

L'ALPic en ligne et le projet d'un Atlas pan-picard informatisé (APPI)

0. Introduction

Les atlas linguistiques constituent, à plus d'un titre, des objets atypiques dans le panorama des outils traitant des langues. Ces ouvrages de référence, tellement essentiels pour la dialectologie et pour la linguistique, présentent en effet des matériaux non linéaires, contrairement aux dictionnaires ou aux monographies de parlars. Leur intégration dans le réseau de la lexicographie informatisée s'en trouve sans doute compliquée, mais elle est vitale pour leur exploitation et leur visibilité. Il s'agit aussi de permettre le décloisonnement entre les domaines géolinguistiques, pour peu que les formats définis pour les différents projets de numérisation soient interoperables. C'est ce à quoi travaille l'équipe du projet APPI (*Atlas pan-picard informatisé*), en cours à l'Université de Lille depuis 2018. Ce projet, financé par l'ANR (<https://anr.fr/Projet-ANR-17-CE38-0002>), envisage la création d'une ressource numérique atlantographique pour le picard en particulier.

Dans cette contribution, nous envisagerons les ressources atlantographiques en domaine picard (1), puis nous présenterons le projet APPI tel qu'il était conçu initialement (2), avant d'évoquer l'intégration dans un modèle informatique préexistant (3), les résultats actuels (4) et l'avenir du projet (5).

1. L'atlantographie en domaine picard

Le picard est atypique au sein de la Galloromania, eu égard à sa situation géographique entre deux espaces nationaux, la France et la Belgique, et à son traitement dans trois atlas linguistiques : l'ALF, l'ALPic en territoire français et l'ALW en territoire belge. Ce parler n'a plus fait l'objet d'enquêtes globales depuis les campagnes pour l'*Atlas linguistique de la France* (ALF).

L'ALF constitue la plus ancienne source de matériaux atlantographiques pour le picard. La caractéristique des matériaux recueillis dans le cadre d'atlas est que ceux-ci sont comparables dans les divers points étudiés, car ils ont été récoltés de façon systématique dans une temporalité (relativement) limitée. La première campagne d'enquête, celle de l'ALF, s'est déroulée entre 1897 et 1901, avec un questionnaire qui, au moment des enquêtes dans le nord de la France et la Belgique, comportait environ 1 400 questions. Le territoire qui nous intéresse y est représenté par une quarantaine de points, dont huit en Belgique et les autres en France. Les cartes ont été publiées en volumes entre 1901 et 1910, puis numérisées sous la forme d'images dans le cadre du projet Cartodialect (voir <http://cartodialect.imag.fr/cartodialect> et Brun-Trigaud/Chagnaud/Seffar & Drapeau dans ce volume).

Par la suite, deux projets complémentaires se sont intéressés au domaine picard, l'ALW en Belgique et l'ALPic en France.

L'*Atlas linguistique et ethnographique picard* (ALPic) fut conçu dans le cadre de la campagne des *Nouveaux atlas linguistiques de France par régions*, lancée par Albert Dauzat (voir par exemple Séguy 1973 : 68-sv.). Après une genèse mouvementée (Carton 2019), deux volumes furent publiés en 1989 et en 1997 par Fernand Carton et Maurice Lebègue. Ils contiennent 660 cartes, éditant une partie des matériaux des 1 150 questions de l'enquête effectuée entre les années 1960 et les années 1980. Un troisième tome est en préparation, sous la direction d'Alain Dawson.

Enfin, l'*Atlas linguistique de la Wallonie* (ALW) présente les matériaux d'enquêtes effectuées entre 1924 et 1959 sur l'ensemble de la Belgique romane. Dix volumes ont paru, éditant environ 1 670 notices, soit la moitié des collections. Le picard et le wallo-picard représentent 105 des 342 points d'enquête de cet atlas, les autres illustrant les parlers wallons, champenois et gaumais (v. <http://alw.philo.ulg.ac.be/publications/liste-des-volumes-publies/>).

Du point de vue du traitement des données, dans l'ALF comme dans l'ALPic, les matériaux sont publiés directement sur les fonds

de carte, sans analyse ni classement. Quelques mentions en marge des cartes signalent d'éventuelles difficultés sémantiques ou des particularités des matériaux. Quant à l'ALW, il s'agit d'un atlas explicatif, dans lequel les cartes sont subordonnées au texte. Dans chaque notice, les matériaux sont intégrés dans un cadre historique galloroman, grâce à un dialogue avec le FEW et avec d'autres ressources lexicographiques. Les cartes apparaissent uniquement lorsque la répartition aréologique des matériaux est intéressante. Contrairement aux pratiques françaises, les cartes deviennent donc une illustration des matériaux et non plus le cœur de la microstructure, dans une démarche tirant l'atlantographie vers la lexicographie.

L'ALPic et l'ALW s'appuient donc sur des méthodes d'édition très différentes, l'une purement atlantographique (du côté français), l'autre intermédiaire entre diatopie synchronique et lexicographie historique (du côté belge).

Contrairement à l'ALF, les deux autres atlas n'avaient pas encore fait l'objet d'une digitalisation¹. Le champ de leur numérisation n'était toutefois pas totalement vierge ; parmi les travaux antérieurs examinant les possibilités de l'informatisation de l'ALW, citons la modélisation des relations à intégrer dans un modèle informatique (Boutier 2008 et Baiwir/Renders 2013) ou la proposition d'utilisation de l'Ontology Web Language (Mazziotta 2008 et 2011). Notre projet s'appuie également sur les réflexions autour de l'informatisation du FEW (Renders 2015) et autour des possibilités techniques de la réunion des matériaux de l'ALPic et de l'ALW ainsi que des verrous d'une telle entreprise (Baiwir 2016, Renders/Baiwir & Dethier 2015).

2. Le projet APPI

Le projet initial se décompose en trois étapes (voir Baiwir / Renders 2019), envisagées comme suit :

- rassembler les données des trois atlas en un corpus homogène pan-picard,
- les rendre disponibles en ligne et interopérables avec d'autres projets,
- enrichir les matériaux par une étymologisation systématique et une mise en relation avec le FEW.

1 — L'ALW possède toutefois un site sur lequel des échantillons sont consultables : <https://alw.philo.ulg.ac.be/>.

2.1 *Un corpus atlantographique pan-picard*

La première étape consiste donc à homogénéiser les matériaux d'un point de vue sémantique et graphique et de les rendre accessibles et interrogeables en ligne (<https://anr-appi.univ-lille.fr/>).

Techniquement, la faisabilité d'un rapprochement des données ALF, ALW et ALPic nécessite de prendre diverses décisions. La première consiste à définir une nomenclature pour le nouvel outil. La dimension onomasiologique des trois atlas est conservée ; nous devons donc identifier des notions communes. Notre grille de départ est constituée par les 660 cartes de l'ALPic 1 et 2². Il s'agit donc d'extraire les notions traitées – ce qui ne correspond pas forcément au titre des cartes – puis de rechercher dans les deux autres ressources les matériaux désignant la même notion.

L'objectif de cette première étape est de permettre, par la réunion de matériaux éparpillés, « une appréhension globale du fait picard qui ne soit plus cloisonnée par les frontières administrative et méthodologique entre la France et la Belgique » (Baiwir/Renders 2019).

2.2 *Une ressource numérique entre atlas et dictionnaire*

La transition vers le numérique nécessite d'intégrer à des ressources géolinguistiques une dimension lexicographique : l'unité d'encodage, la forme linguistique, doit être munie d'une série d'informations selon un programme microstructural fixe. Au sein de celui-ci, le géoréférencement constitue évidemment une donnée essentielle, mais non unique. En effet, des informations grammaticales et sémantiques, éventuellement un découpage morphologique sont prévus dès la première phase d'encodage.

Cette étape nécessite de lever un certain nombre d'obstacles, comme les divergences de systèmes de notation. Selon nous, par souci philologique, il convient de garder accessible la forme originale, mais celle-ci doit évidemment être accompagnée d'une transcription en alphabet phonétique international (API), idéalement automatisée. En outre, une analyse étymologique est souhaitable à moyen terme, car c'est par ce moyen qu'est envisagée la troisième phase du projet.

Dès cette étape, il a été prévu de rendre les ressources accessibles via un site de consultation des cartes originales et des données converties, sous la forme d'une cartographie modulable.

2 — Certaines notions traitées in ALF et ALW seront donc absentes dans un premier temps. Cette décision s'appuie sur la volonté de numériser en priorité l'ensemble de l'ALPic, jusqu'alors absent de la Toile.

2.3 La possibilité d'un réseau roman

Outre des liens sémantiques entre les unités (et donc une interrogation possible par notion), des liens formels permettraient des recherches sémasiologiques ; comment s'organisent sémantiquement tous les lexèmes de telle famille lexicale ? Dans quelles « cartes » trouve-t-on des formes issues de tel étymon ?

L'analyse des unités linguistiques proposée par l'ALW « permet d'inscrire pleinement les matériaux dialectaux dans la discipline de la lexicographie historique, en leur apportant une dimension diachronique dont ils sont dépourvus dans leur forme atlantographique. Le projet APPI fait l'hypothèse que l'ajout de cette épaisseur historique aux données est, dans le domaine galloroman, la solution la plus efficace et la plus pertinente pour intégrer les atlas dans un réseau de ressources lexicographiques numérisées autour du FEW » (Baiwir/Renders 2019).

C'est en effet *via* un « référencement FEW » que les cognats provenant de diverses régions dialectales du monde roman pourraient être connectés. En ce qui concerne les données picardes, elles en sont déjà pourvues dans l'ALW, mais pas dans l'ALF ni dans l'ALPic ; pour ce dernier, le travail a toutefois été entamé dans Carton/Dawson (2010).

Dès l'origine, l'objectif du projet a donc été de munir toutes les formes d'une étiquette étymologique ou, lorsque l'origine est inconnue ou mal assurée, de les intégrer dans le classement onomasiologique du FEW (les volumes des formes d'origine inconnue).

Le modèle, conçu pour pouvoir s'étendre au-delà du parler picard, permettrait donc d'interroger les matériaux malgré leurs disparités formelles de surface.

3. Quel modèle informatique ?

Le projet a commencé en janvier 2018. Les premiers travaux ont permis d'identifier les premières tâches (analyse sémantique des contenus des cartes, comparabilité de celles-ci, etc.) et l'équipe a commencé à élaborer un cahier des charges pour le futur outil informatique. En juillet de la même année, lors du 18^e congrès Euralex à Ljubljana, un rapprochement a eu lieu avec l'équipe du projet Verba Alpina, aux objectifs comparables (voir ce volume, Mutter/Colcuc et https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page_id=4&db=192).

Ce projet, en cours depuis 2014 à la Ludwig-Maximilians-Universität de Munich, a pour ambition la valorisation et l'étude du vocabulaire rural alpin dans toutes les familles de langues de la

zone géographique, par la numérisation des ressources existantes et la récolte de nouveaux matériaux (<https://www.verba-alpina.gwi.uni-muenchen.de/fr/>). Forte d'une expertise dans le traitement des ressources issues d'atlas du domaine alpin (dont l'ALF), l'équipe a, au fil du temps, conçu des outils à destination des collaborateurs du projet (pour encoder les ressources, pour les classer et pour les cartographier) et des utilisateurs de la ressource (avec un module de cartographie « à la carte » et de multiples ressources en ligne).

Très vite, il s'est avéré que l'outil informatique de VerbaAlpina correspondait aux besoins du projet APPI et une convention a été signée en octobre 2018. Les modules d'encodage des formes et de cartographie ont été adaptés et mis à la disposition du projet APPI, grâce à la collaboration des deux équipes dans l'esprit de science ouverte et collaborative que promeut l'équipe de VerbaAlpina (voir ce volume, Colcuc/Mutter).

4. Premiers résultats

Le site du projet est accessible à l'adresse <https://anr-appi.univ-lille.fr/>. L'ALPic, le seul de nos trois atlas exclusivement consacré au picard, a été entièrement numérisé ; les 660 cartes linguistiques sont accessibles sous l'onglet <ALPic>. Un champ de recherche permet de naviguer parmi les intitulés des notices. Les différentes ressources publiées dans les volumes physiques (introduction, tables des matières, informations sur les enquêtes et la transcription, etc.) sont consultables sous l'onglet <outils>.

Les outils intégrant l'ALPic, l'ALW et l'ALF sont au nombre de deux. Le premier est accessible sous l'onglet <macrostructure>, qui aligne les matériaux des trois atlas après une analyse sémantique et morphologique. Pour élaborer cet outil, nous avons défini le contenu sémantique précis des matériaux des cartes de l'ALPic et identifié les notices et les cartes des deux autres ressources traitant la même notion. Il s'agit parfois d'isoler une part des matériaux d'une notice. La notion est alors décrite dans une définition qui s'appuie, lorsque c'est possible, sur celle du TLFi (*Trésor de la langue française informatisé*), avec les précisions ou restrictions qui s'imposent. Lorsque l'équivalence sémantique est assurée, il convient de vérifier la comparabilité des matériaux sur un plan morpho-syntaxique ; s'il s'agit de formes verbales, sont-elles lemmatisées ? Si ce n'est pas le cas, sont-elles au moins en partie comparables ? Les divergences de traitement des données et les points d'attention sont détaillés dans un champ de commentaires. Outre la définition et le commentaire, les notices et les cartes des trois atlas

sont accessibles en mode image, grâce à des accords de coopération avec les équipes partenaires (Service de Dialectologie de l'ULiège pour l'ALW, CartoDialect pour l'ALF). Les cartes de l'ALF sont recadrées autour du domaine qui nous concerne. Actuellement (mai 2020), l'outil comporte un peu plus de 200 notions.

CONCEPT	DÉFINITION	ALPIC	ALW	ALF	COMMENTAIRES
"cru"	Afficher	321 (<i>il fait</i>) <i>humide et froid</i>	3_33 <i>humide</i>	/	Afficher
à bouillons	Afficher	328 (<i>il pleut</i>) <i>à bouillons</i>	3_57_ADD <i>il tombe des gouttes</i>	/	Afficher
à verse	Afficher	327 (<i>il pleut</i>) <i>à verse</i> 327bis (<i>il pleut</i>) <i>à verse</i>	3_58 <i>il pleut à verse</i>	/	Afficher
abeille	Afficher	230 (<i>l'</i>) <i>abeille</i>	/	1 <i>abeille</i>	Afficher
aboyer	Afficher	189 <i>aboyer</i>	/	2 <i>aboyer</i>	Afficher
agneau	Afficher	183 (<i>l'</i>) <i>agneau</i>	/	11 <i>agneau agneaux, agnelle</i>	Afficher
agneler	Afficher	184 <i>agneler</i>	/	/	Afficher
aiguiser	Afficher	308 <i>aiguiser</i>	/	16 <i>aiguiser</i>	Afficher
almanach	Afficher	355 (<i>l'</i>) <i>almanach</i>	3 186 <i>almanach</i>	1434 <i>almanach</i>	Afficher

Illustration 1 : l'outil <macrostructure>

Le second outil est la consultation d'une sélection des données sous l'onglet <plan interactif>. Une vingtaine de concepts sont en ligne, classés thématiquement. Les rubriques « vie et relations sociales », « corps humain et maladies », « ferme, culture et élevage », « terre, plantes et animaux », « maison et ménage », « phénomènes atmosphériques et division du temps » et « morphologie/grammaire » sont pour l'instant représentées (la répartition s'inspire de l'ALW, la seule de nos ressources à organiser ses notices sur un critère sémantique).

Comme il apparaît dans notre exemple (illustration 2 : carte 'maintenant'), ce second outil permet d'étudier, entre autres, les répartitions entre les réalisations en *f*- (en vert) et en *s*- (en bleu) du démonstratif dans une tournure figée de type <à cette heure>. On remarque également qu'au nord du territoire, la zone transitionnelle entre les deux a opté pour un autre type lexical, <maintenant> (en jaune).

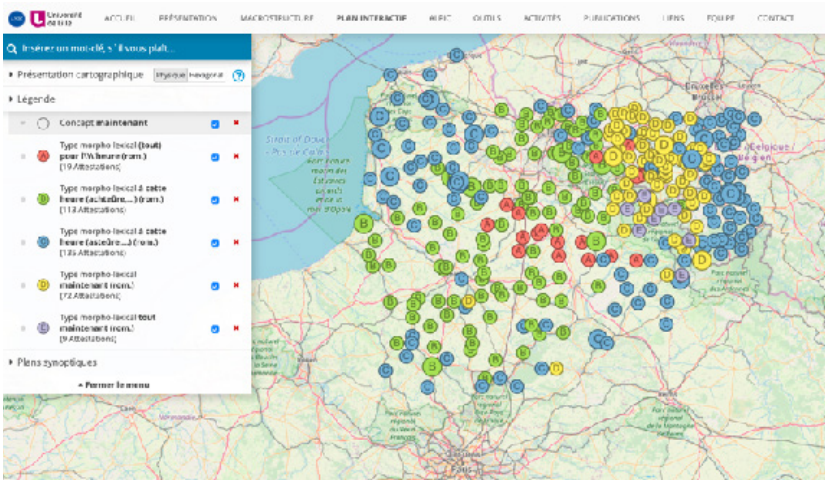


Illustration 2 : ‘maintenant’ dans APPI

Les matériaux sont le fruit de l’encodage manuel des données identifiées à l’étape précédente. L’encodage s’effectue au moyen du système intitulé Betacode, conçu au sein du projet Verba Alpina (https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page_id=13&db=192&letter=B#7). L’objectif du Betacode est de permettre un encodage exhaustif des symboles et des diacritiques des sources sur un clavier standard, sans programme ni application à télécharger. Par la suite, grâce à des tableaux d’équivalences, un basculement vers l’API permettra une meilleure lisibilité des données.

Au niveau de la numérisation, l’unité de traitement est la notion : tous les signifiants rattachés à un concept lors de l’étape de la macrostructure sont encodés et rassemblés dans le système. Toutefois, au niveau de la consultation, d’autres portes d’entrée sont possibles, comme la consultation par « type de base », soit les racines lexicales communes, généralement représentées par leur étymon tel qu’il apparaît dans le FEW. Cet étiquetage permettra ultérieurement l’intégration de liens vers cette ressource, comme le prévoit le projet (voir *supra*). L’interrogation peut également porter sur les « types morpho-lexicaux », afin d’identifier les formants communs à des formes phonétiques diverses, selon un étiquetage par types. Dans l’avenir, un étiquetage plus précis devrait permettre de convoquer certains morphèmes (les suffixes, par exemple) et de les comparer. Enfin, le menu permet également d’isoler les matériaux d’un seul atlas (sous l’onglet « informateur »).

Les données correspondant à plusieurs notions peuvent être affichées ensemble sur une carte, afin de comparer des aires

d'extension. Ainsi, on peut comparer l'aire de phénomènes tels que le maintien de l'occlusion du K- latin à l'initiale suivi de A dans les cartes 'cheval' et 'chemin' (illustration 3).

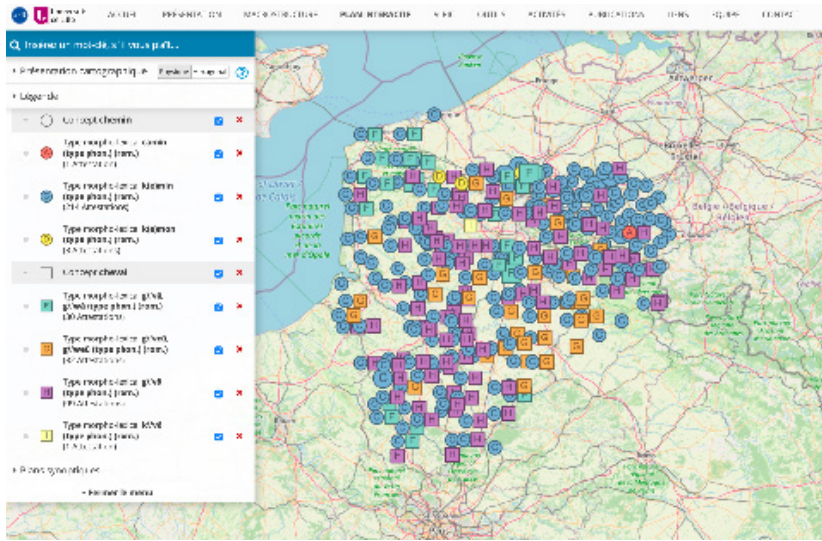


Illustration 3 : les formes en *k*-des types <chemin> 'chemin' et <cheval>, 'cheval' dans APPI

Chaque symbole est cliquable, afin d'accéder au détail des matériaux. Ce détail (v. Illustration 4) se présente sous la forme d'une bulle mentionnant l'unité linguistique, une typisation morpho-lexicale, éventuellement un étymon et la source dont elle est issue. Le travail de transformation automatique en API étant encore en cours, l'affichage actuel respecte les graphies originelles des trois atlas.

Type phonétique	(sans réduction à types)	VA
Type morpho-lexical	gaugae (rom.)	VA
Type de base	GALLICA	VA
Source	Concept	
ALPic 2008* 64 (Prisches)	noix	

Illustration 4 : détail d'une forme désignant la 'noix'

5. Conclusions et pistes

Dans les prochains mois (juin 2020-décembre 2021), l'objectif est d'enrichir la base par l'encodage des données correspondant à 180 notions. Cela permettrait de fournir un tiers des données de l'ALPic interrogeables et enrichies des matériaux ALF et ALW concernant les mêmes notions.

Quelques développements sont en chantier, comme l'interrogeabilité par type morphologique ou la transposition automatique en API, qui soulève des difficultés peu soupçonnables. L'enrichissement des matériaux par adjonction d'informations étymologiques et de liens vers la version informatisée du FEW est également en cours (voir Robecchi dans ce volume). Toutefois, connecter l'APPI au FEW nécessite de définir concrètement les modalités d'implémentation des liens. Une première question concerne les unités à mettre en relation dans chacune des ressources : s'agit-il du lexème, de l'article, de la notice, de la carte ? En fonction de la granularité des unités ainsi identifiées, diverses modélisations informatiques sont possibles à l'avenir – toutefois, dans l'état actuel de la numérisation du FEW, c'est simplement vers

l'entrée que les liens pointeront. Une autre possibilité d'extension serait l'intégration des domaines linguistiques connexes (voir par exemple Brasseur/Baiwir dans ce volume).

Ces enrichissements permettront aux données dialectales picardes d'intégrer pleinement les réseaux lexicographiques en chantier. Grâce au pivot que constitue le FEW, des liens hypertextes seront envisageables avec d'autres ressources, telles que le TLFi ou le DMF. Les corrections et compléments que fournissent les données dialectales pourront, dans un mouvement dialectique, être intégrés dans les projets partenaires. La communauté des chercheurs bénéficiera ainsi de matériaux traités, analysés et exploitables pour documenter les questions de linguistique historique du monde roman. En particulier, des projets pan-romans tels que le DÉRom ou l'ALiR bénéficieront grandement de l'accès à ces données dialectales.

Au niveau picard, une connexion est également possible avec d'autres corpus numériques, tels que Picartext. La ressource vise également à être utilisée par les étudiants et le grand public, en rendant disponibles des ouvrages parfois difficiles d'accès.

En jetant des ponts par-delà les frontières des parlars, des époques ou des États, par-delà les divergences entre les atlas, mais aussi entre les atlas et les autres ressources, nous espérons donner une visibilité plus grande à des matériaux d'une grande richesse, souvent sous-exploités ou ignorés, et participer ainsi à inscrire de plein droit les données dialectales picardes dans le panorama des ressources linguistiques numériques.

Esther BAIWIR et Cécile KAISIN
Université de Lille³

Bibliographie

- ALF = J. GILLIÉRON & E. EDMOND, *Atlas linguistique de la France*, 1 920 cartes, Paris, Champion, 1902-1910.
- ALiR = M. CONTINI (dir.), *Atlas Linguistique Roman*, Vol. I, Tome 1 (Présentation), 232 pages ; Tome 2 (Commentaires), 151 pages ; Tome 3 (Atlas), 14 cartes, Rome, Istituto Poligrafico e Zecca dello Stato, 1996.
- ALPic = F. CARTON & M. LEBÈGUE, *Atlas linguistique et ethnographique picard*, 2 vol. Editions du CNRS, 1989-1998.

3 — Univ. Lille, ULR 1061 - ALITHILA - Analyses Littéraires et Histoire de la Langue, F-59000 Lille, France.

- ALW = L. REMACLE, E. LEGROS, J. LECHANTEUR, M.-T. COUNET, M.-G. BOUTIER, E. BAIWIR, *Atlas linguistique de la Wallonie*, Liège, Université de Liège (10 vol.), 1953.
- E. BAIWIR, « Un type picard par-delà les frontières : le <nom-jeté> », *Les dialectes de Wallonie*, t. 36, 2016, p. 5-24.
- E. BAIWIR & P. RENDERS, « Vingt minutes en Utopie : l'ALWi », in E. Baiwir & P. Renders (éds), *Actes de la 5e Journée liégeoise de Traitement des Sources Galloromanes (Liège, 16 mai 2013)*, Liège, Université de Liège, publication électronique (http://lingwa.philo.ulg.ac.be/trasogal/baiwir_renders_2013.pdf), 9 p, 2013.
- , « Numérisation des données dialectales d'oïl : le projet APPI comme laboratoire », publication en ligne dans le cadre du projet APPI (Atlas pan-picard informatisé, sous la direction d'E. Baiwir), 9 p. (<https://appi.univ-lille.fr/data/medias/baiwirrenders2019>), 2019.
- M.-G. BOUTIER, « Cinq relations de base pour traiter la matière géolinguistique : Réflexions à partir de l'expérience de l'Atlas linguistique de la Wallonie », *Estudis Romànics* 30, p. 301-310, 2008.
- P. BRASSEUR & E. BAIWIR, « Vue d'ensemble de l'ALN: aperçu et perspectives », *Bien dire et bien apprendre* 35, p. 61-70, 2020.
- F. CARTON, « La genèse singulière de l'Atlas linguistique picard », publication en ligne dans le cadre du projet APPI (Atlas pan-picard informatisé, sous la direction d'E. Baiwir), mars 2019, 3 p. (<https://appi.univ-lille.fr/data/medias/carton2019>), 2019.
- F. CARTON & A. DAWSON, *Index lemmatisé et étymologique de l'Atlas linguistique et ethnographique picard, Volumes I et II*, Amiens, Université de Picardie-Jules Verne, Collection du CEP, 2010.
- B. COLCUC & CH. MUTTER, « Les deux plans d'interopérabilité : données primaires (matériel linguistique) et structuration des données à l'exemple du projet Verba Alpina », *Bien dire et bien apprendre* 35, p. 131-146, 2020.
- DÉRom = *Dictionnaire étymologique roman*, <http://www.atilf.fr/DERom>
- DMF = R. MARTIN et S. BAZIN (dir.), *Dictionnaire du Moyen Français*, Nancy, ATILF/CNRS & Université de Lorraine, <http://www.atilf.fr/dmf>, 2015.
- FEW = W. von WARTBURG, *et al.*, *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes*, 25 vol., Bonn/Heidelberg/Leipzig-Berlin/Bâle, Klopp/Winter/Teubner/Zbinden, 1922-2002.
- N. MAZZIOTTA, « Exploitation de l'Ontology Web Language pour la rédaction des notices de l'Atlas Linguistique de la Wallonie », in

- Heiden, S. et Pincemin, B. 2008. *9^{es} journées internationales d'analyses statistiques des données textuelles*. Lyon, 12 au 14 mars 2008, Lyon, Presses universitaires de Lyon, p. 823-835, 2008.
- , « L'informatisation du *Französisches Etymologisches Wörterbuch*. Concepts pour une approche modélisée commune à l'*Atlas Linguistique de la Wallonie* », ZRP 127, p. 36-62, 2011.
- Picartext = Base de données *Picartext*, Centre d'Études des Relations et Contacts Linguistiques et Littéraires (CERCLL – EA 4283), Université de Picardie Jules Verne (<https://www.u-picardie.fr/LESCLaP/PICARTEXT/Public/index.php>).
- P. RENDERS, *L'informatisation du Französisches Etymologisches Wörterbuch. Modélisation d'un discours étymologique*, Strasbourg, ELiPhi-Editions de linguistique et de philologie, 2015.
- P. RENDERS, E. BAIWIR et G. DETHIER, « Automatically Linking Dictionaries of Gallo-Romance Languages Using Etymological Information », in Kosem, I., Jakubiček, M., Kallas, J. (Eds.) *et al.*, *Electronic0020lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*, p. 452-460, 2015.
- M. ROBECCHI, « La place des Atlas dans la rétro-conversion du FEW », *Bien dire et bien apprendre* 35, p. 147-168, 2020.
- J. SÉGUY, « Les Atlas linguistiques de la France par régions », *Langue française* 18 (Les parlers régionaux), p. 65-90, 1973.
- TLF, TLFi = P. IMBS & B. QUEMADA (dir.), *Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècles (1789–1960)*, version informatisée : <http://atilf.atilf.fr/tlf.htm>, 1971–1994.

