilisateurs. Le premier fut Wooldridge, déjà en 1990 (1990 : 239) puis en 1998 : « [l]a seule façon de mettre au jour ce qui concerne le français du XVI^e siècle dans le FEW serait d'informatiser les 25 volumes... puis de les interroger à partir de repères comme « fr. », « mfr. », « 16^e s. », etc. » (Wooldridge 1998 : 211). Il suffit d'avoir consulté une fois l'ouvrage pour comprendre l'intérêt à la fois de celui-ci et de son informatisation. Le FEW, rédigé de 1922 à aujourd'hui (l'équipe de rédaction s'attelle depuis quelques années à la refonte du premier volume, cf. www.atilf.fr/few), est un dictionnaire de référence en linguistique romane, contenant le lexique de tous les parlers galloromans (français, francoprovençal, occitan, gascon et dialectes). Il est néanmoins sous-exploité, en raison de la complexité de ses structures (cf. Büchi & Chambon 1995). Son informatisation, ardemment souhaitée par la communauté scientifique, est censée à la fois résoudre les problèmes d'accessibilité de l'ouvrage et permettre sa mise à jour ; toutefois, un contenu dense et l'existence de nombreux caractères spéciaux non Unicode (plus d'une centaine) rendaient, jusqu'il y a peu, le projet utopique. Après quelques tentatives partielles, une étude de faisabilité fut entamée, financée par le laboratoire ATILF à Nancy (où est hébergée la rédaction du FEW, cf. www.atilf.fr), par la Fondation FEW (Suisse) et, en majeure partie, par l'Université de Liège en Belgique (bourse de doctorat).

Cette étude fut concluante : l'informatisation des 25 volumes du FEW était non seulement souhaitable, mais possible, sous la forme d'un balisage XML inséré a posteriori dans le discours lexicographique (Renders 2011). La réussite de l'entreprise nécessitait toutefois que soient pris en compte trois contraintes fortes, parmi lesquelles l'obligation de respecter les structures du dictionnaire, y compris dans leurs incohérences et leurs défauts, avec l'interdiction formelle de réécrire les articles pour normaliser le tout. La deuxième contrainte était la nécessité de pouvoir automatiser complètement l'insertion du balisage. La troisième contrainte, enfin, consistait à s'assurer que le résultat de l'informatisation répondrait effectivement aux attentes des utilisateurs, notamment en résolvant les divers problèmes d'accès auxquels ils se heurtaient. La prise en compte de ces trois contraintes mène en pratique à l'élaboration d'un compromis qui, seul, permet une conclusion positive.

Sur la base de la méthodologie mise au point dans cette étude, l'informatisation du FEW se déroule en trois phases :

- (1) l'acquisition du texte des 25 volumes, accompagné d'un balisage typographique minimal ;
- (2) le balisage XML complet des informations lexicographiques, de façon totalement automatisée, à l'aide d'un logiciel construit à cet effet ;
- (3) la mise en ligne des articles balisés et leur exploitation via une interface de consultation.

Ces trois phases (développées plus longuement dans Renders à paraître) comportent quelques particularités par rapport à d'autres projets d'informatisation. Pour la première phase, une saisie manuelle du texte a été préférée à une numérisation, cette dernière s'étant avérée peu fructueuse en raison des nombreux caractères spéciaux non reconnus par les logiciels OCR. Une solution de double saisie, déjà utilisée pour d'autres dictionnaires (par exemple le *Deutsche Wörterbuch* des frères Grimm, cf. dwb.uni-trier.de/de/die-digitale-version), a été proposée et apportée par le Center for Digital Humanities de Trèves, qui possède une expertise reconnue dans ce domaine (cf. Kompetenzzentrum.uni-trier.de).

En ce qui concerne le balisage XML, il a pour particularité d'être pensé selon la perspective de l'utilisateur, contrairement par exemple au balisage du *Trésor de la Langue Française informatisé* qui fut automatisé selon les structures du dictionnaire uniquement (cf. Dendien & Pierrel 2003). Cette particularité le distingue également des dictionnaires conçus dès le départ dans une perspective électronique, ces derniers présentant généralement un balisage pensé pour les besoins de leur rédaction. L'automatisation du balisage et sa vérification s'effectuent à l'Université de Liège. Il est prévu que le balisage inséré soit plus tard converti au standard TEI (cf. www.tei-c.org). Le résultat du processus est la création d'articles au format XML qui pourront être exploités sous la forme d'une base de données.

La mise en ligne finale des articles informatisés nécessite quant à elle l'affichage de caractères spéciaux non standards (non Unicode) et, de ce fait, la création d'une police de caractères spécifique comprenant la totalité de ceux-ci. Seule la phase de mise en ligne est directement concernée par cette police : les phases précédentes nécessitent certes la reconnaissance de tous ces caractères spéciaux (sous la forme de codes ou d'entités XML), mais pas leur affichage. L'Atelier National de Recherche Typographique (www.anrt-nancy.fr) a proposé de créer cette police, tandis que l'ATILF se charge de développer l'interface d'interrogation.

Les trois phases sont successives, mais ne requièrent pas obligatoirement le traitement de la totalité du FEW à chaque étape. L'informatisation peut s'effectuer article par article. Actuellement (juillet 2014), trois des 25 volumes sont en cours de traitement. Ces volumes seront très certainement interrogeables en ligne avant que d'autres volumes ne soient saisis : la première phase est, en effet, la plus coûteuse et dépend donc fortement des financements apportés. En raison de cet obstacle financier à une informatisation rapide, il a parallèlement été décidé de mettre le FEW à la disposition de tous en mode image. Les 25 volumes sont accessibles depuis février 2014 à l'adresse <https://apps.atilf.fr/lecteurFEW>. Une possibilité d'interrogation minimale (par étymons et par lexèmes) de ces images est

prévue, en attendant l'interface d'interrogation complète qui accompagnera la mise en ligne du FEW en mode texte.

3 L'utilisation du FEW

Les difficultés d'utilisation étant la raison principale du projet, il nous semblait évident que l'avis de l'utilisateur était à prendre en compte dès le départ, c'est-à-dire non seulement lors du développement de l'interface de consultation (phase 3), mais aussi dans la définition du balisage à insérer dans le dictionnaire (phases 1 et 2).

Afin de rencontrer au mieux les besoins des utilisateurs, il était d'abord nécessaire de les connaître, suivant le conseil donné par Atkins & Rundell (2008 : 5). Plusieurs questions se posaient, concernant d'abord l'utilité du dictionnaire, ensuite les parcours de consultation actuellement suivis dans la version imprimée du FEW et les problèmes rencontrés. Enfin, il s'agissait de s'interroger sur les parcours à mettre en place dans la version électronique pour répondre à ces problèmes. Les attentes des utilisateurs devaient, rappelons-le, dialoguer avec deux autres contraintes : l'obligation de respecter les structures du dictionnaire – c'est-à-dire le produit lexicographique tel qu'il a été pensé lors de sa rédaction et présenté dans la version imprimée – et la nécessité de pouvoir automatiser le balisage XML. Des demandes inconciliables avec ces contraintes seraient d'emblée soit rejetées, soit revues de façon à élaborer un compromis réaliste.

3.1 Le FEW, son utilité, ses utilisateurs

Si l'on en croit son titre, le FEW est un dictionnaire étymologique du français, ce qui pourrait faire croire qu'il est essentiellement utilisé pour connaître l'étymon des lexèmes de la langue française. En réalité, le titre de l'ouvrage est réducteur (cf. Büchi & Chambon 1995 : 947-948). D'une part, l'étymologie-histoire pratiquée par le FEW le mène à donner davantage d'informations que les autres dictionnaires étymologiques. Le FEW présente en effet une étymologie intégrante (cf. Malkiel 1976), c'est-à-dire que l'information étymologique représente le critère organisateur des données. La conséquence est une structure complexe à plusieurs niveaux (super-, macro-, micro- et infrastructure, cf. Büchi 1996 : 5-6). D'autre part, le domaine couvert par le FEW dépasse la langue française pour embrasser de façon presque exhaustive la totalité des lexèmes du domaine galloroman. Il s'agit donc d'un ouvrage de référence pour tous les parlers et dialectes concernés.

L'étendue du domaine linguistique pris en compte explique que chaque lexème soit associé dans le FEW à une étiquette géolinguistique, qui précise l'état de langue (ancien français, moyen français, français moderne ; ancien gascon etc.) ou le dialecte (lorrain, champenois, picard etc.) auquel il appartient. Des références bibliographiques complètent et précisent la chronologie suggérée par l'information géolinguistique. Le FEW sert donc prioritairement à étymologiser, localiser et dater un lexème

dans un sous-domaine linguistique, même si d'autres utilisations sont possibles, par exemple pour connaître le sens d'un mot, sa graphie, sa forme phonique ou sa catégorie grammaticale.

La nature des données contenues dans le FEW, à savoir le lexique des parlers du domaine galloroman, explique qu'il soit utilisé en linguistique historique, dans les études concernant le lexique du français et des autres langues ou dialectes du domaine galloroman. Il est systématiquement utilisé par les étymologistes des autres langues romanes, ainsi que des langues non romanes. De manière générale, le FEW constitue une référence pour l'étude historique de toute langue qui a été en contact étroit avec le français. Mais pour incontournable qu'elle soit, cette référence est toujours un simple outil au service du chercheur, dont l'objet d'étude n'est évidemment jamais le FEW en lui-même. Ainsi, dialectologues, philologues, éditeurs, lexicographes consulteront avidement le FEW, mais pour mieux construire leur objet propre — nous y reviendrons. Diverses catégories d'utilisateurs consultent donc l'ouvrage, avec des besoins variés et avec des ressources différentes face à la complexité du discours lexicographique. Ce sont majoritairement des spécialistes en leur domaine, mais une partie est constituée par les étudiants et par un public d'amateurs. Tous, y compris les spécialistes, rencontrent des difficultés de consultation et souhaitent une informatisation rapide de l'ouvrage.

3.2 Parcours de consultation et de lecture

L'avis des utilisateurs a pu être recueilli de diverses manières, d'abord via les publications scientifiques des disciplines concernées (voir par exemple Rey 1971 : 103-104 ; Roques 1991 : 94 ou encore Wooldridge 1998 : 211 ; cf. Renders 2011 : 8-15), ensuite via la diffusion d'un questionnaire au sein de la communauté internationale des chercheurs en linguistique française et romane lors du *XXVe Congrès International de Linguistique et Philologie Romanes* (cf. Renders 2010 ; pour les résultats, Renders 2011) et, enfin, via de nombreuses rencontres individuelles. Il a ainsi été possible d'obtenir un aperçu des pratiques actuelles d'utilisation du FEW et, dans un second temps, de connaître les souhaits des utilisateurs dans l'optique d'un FEW informatisé, souhaits en relation étroite avec les difficultés qu'ils rencontrent dans la consultation de la version imprimée. Les attentes exprimées sont révélatrices de la façon dont les utilisateurs voudraient exploiter le FEW et, plus généralement, de la façon dont ils perçoivent l'ouvrage et ses structures.

L'analyse des comportements d'utilisation du FEW a distingué deux activités distinctes et successives lors de l'utilisation du FEW : d'une part, la consultation, opération consistant à repérer dans le dictionnaire l'endroit où se trouve l'information que l'on recherche ; d'autre part, la lecture, opération consistant à s'approprier de façon complète l'information recherchée ainsi que l'analyse qu'offre le FEW en rapport avec cette information. Pour chacune de ces deux opérations, l'étude a montré des divergences entre les parcours actuels, permis par la version imprimée, et les parcours souhaités dans une version électronique.

En ce qui concerne la consultation, les souhaits des utilisateurs d'un futur FEW numérique induisent des itinéraires totalement nouveaux par rapport aux itinéraires traditionnels. En effet, dans la version

imprimée, les points d'entrée dans le dictionnaire sont réduits aux étymons-vedettes (à condition de connaître la langue de l'étymon, qui détermine la partie superstructurelle et donc le volume à consulter) et aux lexèmes (à condition soit d'avoir une idée de leur étymon, soit de trouver ces lexème – ou des lexèmes apparentés – dans les index du FEW qui ne sont nullement exhaustifs : cf. ATILF 2003, qui remplace les divers index situés en fin de volume). En pratique, l'utilisation du FEW s'apparente souvent à un jeu de piste.¹ Dans la perspective d'une version électronique, les points d'entrée ne seraient toutefois plus réduits aux seuls lexèmes et lemmes (qui deviendraient en outre plus facilement repérables), mais s'étendraient fructueusement à tout type d'information présent dans le discours lexicographique (étiquettes géolinguistiques, sources bibliographiques, dates etc.). Ce mode de consultation « transversale », qui mène à plusieurs endroits dans le dictionnaire, est impossible dans la version papier du FEW, mais très attendu dans l'optique de son informatisation.

En ce qui concerne la lecture, l'étude a mis en évidence la complexité des itinéraires traditionnels, due à la nécessité, pour s'approprier l'analyse approfondie des données fournies par l'ouvrage, de mettre ces données en relation et en contexte à plusieurs niveaux. Par ailleurs, un aller-retour est constamment requis entre le dictionnaire et son *Complément*, qui explicite les nombreuses abréviations géolinguistiques et bibliographiques propres au FEW. Il est intéressant de constater que ces parcours présentent des variantes selon l'expérience qu'a l'utilisateur des structures du dictionnaire, selon ses compétences en linguistique française et dialectale et selon son besoin d'explicitation des abréviations. Les difficultés d'accès de l'ouvrage expliquent que de nombreux souhaits soient émis dans l'optique d'une informatisation, tels que la résolution des nombreuses abréviations, l'explicitation des sigles bibliographiques et des sources, la traduction des termes allemands ou, encore, la mise en évidence du plan des articles longs. La plupart de ces besoins se résolvent par des mises en relation (avec le commentaire de l'article, avec le *Complément*, avec des outils externes) qui, certes, sont possibles dans la version imprimée du FEW, mais seraient grandement facilitées par une informatisation de son contenu. Il ne s'agit donc pas de modifier les itinéraires de lecture classiques de la version imprimée, mais de faciliter la mise en relation de données qui, dans le discours lexicographique, ne sont pas situées côte à côte. Ce faisant, on ouvre la voie à des parcours de lecture hypertextuels qui n'étaient pas identifiés comme tels dans l'analyse des comportements d'utilisation du FEW papier, mais qui étaient sous-jacents.

3.3 Deux visions du FEW

L'analyse des parcours effectifs et des parcours souhaités par les utilisateurs a fait apparaître deux modes a priori contradictoires de consultation et de lecture du dictionnaire. Les difficultés de lecture s'expliquent en effet par la vision de l'article du FEW comme un discours construit et structuré, dans lequel chaque information est à mettre en relation avec celles qui l'entourent. Rappelons que les ar-

1 Il faut ajouter à ces difficultés d'accès le problème des classements multiples, cf. Baldinger 1980.

tics du FEW classent et hiérarchisent les données différemment, de façon à retracer l'histoire particulière de chaque famille lexicale, ce qui fait de chaque article une monographie à part entière :

L'ouvrage se présente, en fait, comme un ensemble structuré de monographies, dont la forme lexicographique n'est qu'un auxiliaire au service de la « visée globalisante » (Swiggers 1990 : 347) de Wartburg, qui l'anime et la domine. (Büchi & Chambon 1995 : 952)

Cette particularité explique que les articles du FEW se lisent davantage qu'ils ne se consultent. Il s'agit de ce que nous appelons la *dimension monographique* (ou dimension M) du FEW, qui n'est accessible que par l'opération de lecture. Les demandes, partagées par les utilisateurs, d'un plan de l'article et d'une traduction du commentaire sont tout à fait liées à cette dimension monographique du FEW : elles visent à atteindre plus aisément le classement des données et l'analyse qui en découle.

L'enthousiasme des utilisateurs pour l'informatisation du FEW est toutefois davantage à expliquer par une autre vision de l'ouvrage, que nous appelons la *dimension thesaurus* (ou dimension T) du FEW. Dans cette vision du FEW comme un thesaurus, les utilisateurs sont intéressés par la masse de données qui s'y trouve et par les informations qui sont associées à chaque lexème. C'est le lexème, et non plus l'article, qui constitue leur centre d'intérêt. Le FEW est en effet le seul dictionnaire où se trouvent rassemblés tous les lexèmes des langues et dialectes du domaine galloroman, ce qui en fait un ouvrage des plus précieux dans de nombreuses sous-disciplines linguistiques. Cette dimension thesaurus est à l'œuvre lorsque l'utilisateur imagine des modes de consultation transversale, qui permettraient d'accéder directement à un groupe de lexèmes partageant un point commun malgré leur dispersion lexicographique due à leur appartenance à des familles lexicales différentes. Les besoins de mise en relation avec le *Complément* et avec d'autres dictionnaires, qui permettraient d'accéder directement à l'intégralité des références bibliographiques associées à un lexème, font partie de cette vision du FEW comme thesaurus. Enfin, c'est cette dimension qui explique que le FEW se consulte généralement à partir des lexèmes, alors que les entrées de la nomenclature sont des étymons.

Les deux dimensions dégagées ci-dessus ne sont pas inconciliables, mais étroitement imbriquées et complémentaires dans la construction du discours lexicographique. Si la consultation gagne à être envisagée dans une dimension thesaurus où chaque lexème est individualisé et accessible séparément, la lecture, quant à elle, ne peut s'effectuer que dans une dimension monographique, c'est-à-dire dans une mise en relation et une contextualisation des données.

Ces besoins, exprimés par les utilisateurs, ne sont pas résolus par les itinéraires de consultation et de lecture permis par le discours lexicographique sous sa forme actuelle. Les difficultés d'accès aux données du FEW ont toujours été attribuées à la présentation condensée et hautement structurée du discours lexicographique ; l'étude montre que ces difficultés proviennent également, si pas davantage, du fait que les utilisateurs veulent consulter le FEW dans une optique plus large que celle pour laquelle il a été conçu.

4 La prise en compte de l'utilisateur dans la version électronique

Partant de ce constat, il s'est avéré essentiel de résoudre l'inadéquation entre la conception initiale du dictionnaire et l'utilisation qui en est souhaitée, à la fois en tant que recueil de monographies et en tant que thesaurus. En dimension M, il s'agit de faciliter la lecture de l'article en tant qu'ensemble structuré et autonome et d'optimiser les itinéraires de mise en relation entre les informations (à la fois au sein d'un article et hors article). En dimension T, il s'avère nécessaire d'ouvrir l'accès aux nouveaux itinéraires de consultation attendus par la communauté (consultation transversale). Enfin, les deux dimensions sont concernées par la nécessité de permettre la mise à jour du FEW (intégration des ajouts et corrections apportés ailleurs) et sa mise en réseau avec d'autres ressources lexicographiques. Ces nouveaux parcours doivent être pris en compte non seulement au moment de créer l'interface de consultation, mais aussi dans la modélisation préalable et la formalisation XML du discours lexicographique.

4.1 Dimension monographique : itinéraires de lecture

Rétablir l'autonomie de l'article en dimension M nécessite tout d'abord de permettre son extraction hors du volume imprimé, tout en conservant les informations qui étaient données par le contexte physique et la situation de l'article au sein de l'ensemble. Rappelons que le FEW possède une superstructure divisant l'ouvrage en fonction de la langue d'origine des lexèmes et que les étymons-vedettes ne sont dès lors pas nécessairement partout classés par ordre alphabétique. La solution informatique consiste à « redescendre » au niveau de l'article les informations données dans les niveaux supérieurs (essentiellement le numéro de volume ainsi que la page où débute l'article). En pratique, ces informations sont automatiquement explicitées au début de l'article, dans les attributs des balises XML identifiant l'article et la colonne. Par exemple, la version XML de l'article MASCŪLĪNUS (FEW 6/1, 424b) commence ainsi :²

```
<art book="1" ici="1" id="0" lang="german" type="doc-com" volume="6">
<col pg="424" s="b"/>
```

Afin de faciliter la lecture de l'article en tant qu'ensemble structuré, il a été décidé de proposer au lecteur un plan résumant cette structure. Ce résumé ne consiste pas en une réécriture (impossible à automatiser et donc inconcevable, conformément au respect des trois contraintes précitées), mais, plus simplement, en l'affichage automatique de la première unité lexicale de chaque paragraphe, précédée du marquage alphanumérique situant le paragraphe dans la numérotation microstructurale de l'article (réexplicitée si nécessaire) et, lorsqu'il existe, du marqueur textuel (titre de section) explicitant le

2 L'attribut *ici* (*in-column index*) indique l'ordre de l'article dans la colonne, *lang* la métalangue utilisée dans le commentaire; l'attribut *type* indique si l'article est divisé en une partie documentaire (*doc*) et une partie de commentaire (*com*) (cf. Büchi 1996 : 78).

critère de regroupement des lexèmes au sein du paragraphe. Le plan de l'article CHOCOLATL (FEW 20, 63b-64a) se présente par exemple ainsi dans sa version XML :

```
<!--article map
```

```
1 Mfr. chocholate m. „breuvage fait avec des amandes de cacao“ (1598)
```

```
1 Ablt. — Nfr. chocolatière f. „vase où l'on prépare, où l'on sert le chocolat en boisson“ (seit 1680)
```

```
2 Nfr. chicolate f. „chocolat“ (1658)
```

```
-->
```

Ce plan, affiché en tête d'article, donne immédiatement au lecteur une vision synthétique de la structure de l'article (classement des lexèmes au sein des différentes subdivisions) : ici, trois paragraphes structurés en deux parties numérotées, la première partie reprenant des dérivés (*Ablt.* pour *ableitung*). Ce faisant, il aide à saisir d'un coup d'oeil l'organisation générale de la famille lexicale traitée.

Enfin, les itinéraires de lecture sont également facilités par la création de liens hypertextuels internes et externes. Les liens internes concernent d'une part les notes et appels de note (identifiés et associés de façon à faciliter le passage de l'un à l'autre), d'autre part les références renvoyant, dans le commentaire, au marquage alphanumérique structurant les matériaux. Par exemple, toujours dans l'article CHOCOLATL, le commentaire explique l'origine étymologique de chacune des deux parties numérotées. Tant les marqueurs alphanumériques que les références à cette numérotation (*<pref>*) ont été automatiquement reconnus et balisés, ce qui permettra la création de liens hypertextuels :

```
<pref id="1">1</pref> ist aus <lang>sp.</lang> <form><i>chocolate</i></form> entlehnt [...]. Unklar ist auch das verhältnis von <pref id="2">2</pref> zu <pref id="1">1</pref>. Der erste beleg von <pref id="2">2</pref> kommt von den kleinen Antillen, wodurch wahrscheinlich gemacht wird, dass diese form sich selbständig verbreitet hat.
```

Les liens externes concernent d'une part les renvois à d'autres articles du FEW (mis en oeuvre par le balisage automatique des étymons, volumes et pages), d'autre part les renvois externes au dictionnaire. Le balisage des sigles bibliographiques permet en effet de créer un lien hypertextuel vers leur explicitation (fournie dans la base de données contenant le *Complément* au FEW) et vers la ressource électronique si cette dernière existe.

Ces trois nouveautés apportées par la version électronique (autonomie de l'article, résumé de sa structure, liens hypertextuels) permettent de faciliter et d'optimiser les parcours de lecture du FEW dans sa dimension monographique.

4.2 Dimension thesaurus : itinéraires de consultation

En dimension T, il s'agit en priorité de permettre une consultation du FEW via les lexèmes (unités lexicales) et les informations qui y sont associées. L'autonomie de chaque unité lexicale est rétablie en balisant au sein d'un même élément XML les informations qui la composent (étiquette géolinguistique, signifiant, catégorie grammaticale, définition, références) et en rétablissant celles de ces informations qui seraient implicites, cas de figure très fréquent dans le FEW puisque certaines informa-

tions (étiquette géolinguistique, catégorie grammaticale, signifiant, définition) ne sont jamais répétées si elles ont déjà été citées dans l'unité précédente (cf. Büchi 1996 : 117 et Renders 2011 : 76-81). Le principe de « redescende » des informations s'applique donc également ici. Dans l'article ACCUSATIVUS par exemple (FEW 24, 94b), qui comporte uniquement deux lexèmes *accusatif*, l'un substantif masculin (« cas auquel on met le complément direct »), l'autre adjectif (« qui concerne l'accusatif »), l'étiquette géolinguistique et le signifiant associés au deuxième lexème n'ont pas été répétées. La version XML rétablit ces informations grâce à l'insertion automatique d'une balise <imp>, identifiant le type d'information manquant et son contenu implicite :

```
<unit><imp contents="Mfr." type="geoling"/><imp contents="accusatif" type="form"/><gram>adj.</gram> <def>„qui concerne l'accusatif“</def>
```

Ce balisage permet d'extraire de chaque article toutes les unités lexicales qu'il contient et de leur rendre leur autonomie, indépendamment de leur insertion dans le discours monographique. Une liste de toutes les unités est systématiquement créée pour chaque article ; cette liste attribue en outre à chaque lexème son étymon (étymon-vedette de l'article) et sa référence FEW. Toujours pour l'article ACCUSATIVUS, la liste créée est la suivante :

```
<fiche etymon="accusativus" lang="Mfr." lang="nfr." forme="accusatif" gram="m." def="„cas auquel on met le complément direct““ ref="(seit ca. 1170, EdConf, FrMod 21, 217)" N="FEW 24/1, 94b, ici 1, §1, u1"></fiche>
```

```
<fiche etymon="accusativus" lang="(imp.) Mfr." forme="(imp.) accusatif" gram="adj." def="„qui concerne l'accusatif““ ref="(1380, Aalma 98; Pom 1671-1700; Lar 1866-1948)" N="FEW 24/1, 94b, ici 1, §1, u2"></fiche>
```

Le balisage rend ainsi possible une consultation transversale du FEW via les informations associées aux lexèmes : consultation par parler (étiquette géolinguistique), par catégorie grammaticale, par élément de définition, par signifiant, par référence bibliographique, par étymon. La consultation par critères chronologiques nécessite en outre que les dates correspondant à chaque sigle bibliographique soient explicitées, ce qui s'effectue via le *Complément* au FEW.

Enfin, des consultations à cheval entre la dimension monographique et la dimension thesaurus, à savoir selon la langue d'origine des lexèmes (souvent implicite dans le FEW, car associée à la partie superstructurelle où se trouve l'article) ou selon le type de descendance (emprunts ou lexèmes héréditaires) sont également rendues possibles par le balisage.

4.3 Sortir du FEW

Comme évoqué au point 3.1., les utilisateurs du FEW sont divers, de même que leurs objets d'étude. Le parcours dialectique entre le FEW et ces diverses entreprises peut être illustré par un exemple qui nous est familier, celui de l'*Atlas linguistique de la Wallonie* (ALW). Tout au long de l'analyse des matériaux de l'ALW, les rédacteurs tissent un dialogue entre leur ouvrage et le formidable outil qu'est le FEW

: d'une part, en extrayant de celui-ci les données servant à éclairer les matériaux wallons, d'autre part, en amenant divers compléments, amendements ou ajustements. Au fil de ce cheminement se développe « une véritable réévaluation de l'état de l'art représenté par le FEW » (Chauveau & Buchi 2011 : 12). Si les spécialistes s'accordent à reconnaître un certain intérêt aux compléments apportés par l'ALW au dictionnaire de Wartburg, ceux-ci sont de toute évidence difficilement accessibles. Cette conclusion peut être étendue à d'autres ouvrages, qu'ils soient atlantographiques, lexicographiques ou philologiques. Dès lors, il nous semble que prendre en compte l'utilisateur du FEW passe également par une intégration, à quelque niveau que ce soit, de ses apports à la construction d'un savoir commun. Le balisage de l'œuvre selon les modalités exposées ci-dessus permettra une navigation optimisée et des consultations transversales internes. Ce balisage pourrait ensuite être exploité pour développer des liens externes vers divers projets apparentés, tels que l'ALW pour des apports ponctuels ou des entreprises telles que DEAF, Godefroy, TLFi etc. pour des apports plus systématiques. Ces divers ouvrages intégrant déjà dans leur programme lexicographique des références au FEW, leur mise en réseau nécessite uniquement que le FEW informatisé dispose d'url pérennes qui puissent servir à la fois de référence et de lien hypertextuel. Toutefois, l'intégration des apports externes pourrait aller plus loin qu'une simple mise en réseau et conduire à un FEW évolutif. Cette idée encore utopique est en phase avec la dimension T du FEW, puisqu'elle permettrait des consultations basées sur des critères (géolinguistiques, chronologiques, étymologiques ou autres) qui correspondraient à l'état actuel de la science et non à une version périmée. Le balisage du FEW en dimension T, et plus particulièrement le rétablissement de l'autonomie des unités lexicales, devrait permettre ces consultations sans que la dimension M du FEW n'en soit affectée.

5 Conclusion

Loin de constituer une modélisation théorique et gratuite du dictionnaire, les deux dimensions décelées dans la conception et l'utilisation effective du FEW d'une part, mais également dans l'exploitation que rêvent d'en faire ses utilisateurs, ont très concrètement guidé la réflexion dans le cadre du projet. Nous espérons avoir montré comment ces deux dimensions s'articulent et devront continuer à le faire, de la façon la plus explicite possible, dans la future version électronique.

En outre, nous avons montré en quoi les fonctionnalités permettant l'exploitation du dictionnaire dans ses deux dimensions ne concernaient pas seulement la création de l'interface utilisateur, mais devaient être prévues dès la phase de modélisation du discours lexicographique. Aborder les structures du dictionnaire selon le point de vue de l'utilisateur est la condition nécessaire pour pouvoir rendre compte de la façon dont le FEW est, non plus conçu et rédigé, mais perçu et utilisé.

Bien entendu, c'est au niveau de l'interface, toujours en développement à l'heure où nous écrivons ces lignes, que l'utilisateur prendra pleinement conscience des potentialités des nouveaux parcours qui

lui sont offerts. Gageons cependant que cet avant-goût aiguisera encore davantage l'appétit des chercheurs!

6 Bibliographie

- ALW = Remacle, L. et al. (1953-). Atlas linguistique de la Wallonie. Tableau géographique des parlers de la Belgique romane d'après l'enquête de Jean Haust et des enquêtes complémentaires. Liège: Vaillant-Carmagne.
- ATILF (2003). *Französisches Etymologisches Wörterbuch. Index A-Z*. Paris: Champion.
- Atkins, S.B.T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Baldinger, K. (1980). Etymologies doubles dans le FEW. In H. J. Izzo (éd) *Italic and Romance: Linguistic Studies in Honor of Ernst Pulgram*. Amsterdam: Benjamin, pp. 189-194.
- Büchi, E., Chambon, J.-P. (1995). Un des plus beaux monuments des sciences du langage : le FEW de Walter von Wartburg (1910-1940). In G. Antoine, R. Martin (éd.) *Histoire de la langue française, 1914-1945*. Paris: CNRS Editions, pp. 935-963.
- Büchi, E. (1996). *Les structures du Französisches Etymologisches Wörterbuch. Recherches métalexigraphiques et métalexicologiques*. Tübingen: Niemeyer.
- Chauveau, J.-P., Buchi, E. (2011). État et perspectives de la lexicographie historique du français. In *Lexicographica. International Annual for Lexicography*, 27, pp. 101-122.
- Complément = Chauveau, J.-P., Greub, Y. & Seidl, C. (2010). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes. Supplement zur 2. Auflage des Bibliographischen Beiheftes*. Bâle: Zbinden.
- DEAF = Baldinger, K. et al. (1974-). *Dictionnaire Étymologique de l'Ancien Français*. Québec/Tübingen/Paris: Presses de l'Université Laval/Niemeyer/Klincksieck. Site internet : <http://deaf-server.adw.uni-heidelberg.de>.
- Dendien, J., Pierrel, J.-M. (2003). Le trésor de la langue française informatisé. Un exemple d'informatisation d'un dictionnaire de langue de référence. In *Traitement automatique des langues (TAL)*, 43 (2), pp. 11-37.
- FEW = von Wartburg, W. et al. (1922-2002). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes* (25 vol.). Bonn/Heidelberg/Leipzig-Berlin/Bâle: Klopp/Winter/Teubner/Zbinden.
- Godefroy = Godefroy, F. (1881-1895). *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IXe au XVe siècle* (8 vol.). Paris: Vieweg.
- Malkiel, Y. (1976). *Etymological dictionaries. A tentative typology*. Chicago-London: The University of Chicago Press.
- Renders, P. (2010). L'informatisation du *Französisches Etymologisches Wörterbuch* : quels objectifs, quelles possibilités ? In M. Iliescu, H. Siller-Runggaldier & P. Danler (eds.) *Actes du XXVe Congrès International de Linguistique et de Philologie Romanes* (Innsbruck, 3-8 septembre 2007). Berlin/New York: De Gruyter, vol. 6, pp. 311-320.
- Renders, P. (2011). Modélisation d'un discours étymologique. Prolégomènes à l'informatisation du *Französisches Etymologisches Wörterbuch*. PhD Thesis. Université de Liège, Liège, BE.
- Renders, P. (à paraître). Mise en ligne, mise à jour et mise en réseau du *Französisches Etymologisches Wörterbuch*. In D. Trotter, A. Bozzi & C. Fairon (éds.) *Actes du XXVIIe Congrès international de linguistique et de philologie romanes* (Nancy, 15-20 juillet 2013). Nancy: ATILF.
- Rey, A. (1971). Le dictionnaire étymologique de W. von Wartburg : structures d'une description diachronique du lexique. In *Langue française*, 10, pp. 83-106.

- Roques, G. (1991). L'articulation entre étymologie et histoire de la langue. In *Travaux de linguistique*, 23, pp. 91-95.
- Swiggers, P. (1990). Lumières épistolaires sur l'histoire du F.E.W. : Lettres de Walther von Wartburg à Hugo Schuchardt. In *Revue de linguistique romane*, 54, pp. 347-358.
- TLFi = CNRS/Université Nancy2/ATILF (2004). *Trésor de la langue française informatisé* (cédérom). Paris: CNRS Éditions (site internet : <http://www.cnrtl.fr/definition>).
- Wooldridge, T. R. (1990). Le FEW et les deux millions de mots d'Estienne-Nicot : deux visages du lexique français. In *Travaux de linguistique et de philologie*, 28, pp. 239-316.
- Wooldridge, T. R. (1998). Le lexique français du XVIe siècle dans le GDFL et le FEW. In *Zeitschrift für romanische Philologie*, 114, pp. 210-257.