



APPI: L'ALPic en ligne et en reseau

Esther Baiwir, Cécile Kaisin

► To cite this version:

Esther Baiwir, Cécile Kaisin. APPI: L'ALPic en ligne et en reseau. Les dialectes de Wallonie, 2021, 37, pp.43-54. hal-04405381

HAL Id: hal-04405381

<https://hal.science/hal-04405381>

Submitted on 19 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

APPI : L'ALPic en ligne et en réseau

0. Introduction

Le merveilleux outil qu'est l'ALPic (*Atlas linguistique et ethnographique picard*), présenté dans ce volume par l'une de ses principales chevilles ouvrières, Fernand Carton (2021), méritait d'être numérisé et accessible en version virtuelle. Le projet ANR APPI (*Atlas pan-picard informatisé*), en cours à l'Université de Lille depuis 2018, s'est attelé à cette numérisation, ainsi qu'à la valorisation des matériaux dans un cadre plus large, incluant toutes les ressources atlantographiques traitant du picard.

Dans cette contribution, nous présenterons brièvement ces ressources (1), puis le projet APPI (2).

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 campagne d'enquête 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 1400 questions. Le territoire qui nous intéresse y est représenté par une quarantaine de points, dont 8 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 et 2021), 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 1150 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 1670 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).

2. Le projet APPI

Le projet ANR APPI (<https://anr.fr/Projet-ANR-17-CE38-0002>), en cours à l'Université de Lille depuis 2018, a pour ambition de numériser conjointement les données picardes des trois atlas évoqués *supra*, dans le cadre d'une analyse sémantique des matériaux, mais également d'envisager des liens avec d'autres projets et ressources. Grâce à la collaboration avec le projet VerbaAlpina de l'Université de Munich, une plateforme informatique a été conçue, accueillant la numérisation de toutes les cartes de l'ALPic ainsi que les matériaux rassemblés des trois atlas pour une série de notions.

La plateforme est consultable à l'adresse <https://anr-appi.univ-lille.fr/>. L'ALPic y est entièrement consultable ; les 660 cartes linguistiques sont accessibles sous l'onglet <ALPic>. 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>.

Quant aux outils intégrant l'ALPic, l'ALW et l'ALF, ils sont au nombre de deux. Le premier consiste en l'alignement des maté-

riaux des trois atlas, à l'issue d'une analyse sémantique et morphologique. Notre nomenclature de départ est constituée par les 660 cartes de l'ALPic 1 et 2¹. À partir de cette liste, il s'agit d'extraire les notions traitées – ce qui ne correspond pas forcément au titre des cartes (voir Baiwir à paraître) – puis de rechercher dans les deux autres ressources les matériaux désignant la même notion, afin d'assurer leur comparabilité. 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.

L'objectif de cette étape d'analyse 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). Ce premier outil est accessible sous l'onglet <macrostructure>.

¹ 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.

Macrostructure

Alignement d'une part des matériaux de l'ALPic, l'ALW et l'ALF sur une base sémantique (environ 200 notions).

Les concepts sont munis d'une définition qui s'appuie sur celles du TLFi (*Trésor de la langue française informatisé*). Les notices et cartes des trois atlas sont accessibles en mode image, grâce à l'accord de nos équipes partenaires (dont celle du projet CartoDialect pour l'ALF, dont nous ne reproduisons que la zone la plus septentrionale).

Une zone de commentaires détaille, le cas échéant, les divergences de traitement des données selon les atlas, qu'elles soient sémantiques ou morphologiques, ainsi que divers points d'attention.

Requête (concept seulement) :

CONCEPT	DÉFINITION	ALPIC	ALW	ALF	COMMENTAIRES
à bouillons	Afficher	328 (il pleut) à bouillons	3_57_ADD il tombe des gouttes	/	Afficher
à verse	Afficher	327 (il pleut) à verse 327bis (il pleut) à verse	3_58 il pleut à verse	/	Afficher
abeille	Afficher	230 (l')abeille	/	1 abeille	Afficher
aboyer	Afficher	189 aboyer	/	2 aboyer	Afficher
agneau	Afficher	183 (l')agneau	/	11 agneau agneaux, agnelle	Afficher
agneuler	Afficher	184 agneuler	/	/	Afficher
aiguiser	Afficher	308 aiguiser	/	16 aiguiser	Afficher
almanach	Afficher	355 (l') almanach	3_186 almanach	1434 almanach	Afficher
altéré (d'un œuf)	Afficher	213 (l')œuf couvé	/	/	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 la base d'un critère sémantique).

Comme il apparaît dans notre illustration 2 (carte 'maintenant'), ce second outil permet d'étudier, par exemple, 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 ces deux réalisations a opté pour un autre type lexical, <maintenant>.

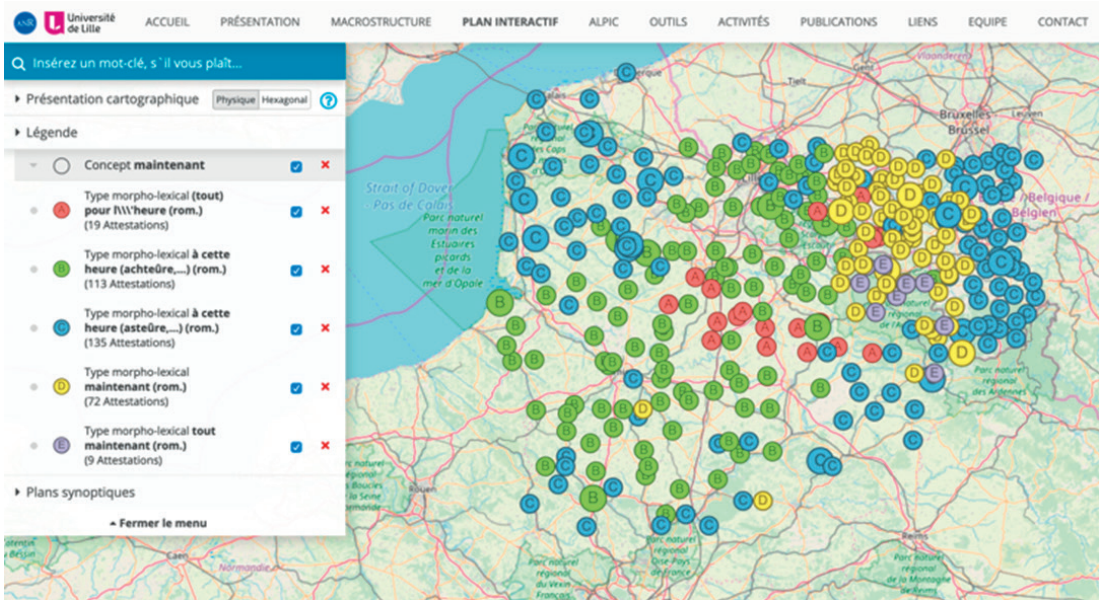


ILLUSTRATION 2 : 'MAINTENANT' DANS APPI

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 VerbaAlpina (https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page_id=13&db=192&letter=B#7), et partagé dans l’esprit de science ouverte et collaborative que promeut l’équipe de VerbaAlpina.

Au niveau de la numérisation, l’unité de traitement est la notion :

tous les signifiants rattachés à une notion 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 », c'est-à-dire par 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. 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 du K- latin à l'initiale suivi de A dans les cartes 'cheval' et 'chemin' (illustration 3).

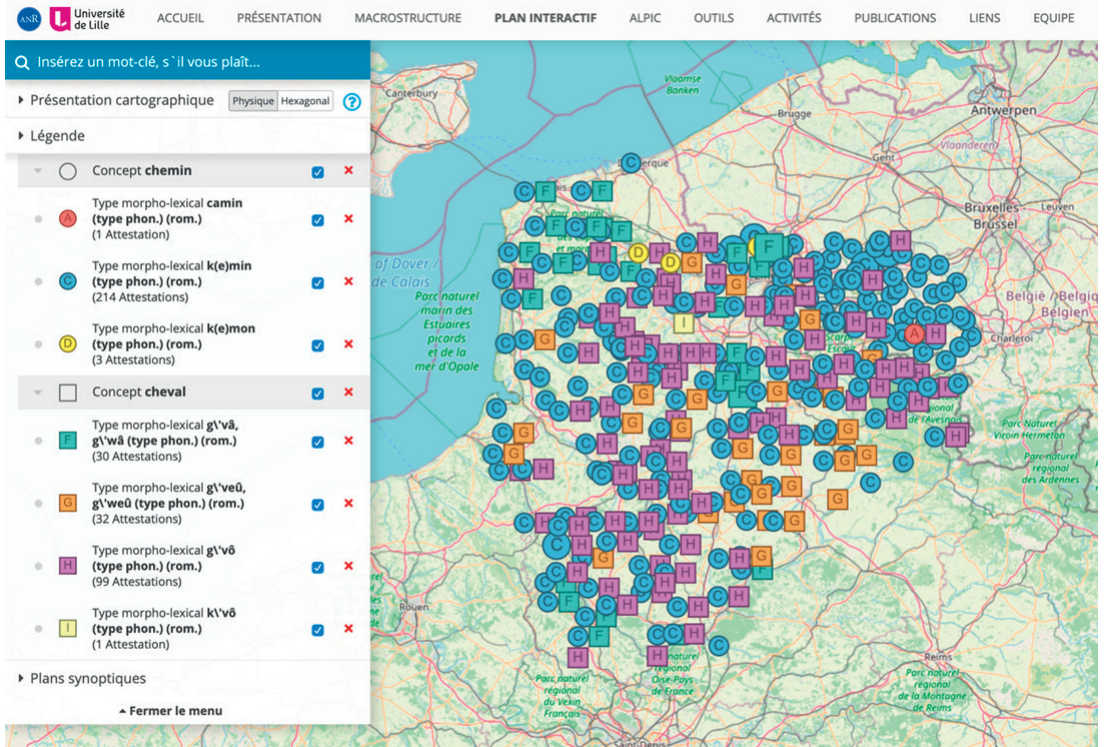


ILLUSTRATION 3 : LES FORMES EN K- DES TYPES <CHEMIN> 'CHEMIN' ET <CHEVAL> 'CHEVAL' DANS APPI

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.



ILLUSTRATION 4 : DÉTAIL D'UNE FORME DÉSIGNANT LA 'NOIX'

3. Conclusions et pistes

L'encodage des données, toujours en cours, devrait à terme intégrer au moins un tiers des données de l'ALPic, interrogeables en ligne 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. L'enrichissement des matériaux par adjonction d'informations étymologiques et de liens vers la version informatisée du FEW est également en cours (v. par exemple Robecchi 2020).

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 parlers, 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 Lille²

² Univ. Lille, ULR 1061 - ALITHILA - Analyses Littéraires et Histoire de la Langue, F-59000 Lille, France. Cet article est une version condensée de l’article paru in BDBA (Baiwir & Kaisin 2020).

Bibliographie

ALF = GILLIÉRON (Jules) & EDMOND (Edmont), *Atlas linguistique de la France*, 1920 cartes, Paris, Champion, 1902-1910.

ALiR = CONTINI (Michel) 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 = CARTON (Fernand) & LEBÈGUE (Maurice), *Atlas linguistique et ethnographique picard*, 2 vol., Éditions du CNRS, 1989-1998.

ALW = REMACLE (Louis), LEGROS (Élisée), LECHANTEUR (Jean), COUNET (Marie-Thérèse), BOUTIER (Marie-Guy), BAIWIR (Esther), *Atlas linguistique de la Wallonie*, Liège, Université de Liège, 10 vol., 1953-.

BAIWIR (Esther), « Un type picard par-delà les frontières : le <nom-jeté> », *Les dialectes de Wallonie*, t. 36 (2016), p. 5-24.

—, « Quel sens ont les unités lexicales des atlas linguistiques? Une exploration sémantique dans le domaine (gallo)roman », Actes du colloque intitulé *Nouveaux regards sur la variation dialectale*, Paris, novembre 2019, Société de Linguistique romane, à paraître.

— & KAISIN (Cécile), « L'ALPic en ligne et le projet d'un Atlas pan-picard informatisé (APPI) », *Bien dire et bien apprendre* 35 (2020), p. 47-59.

— & RENDERS (Pascale), « Numérisation des données dialectales d'oïl : le projet APPI comme laboratoire », publication en ligne dans le cadre du projet APPI (Atlas pan-picard informatisé), 2019, 9 p. (<https://appi.univ-lille.fr/data/medias/baiwirrenders2019>).

CARTON, (Fernand), « La genèse singulière de l'Atlas linguistique picard », publication en ligne dans le cadre du projet APPI (Atlas

pan-picard informatisé), mars 2019, 3 p. (<https://appi.univ-lille.fr/data/medias/carton2019>).

—, « Soixante ans après... Message d'un enquêteur », *Bien dire et bien apprendre* 35 (2020), p. 5-7.

—, « La genèse singulière de l'*Atlas linguistique picard* », *Dialectes de Wallonie* 37 (2021), p. 38-42.

DÉRom = *Dictionnaire étymologique roman* <http://www.atilf.fr/DERom>

DMF = MARTIN (Robert) et BAZIN (Sylvie) dir., *Dictionnaire du Moyen Français*, Nancy, ATILF/CNRS & Université de Lorraine, 2015, <http://www.atilf.fr/dmf>.

FEW = WARTBURG (Walther von) *et al.*, *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes*, 25 vol., Bonn/Heidelberg/Leipzig-Berlin/Bâle, Klopp/Winter/Teubner/Zbinden, 1922–2002.

Picartext = Base de données *Picartext*, CERCLL, Université de Picardie Jules Verne <https://www.u-picardie.fr/LESCLaP/PICARTEXT/Public/index.php>.

ROBECCHI (Marco), « La place des Atlas dans la rétro-conversion du FEW », *Bien dire et bien apprendre* 35 (2020), p. 147-167.

SÉGUY (Jean), « Les Atlas linguistiques de la France par régions », *Langue française* 18 (1973), p. 65-90.

TLF, TLFi = IMBS (Paul), QUEMADA (Bernard) dir., *Trésor de la langue française. Dictionnaire de la langue du XIXe et du XXe siècles (1789–1960)*, 1971–1994. Version informatisée <http://atilf.atilf.fr/tlf.htm>

VerbaAlpina = KREFELD (Thomas) & LÜCKE (Stephan) éd., *VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit*, München, 2014–, online, <https://dx.doi.org/10.5282/verbaalpina>