



HAL
open science

Le picard dans le projet SyMiLa

Anne Dagnac

► **To cite this version:**

Anne Dagnac. Le picard dans le projet SyMiLa. Les dialectes de Wallonie, 2021, 37, pp.55-77.
hal-03696051

HAL Id: hal-03696051

<https://hal.science/hal-03696051>

Submitted on 15 Jun 2022

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.

Le picard dans le projet SyMiLa

Anne Dagnac (CLLE – CNRS et Université Toulouse 2)

Version auteur préfinale, mai 2021

Paru dans : Baiwir, E. (éd), 2021, *Hommage à Fernand Carton : mélanges*. Liège : Société de langue et de littérature wallonnes, Les Dialectes de Wallonie, n°37. ISSN 0773-7688, 37, pp. 55-77.

Dans cet article, nous présentons le projet SyMiLa (Syntactic Microvariation of the Romance Languages of France / Microvariation syntaxique dans les langues romanes de France :), et la place qu'y tient le picard. Ce projet, qui a reçu un financement de l'ANR (Agence Nationale de la recherche) de février 2013 à janvier 2017 (ANR-12-CORP-0014), comprend deux axes, ou « sous-projets », SyMiLa ALF et SyMiLa-NES, dont les données ont vocation à être réunies dans des bases de données soeurs, accessibles sur un unique site web (<http://symila.univ-tlse2.fr/>) et, à terme, annotées et interrogeables de manière identique.¹

SyMiLa-ALF a pour but de rendre accessible et d'analyser les données syntaxiques de l'*Atlas linguistique de la France* (ci-après ALF), d'Edmond Edmont et Jules Gilliéron : leurs enquêtes conduites de 1897 à 1900 sont la source de l'atlas papier à base lexicale que nous connaissons aujourd'hui, publié entre 1902 et 1910. Il compte 639 points d'enquête principalement répartis sur les régions gallo-romanes de France (hors Corse), avec quelques points dans les zones francophones de Belgique, de Suisse, d'Italie et des îles Anglo-Normandes. L'enquêteur, Edmond Edmont, étant natif de Saint-Pol-sur Ternoise (il est lui-même le témoin du point 284A, celui du point 284B étant sa mère), les enquêtes sont particulièrement denses dans le domaine linguistique picard (une cinquantaine de points) ; à l'usage, elles s'avèrent aussi particulièrement fiables. SyMiLa-NES (Nouvelles Enquêtes Syntaxiques) a quant à lui pour objectif d'effectuer de nouvelles enquêtes, cette fois à cible spécifiquement syntaxique : à ce jour, 18 enquêtes ont été réalisées en domaine picard.

Dans un premier temps, nous détaillons le triple objectif initial du projet et les méthodes prévues pour y parvenir. Nous faisons ensuite un bilan, centré sur le cas du picard, de l'état d'avancement de chaque sous-projet à l'issue du financement, sur les problèmes pratiques et techniques rencontrés, sur les solutions retenues et les modalités envisageables pour sa poursuite.

1. Le sous-projet SyMiLa ALF

Pour mémoire, l'ALF est issu des enquêtes de terrain conduites par Edmond Edmont à la toute fin du XIX^e siècle sous la houlette de Jules Gilliéron, selon des modalités bien connues des dialectologues, et dont les limites ont été abondamment commentées. La méthodologie repose en effet sur la traduction par des patoisants d'un questionnaire en français, aujourd'hui perdu ; ce dernier, qui a pu être reconstitué dans le cadre du projet à partir des carnets d'enquêtes, était proposé oralement à la traduction par l'enquêteur, qui notait essentiellement la première réponse, plus « spontanée » (le « premier jet »)... mais aussi plus exposée aux effets de calque. En l'absence, bien souvent, de possibilités de recoupement direct, les données obtenues se révèlent souvent imparfaites mais néanmoins précieuses. Nos recoupements avec des corpus nous font partager, pour le picard du moins, les conclusions de Séguy (1973) sur les données de l'*Atlas Linguistique et ethnographique de la Gascogne* (ALG) : la présence, dans les traductions, d'une structure divergente du français standard, en particulier quand elle est systématique dans une zone géographique, est un excellent indicateur de son appartenance robuste à la grammaire dudit dialecte. En revanche, l'absence d'une construction divergente dans la traduction ne permet pas de trancher entre absence dans le dialecte ou effet de calque.

1.1. La base SyMiLa-ALF : principes

Les carnets d'enquête, établis pour chaque point d'enquête par Edmond Edmont, sont archivés par cartons à la BNF. Dans le cadre de SyMiLa, nous en avons fait numériser une partie, qui sont désormais accessibles en pdf sur le site des Archives et Manuscrits de la BNF (<https://archivesetmanuscrits.bnf.fr/ark:/12148/cc116301>). Ils ont la forme illustrée en Figure 1 :

bwiq (vieux), bwi	foel	ē nōyō d'pēg
būryēn	plāt	100 gl ānā-si n'ō œ
rōq	100 " (Incunna)	ghāmē d'frū
rōzyē	kōpē du lō pān	ōz ērēt du vīn
būcō	fwār d'fācēn	kōm sēz āp i
lique jēnwāf, m. (A)	jē m' sū āsi dzū	n'n'ēt wāt kērkē
pāmyē	100 ēn āp, āpōyē kōt	100 si ō n' mējō
pwāryē	100 ēcl āp (B). ēn brāk	pwā nū prōn, āl
qōqē	pārit āl ō kēē	mūzjō bētō

Figure 1 : Extrait du carnet d'enquête d'E. Edmont pour le point 276 (Ligny-Saint-Flochel, Pas-de-Calais)

Les données du carnet, retravaillées et réorganisées, ont donné naissance à l'ALF, un atlas papier en 9 tomes, comportant 1920 cartes : 1421 cartes couvrant l'ensemble du territoire considéré, 499 quasi exclusivement² la moitié sud ou le quart sud-est. Les cartes sont ordonnées en suivant l'ordre alphabétique du mot qu'elles illustrent ; par exemple, la carte 1 est *abeille*, les cartes 2 *aboyer*, tandis que les cartes 82 à 103 regroupent diverses formes du verbe *avoir* (82 : *avoir*, 83 : *J'ai / j'en ai / je l'ai*, 84 : *tu as*, 85 : *l'as-tu* ; ... ; 98 : *tu aurais* ; ... ; 103 : *il y a eu*). La carte 1421 est *vriille*. L'ordre alphabétique redémarre pour les demi-cartes puis pour les quarts de carte. Sur une carte, à chaque point géographique est associée la transcription en alphabet Rousselot adapté par Gilliéron de la réponse du témoin, avec parfois quelques variantes, et des commentaires occasionnels en haut de la carte. La figure 2 illustre la carte n° 51, « sous un arbre », classée alphabétiquement à « arbre », pour le domaine picard. La figure 3 représente les données du point 284 (A), dont Edmond Edmont est lui-même le témoin.

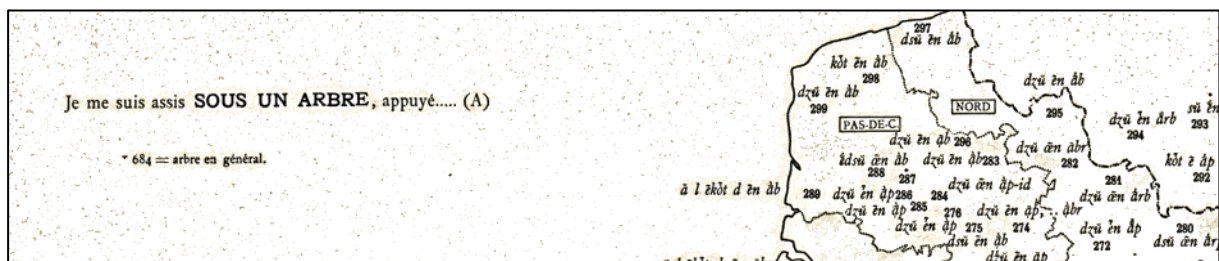


Figure 2 : Extrait de la carte 51 de l'ALF, « sous un arbre »

dzū ān āp
284

Figure 3 : Point 284 : Saint-Pol-sur-Ternoise – ville, dans la carte 51

L'angle supérieur gauche de la carte, qui indique, en gras, son titre, « sous un arbre », porte parfois un contexte d'où est extrait l'élément représenté : ici, on a *Je me suis assis SOUS UN ARBRE, appuyé...* (A). Une confrontation entre ces indications et les carnets d'enquêtes, effectuée par Guylaine Brun-Trigaud, a permis d'établir que certains mots ou groupes de mots de l'atlas ont en effet, à l'origine, été élicités sous forme de phrase puis extraits pour la publication papier alphabétique. L'extrait de carnet présenté en figure 1 montre en effet qu'au-delà des mots isolés de la colonne 1, ou des groupes de la colonne 2 (couper du bon pain, faire des manières), ce sont des phrases entières, voire des séquences, que le témoin a traduites : « je me suis assis sous un arbre, appuyé contre le tronc. Une branche pourrie m'est tombée sur la figure et m'a fait saigner du nez » (bas de la colonne 2), puis, colonne 3 : « cette année-là, il y a eu beaucoup de fruits. Vous auriez dû voir comme les arbres en étaient chargés », puis « si vous ne mangez pas vos prunes, elles moisiront bientôt ».

Il est ainsi possible de travailler sur de la syntaxe dialectale en raboutant les éléments dispersés dans l'atlas. Ainsi, outre la carte 51 de la figure 2, on peut consulter les cartes 500 (Je me suis), 63 (assis), 48 (appuyé) et 1334 (contre le tronc), afin de reconstituer la phrase élicitée à l'origine : « Je me suis assis sous un arbre, appuyé contre le tronc ». Bien sûr, à partir des 9 tomes de l'atlas papier, reconstituer chaque phrase pour chaque point demande une certaine manutention et une patience certaine...

Le projet SyMiLa-ALF a donc recensé l'ensemble des phrases et les cartes correspondantes (cf. http://symila.univ-tlse2.fr/alf/questionnaire_alf_reconstitue) : nous avons dénombré 181 phrases complètes, correspondant à 525 cartes. Nous avons également entrepris de saisir manuellement les données correspondantes, carte par carte sous Excel, en alphabet phonétique international, plus maniable,³ et dans une police Unicode. Les données des fichiers-cartes ont ensuite été fusionnées en gérant les problèmes d'ordre des mots, les variantes et reformulations, et versées dans la base. Ainsi, pour le point 284A (point 0 dans la base), on trouve :

(1) [eʒ m ε asi dzu ðn 'a:p apoje kõt eʃ tr'õ]
 JE M'AI ASSIS DESSOUS UN ARBRE APPUYÉ CONTRE LE TRONC⁴

Sur les 181 phrases recensées, 49 ont à ce jour été saisies dans la base (soit 27% du total, 225 cartes sur 527). Elles ont également été enrichies, notamment par des annotations morpho-syntaxiques (propriétés des phrases, étiquettes morphosyntaxiques dans une version adaptée d'Eagles, lemmatisation), pour faciliter à terme la recherche dans la base. Les formes occitanes sont également accompagnées de « gloses », c'est-à-dire d'une graphie standard, pour faciliter les recherches lexicales, et les formes d'oïl et de francoprovençal d'une proposition de glose établie par Patrick Sauzet, responsable général du projet SyMiLa : elles demandent encore à être systématiquement vérifiées pour ces dialectes.

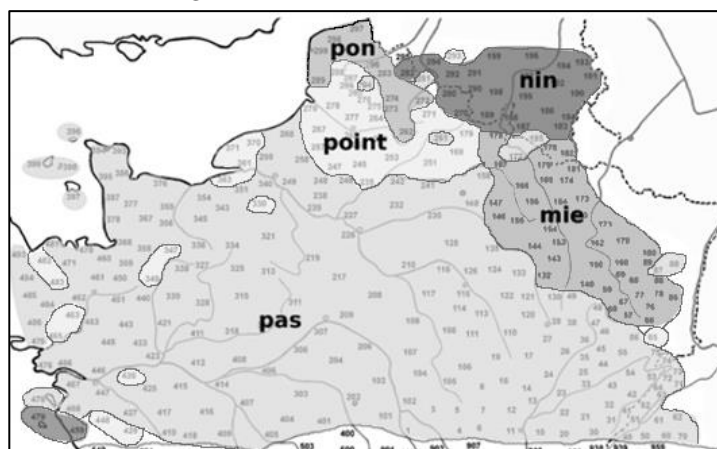
1.2. Quelques exploitations possibles

Le travail effectué permet un accès bien plus lisible aux phénomènes syntaxiques que ne le permettrait l'atlas papier. Bien sûr, il faut manipuler ces données avec prudence, et les recouper avec d'autres sources. Outre les biais possiblement induits par la traduction spontanée, le questionnaire est, du point de vue syntaxique, un filet à larges mailles, ne ciblant en profondeur aucun phénomène particulier. Néanmoins, notamment lorsque des données distinctes du stimulus français (et donc non susceptibles d'être des calques) apparaissent dans des zones cohérentes, elles signalent l'intérêt syntaxique d'une construction. Nous en donnons ici quelques illustrations.

1.2.1 Phrases et cartes isolées

La consultation de cartes ou de phrases isolées peut faire apparaître une répartition géographique homogène de certains phénomènes, voire suggérer des évolutions diachroniques.

Ainsi, la phrase « j'ai cru qu'ils ne viendraient pas » contient une négation verbale, représentée sur la carte 897 de l'ALF (« *J'ai cru QU'ILS NE viendraient PAS* »). Elle fait clairement apparaître la



distribution régionale du marqueur post verbal usuel pour le domaine d'oïl, illustré en figure 4 : 4 formes s'y répartissent le territoire, *MIE* à l'est, *NIN* au nord-est, *POINT* et sa variante septentrionale *PON* dans une zone englobant le domaine picard (et sporadiquement à l'ouest, résiduellement à l'est) et enfin la forme standard *PAS*.

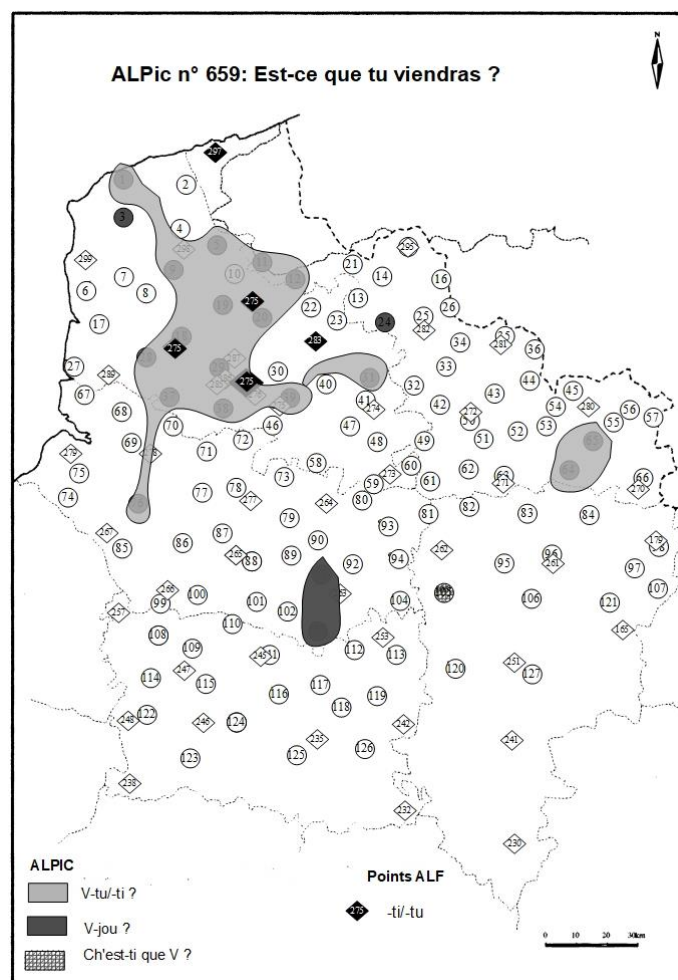


Figure 5 : Particules interrogatives verbales, d'après la carte 659 de l'ALPIC

Ce sont là deux apports majeurs à verser au dossier ouvert par Emrik dans sa démarche diachronique et philologique.

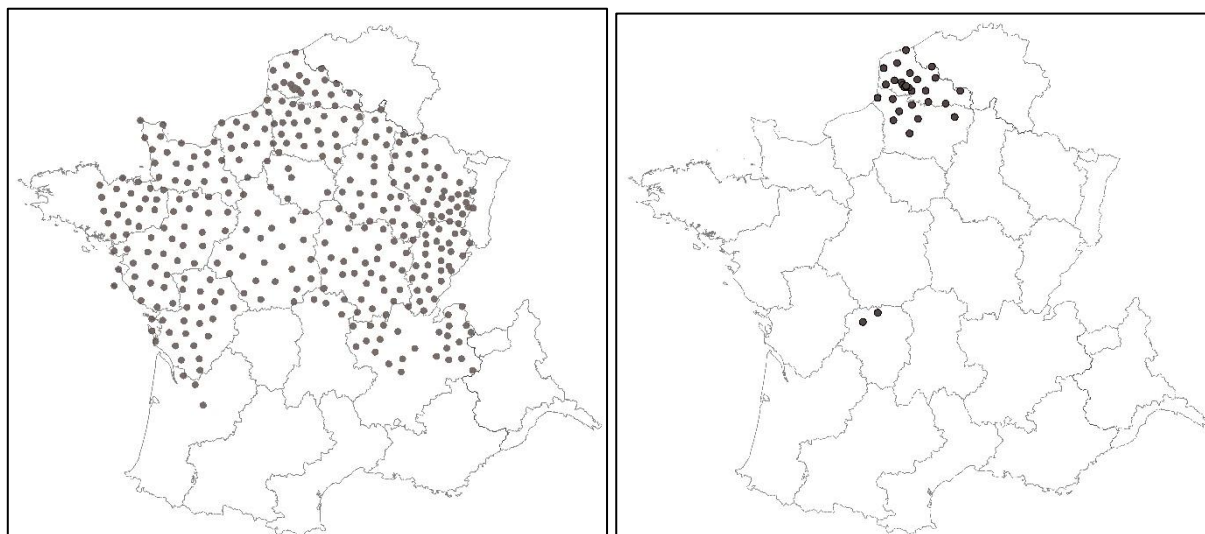
1.2.3. Comparaisons internes

Lorsque, par un heureux hasard, nous disposons de phrases faisant apparaître des phénomènes liés, des comparaisons internes à la base de données et à l'atlas deviennent possibles. Certaines, comme la forme des interrogatives, donnent des résultats difficiles à cartographier lisiblement. D'autres en revanche, semblent éclairer certains phénomènes. C'est le cas par exemple du fameux « redoublement du sujet », tantôt attribué au picard et à certains dialectes du français (Auger 2003), tantôt au français oral (par exemple Blasco 1994), distingué de (Poletto 2000, Legendre et Culbertson 2008) ou assimilé à la dislocation à gauche.

L'ALF, sur ce point, donne des indications intéressantes. On y trouve en effet des phrases-stimulus disloquées, comme « Moi, je ne les aide pas ». La plupart des dialectes d'oïl (et même quelques locuteurs d'oc) reproduisent cette dislocation : le pronom fort est repris par un pronom clitique, cf. figure 6, où les points en gris indiquent la coexistence d'un pronom fort et d'un clitique sujet. Néanmoins, on trouve aussi des phrases-stimulus à sujet nominal non disloqué, comme *Le chasseur va à la chasse pour chasser*, d'interprétation potentiellement générique et thétique, moins propice à la dislocation. Or, dans

ce cas, un pronom clitique apparaît néanmoins de manière très systématique en domaine picard et seulement ponctuellement dans le reste de la gallo-romania (cf. Figure 7).

Est-ce à dire que le redoublement du sujet est systématique en picard, comme on peut l'entendre ? Les données de l'ALF invitent à la prudence sur ce point : une étude de plusieurs phrases



Carte 12 (« Moi, je ne les aide pas »)

de l'ALF à sujet nominal donne, pour le domaine, les résultats illustrés en figure 8.

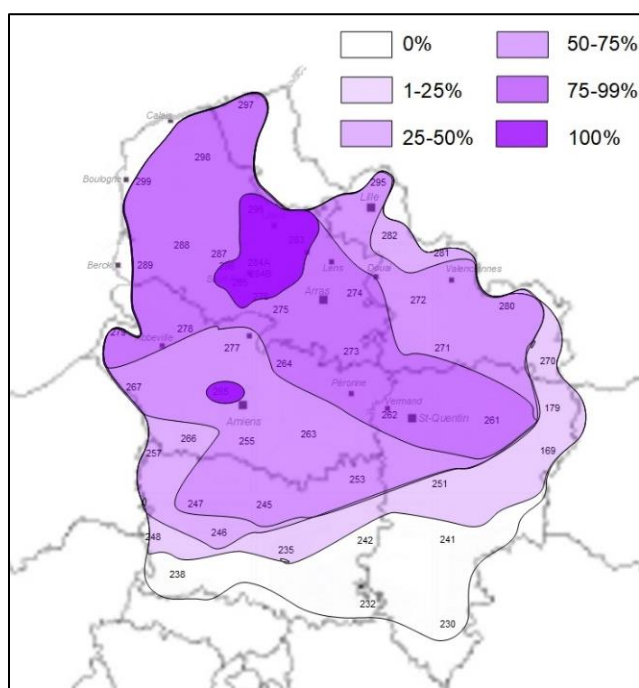


Figure 8 : taux de redoublement du sujet nominal par un clitique à partir de 5 cartes de l'ALF

Le tableau qui en ressort est plutôt celui d'un phénomène à systémativité décroissante à mesure que l'on s'éloigne d'un centre géographique correspondant grosso modo, comme dans l'exemple précédent, au Ternois. Les conditions d'enquête pourraient avoir joué, mais il serait intéressant de confronter ce résultat à des études de corpus géographiquement localisés de diverses périodes afin d'explorer la diachronie du phénomène.

1.3. Bilan et prolongements

Le premier travail effectué sur le volet ALF du projet SyMiLa confirme l'intérêt de disposer de données sérielles accessibles et facilement manipulables. D'un point de vue pratique, il reste à prolonger ce travail. La priorité est de compléter la saisie, puisque 70% des cartes sont encore en attente. Les taux d'erreurs lors des tests de saisie étaient très faibles : il n'en reste pas moins qu'une vérification systématique est nécessaire. Si le projet ECLATS (<http://eclats.imag.fr/>), dont un des objectifs est de numériser les atlas par OCR, aboutit, les deux projets pourraient permettre des vérifications réciproques. Les étiquetages, qui sont générés en partie automatiquement, doivent également être vérifiés, et les documents explicatifs améliorés. Ces problèmes sont faciles à régler : ils ne demandent que du temps et/ou de la main d'œuvre compétente, et un nouveau financement pourrait y pourvoir.

D'autres prolongements sont plus lourds, car ils requièrent la présence d'un informaticien à la fois pour le développement, mais également pour la maintenance de la base et du site : il semble peu prudent de continuer à recourir à un financement sur projet, nécessairement à court terme, pour ce type d'opération. Les postes pérennes de personnels de soutien à la recherche se raréfiant, les modalités permettant ce prolongement restent à débattre, d'autant que le problème se pose pour toutes les initiatives de création de ressources financées sur projet. Pourtant de nombreuses améliorations restent à apporter à la base. C'est le cas pour l'interface de recherche, qui ne permet pas l'affichage simplifié des résultats ou celui des langues/dialectes associés aux points d'enquête, ainsi que du format de rapatriement des résultats. C'est également le cas de la fonction Recherche elle-même. Actuellement, elle est limitée. Elle permet de choisir tous les lieux ou un seul, toutes les phrases ou une seule, ou encore les phrases associées à une propriété syntaxique générale (ex : « interrogative ») ou contenant un type de partie du discours, comme l'illustre la Figure 9.

Figure 9 : Aspect actuel de la fonction Recherche dans SyMiLa-ALF

Outre le fait que la fonction Recherche est présentée de manière peu ergonomique, elle ne permet pas d'exploiter certaines informations pourtant disponibles dans la base, comme la limitation des recherches à une langue ou un dialecte, ou des croisements de propriétés morphosyntaxiques.

Plus délicate encore est sans doute la généralisation des « gloses » par une graphie standard : si l'occitan et quelques langues d'oïl en disposent, c'est loin d'être le cas pour toutes. Leur établissement est dans ce cas un projet de recherche en soi.

2. Le sous-projet SyMiLa-NES

Cette partie du projet vise à créer un nouveau corpus d'enquêtes, ciblé sur cinq phénomènes syntaxiques. Nous présentons ici d'abord les principes de SyMiLa-NES, puis son état d'avancement.

2.1. Principe des nouvelles enquêtes

Afin d'avoir une profondeur diachronique, le choix a été fait de garder, dans ses grandes lignes, la méthodologie d'enquête de l'ALF : il s'agit d'une tâche de traduction à partir qu'un questionnaire unique. Le premier jet n'est néanmoins pas systématiquement favorisé (cela dépend de fait du confort du témoin vis-à-vis de l'exercice) et certaines phrases peuvent donner lieu à des jugements de grammaticalité, bien que tous les enquêteurs n'aient pas été aussi à l'aise avec cet aspect. Le

questionnaire est centré sur les cinq phénomènes suivants : syntaxe de base (ordre des mots), réalisation des pronoms personnels (forme et place, combinaisons), négation et concordance négative, structures interrogatives et accords de genre et de nombre au sein du groupe nominal.

Les phrases sont établies afin de couvrir le champ de ces constructions à la fois empiriquement (par exemple, toutes les personnes devraient être obtenues pour les pronoms personnels) et dans une perspective théorique (présence des contextes réputés cruciaux pour l'analyse d'un phénomène). À partir de ces contraintes syntaxiques (ex : interrogative directe totale avec un sujet nominal), les phrases correspondantes ont été conçues pour s'insérer dans de courts dialogues et une séquence narrative (« Conte de l'âne peureux ») : l'objectif était d'une part de placer le témoin dans une situation un peu plus naturelle qu'avec une traduction de liste de phrases déconnectées, et d'autre part de fournir des contextes aux items syntaxiques. Le questionnaire se termine par une série de phrases isolées, dont nous faisons l'hypothèse qu'au moment de les aborder, le témoin se sera familiarisé avec l'exercice, ainsi que par la conjugaison de trois verbes.

Un exemple du dispositif est donné dans la Figure 10. En gras sont les items ciblés en priorité ; la colonne de droite indique à l'enquêteur le type d'information que l'on souhaite obtenir. Les variantes précédées d'une étoile sont agrammaticales en français: l'objectif de l'enquêteur est de vérifier si elles sont possibles ou non dans le parler du témoin, et au besoin quelles sont ses stratégies de contournement.

Id Phrase	Phrase à traduire	Objectif
Séquence 1 : Les présentations.		
<i>M. Lefort, accompagné de son fils et de sa nièce, est abordé par un homme chez le boulanger.</i>		
1	Bonjour Monsieur Lefort.	remplissage
2	Bonjour. Est-ce que nous nous sommes déjà rencontrés ?	'Nous'= je+tu - forme du réfléchi P4 inclusif
3	Vous et moi, nous nous connaissons déjà.	'Vous' de politesse - forme du réfléchi P4 inclusif
4	Ah bon, nous nous connaissons ?	question totale intonative ('given') P4 - forme du réfléchi P4 inclusif
5	Oui, mais ce monsieur et moi, nous ne nous connaissons pas encore.	'Nous'=je+il - forme du réfléchi P4 exclusif
6	Vous me présenterez à lui , plus tard ?	P1objet+P3.dat fort en phrase assertive
7	Var. : *Vous me lui présenterez plus tard ?	Vérifier l'agrammaticalité de chaîne P1objet+P3.dat (me lui /lui me) en phrase assertive
8	Présente- moi à lui tout de suite, plutôt.	P1objet+P3.dat fort à l'impératif
9	Var. : *Présente-moi-lui plus tard ?	Vérifier l'agrammaticalité de chaîne clitique P1objet+P3.dat (me lui /lui me) à l'impératif
10	C'est mon fils.	'c'est' avec attribut nominal défini
11	Et la jeune fille à côté de lui est ma nièce .	structure attributive identificationnelle (Rq: essayer de l'obtenir sans dislocation)
12	On me l'a déjà présentée.	chaîne clitique P1dat+P3objet en phrase assertive
13	Var. : *On l'a déjà présentée à moi.	vérifier (a)grammaticalité de P3objet+ P1.dat fort

Figure 10 : Extrait du questionnaire commenté SyMiLa-NES

Les enquêtes elles-mêmes ont été faites là où nous avons pu trouver des témoins et des enquêteurs disponibles, sans réelle planification géographique ou sélection sociologique des témoins : les conditions actuelles d'une telle entreprise ne nous le permettaient tout simplement pas. Selon les cas, l'enquête a été conduite entièrement ou partiellement : dans ce dernier cas, seule la séquence narrative, le « Conte de l'âne peureux », axée sur la négation, a été proposée à la traduction. Il était prévu que la traduction du questionnaire soit suivie d'un entretien libre en dialecte : tous les enquêteurs ne l'ont pas conduit, souvent en raison de la fatigue du témoin ou du temps restant disponible pour l'entrevue. Bien sûr, toutes les enquêtes ont fait l'objet d'un enregistrement audio, et de documents administratifs (fiches biographiques des témoins, formulaires de consentement).

La transcription est prévue sous Praat, avec segmentation du son et transcription graphique ou phonétique alignée sur le son, comme illustré en Figure 11.

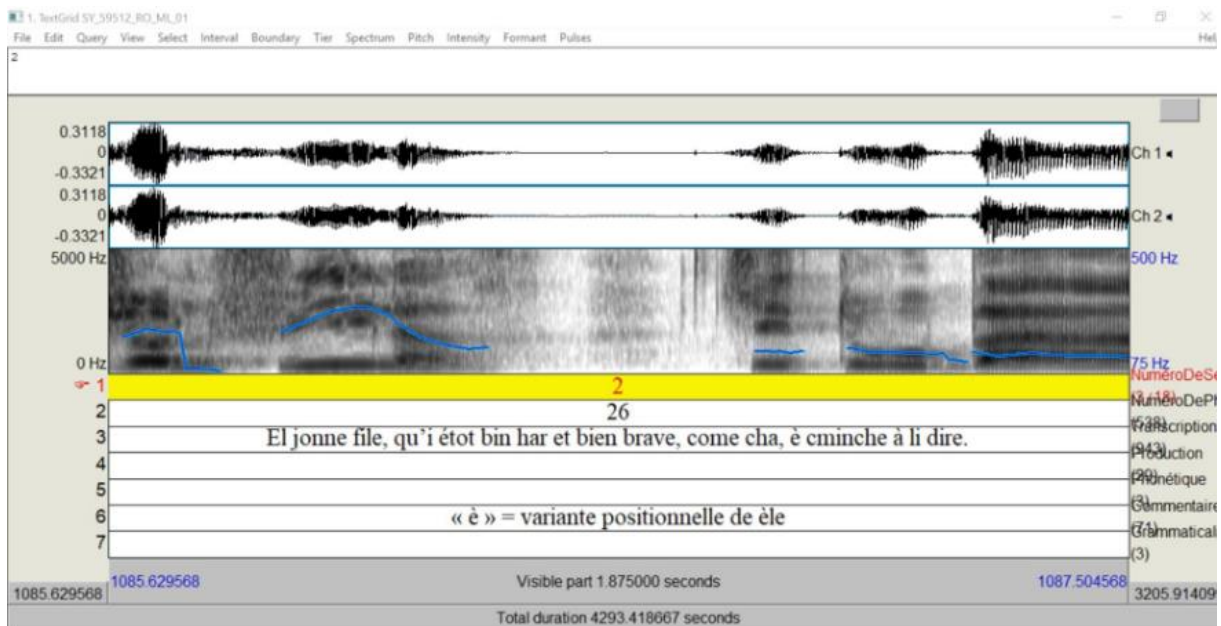


Figure 11 : Extrait de la transcription d'une enquête (Roubaix, enquête SY_59512_RO_ML_01), enquête et transcription par Liudmila Smirnova.

2.2. Avancement de la partie SyMiLa-NES

A ce jour, 107 nouvelles enquêtes syntaxiques ont été menées dans le cadre de SyMiLa, bien que sous des formes et à des états divers. Sur ces 97, 50 ont été vérifiées comme complètes et sans problème technique de son par exemple, à quoi s'ajoutent des enquêtes partielles, soit parce que l'état d'attrition linguistique chez le témoin n'a pas permis de passer l'ensemble du questionnaire, soit parce que les conditions matérielles ne le permettaient pas, soit par choix de l'enquêteur (certains se sont concentrés sur la partie narrative essentiellement ciblée sur les faits de négation, et sur les deux séquences précédentes). Leur répartition géographique est donnée dans la figure 12. Comme on peut le voir sur les figures 12 et 13, l'essentiel des enquêtes est en France, mais 13 ont été conduites en Italie (Aoste et Piémont) et 2 en Belgique, dans le Hainaut.

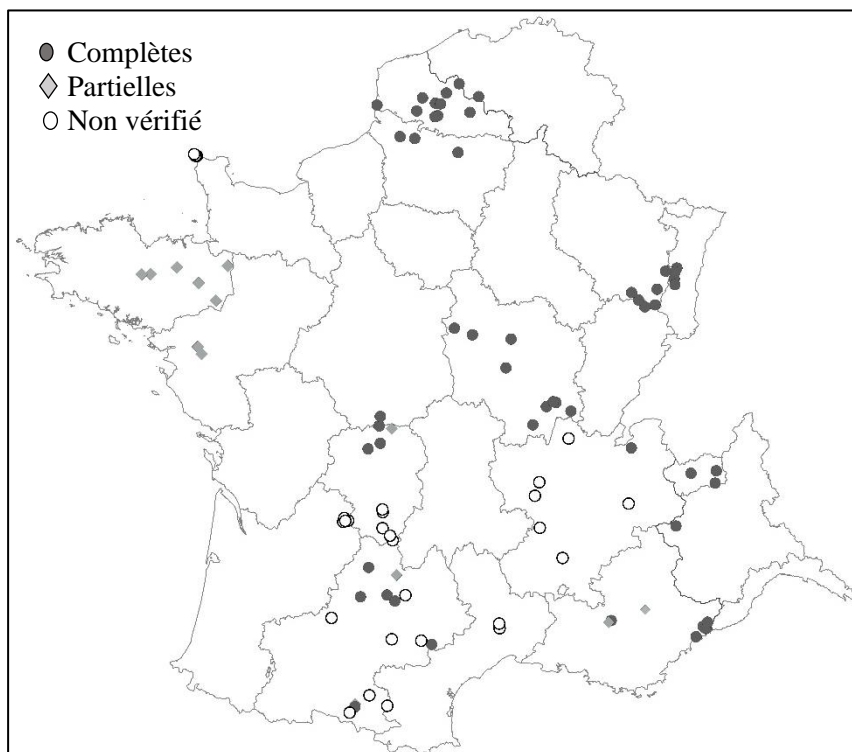


Figure 12 : Répartition des enquêtes SyMiLa-NES effectuées à ce jour

Comme on le voit, les enquêtes en domaine picard, toutes conduites et transcrites par Liudmila Smirnova, sont donc particulièrement bien représentées et constituent une part importante des enquêtes complètes – de même que celles conduites dans les Vosges et le Morvan par Benjamin Massot, et celles du Val d’Aoste conduites par Marc-Olivier Hinzelin (une dizaine, fait occulté sur la carte, du fait de l’échelle, par la proximité géographique des témoins). Elles sont au nombre de 16, dans les localités suivantes, suivi de leur code officiel géographique :

Aisne : Etreillers (02296) ;

Nord : Denain (59172), Roubaix (59512), Wavrin (59653) ;

Pas-de-Calais : Arras (62041), Auchel (62048), Berck (62108, 2 enquêtes), Bully-les_Mines (62186), Dainville (62263), Lens (62498), Maisnil (62539) ;

Somme : L’Etoile (80296), Villers-Bocage (80798) ;

Hainaut belge : Wiers (BEL-0057064), Tournai (BEL-0057081).

À cela s’ajoute une enquête partielle non transcrite, centrée sur le « Conte de l’âne peureux », aimablement conduite par Julie Auger dans le Vimeu.

Parmi l’ensemble des enquêtes, à ce jour, 28 seulement sont entièrement transcrites, essentiellement en domaine d’oïl. Toutes les enquêtes picardes sont transcrites, comme l’illustre la figure 13.

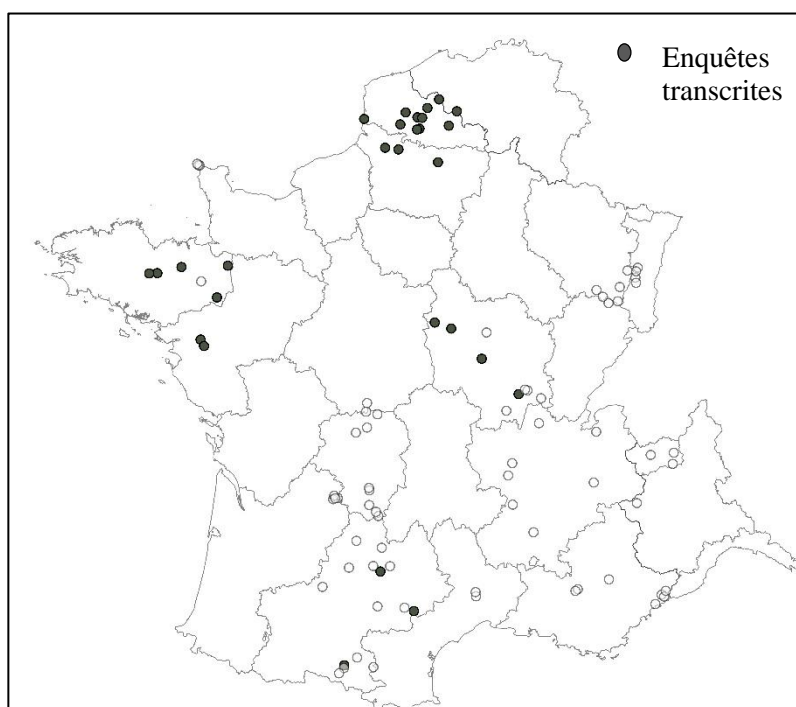


Figure 13 : Transcription des enquêtes SyMiLa-NES à ce jour

Le picard occupe donc une place de choix dans le sous-projet : dans le domaine, les enquêtes sont à la fois relativement nombreuses, complètes et transcrites, ce qui permettra à terme d’affiner la description et l’analyse de sa syntaxe et de ses variations. Malheureusement, ce travail est pour l’instant non accessible publiquement : un dysfonctionnement technique de la base de données empêche actuellement la mise en ligne de toutes les enquêtes transcrites, avec le son associé.

3. Bilan

Beaucoup a été accompli dans la durée de financement du projet SyMiLa, notamment grâce à la possibilité qu’il a offerte de recruter des postdoctorants et des personnels de soutien à la recherche, et, pour ce qui est des enquêtes, grâce à la contribution bénévole et efficace de collaborateurs extérieurs ayant une excellente connaissance de leur terrain – et, évidemment, grâce à l’incalculable participation des témoins. Une base de données à la structure spécifique au matériel à traiter et aux objectifs soit initiaux soit renforcés en cours de projet, et d’une complexité certaine, a été conçue et mise en place. Une part non négligeable des données syntaxiques de l’ALF a été saisie, et a rendu possible tant l’analyse de certains phénomènes que des modalités de cartographie nouvelles. La base a le potentiel d’accueillir de nouvelles données, par exemple issues d’atlas régionaux. Pour ce qui est nouvelles enquêtes, un

questionnaire unifié, ciblant en particulier des phénomènes spécifiques (négation, interrogatives, clitiques, syntagme nominal), a été conçu et a donné lieu à un nombre important d'enquêtes au vu du temps imparti. Que ce soit dans son volet ALF ou dans son volet NES, le picard constitue un des points forts du projet.

Beaucoup reste également à faire à l'issue de la période de financement. Certains aspects ne nécessitent essentiellement que du temps : poursuite de la saisie, vérifications, transcription des enquêtes, retour aux témoins, analyses syntaxiques, voire augmentation de la couverture géographique des enquêtes. D'autres aspects sont plus problématiques, en particulier pour ce qui dépend de l'infrastructure numérique. On touche là les limites du financement de la recherche par projets et de la réduction des postes pérennes notamment dans l'appui à la recherche. Tout ce qui touche soit aux dysfonctionnements de la BDD, comme la mise en ligne des enquêtes, soit à son évolution, comme l'amélioration de sa présentation, de son ergonomie ou de ses fonctions (recherche notamment) est tributaire de l'implication à long terme de développeurs que pour l'instant nous n'avons pas, tandis que le caractère transrégional du projet rend plus complexe le recours à des soutiens ou à des infrastructures régionales. Enfin, si la pérennité des données numériques primaires (données saisies, sons, transcriptions) semble garantie pour une dizaine d'années grâce au service d'archivage à long terme d'HumaNum, celle de la base de données elle-même, qui peut être impactée par l'évolution des outils utilisés pour son développement et sa mise en ligne, ne l'est pas.

Références

- AUGER (Julie), « Les pronoms clitiques sujets en picard: Une analyse au confluent de la phonologie, de la morphologie et de la syntaxe », dans *Journal of French Language Studies*, n° 13-1 (2003), pp. 1-22. DOI: [10.1017/S0959269503001066](https://doi.org/10.1017/S0959269503001066)
- AUGER (Julie), « Le redoublement des sujets en picard », dans *Journal of French Language Studies*, n° 13-3 (2003), pp. 381-404. DOI: <https://doi.org/10.1017/S0959269503001200>
- BAIWIR (Esther). « La particule interrogative JOU à Amiens et dans la Somme : typologie des emplois », dans *Nord*, n°70 (2017), pp.79-87.
- BLASCO (Mylène), « Les énoncés à “redoublement” et “dislocation” en français contemporain. Analyse en double marquage », dans *L'Information Grammaticale*, n° 63 (1994), pp. 42-44. DOI : <https://doi.org/10.3406/igram.1994.3085>
- BURNETT (Heather), « [Sentential Negation in the Northeastern Gallo-Romance Dialects: Insights from the Atlas Linguistique de la France](#) », dans *Journal of French Language Studies*, n°29 (2019), pp. 189-207. DOI: <https://doi.org/10.1017/S0959269519000218>
- CARTON (Fernand) & LEBÈGUE (Maurice), *Atlas linguistique et ethnographique picard*, Vol. 2. Paris, CNRS, 1998.
- DAGNAC (Anne), (2013) « How do you double your C ? Evidence from an Oïl dialect », dans PIÑON (Christopher), dir, *Empirical Issues in Syntax and Semantics*, n° 9. <http://www.cssp.cnrs.fr/eiss9/index.html> (en ligne).
- DAGNAC (Anne), « *Pas, mie, point* et autres riens : de la négation en picard », dans PITAR (Mariana) & GOES (Jan), éd., *La négation: études linguistiques, pragmatiques et didactiques*. Arras, Artois Presses Université, 2015, pp. 129-152.
- DAGNAC (Anne) & BURNETT (Heather), (2016). Concordance négative optionnelle : contrastes forts et faibles entre picard et québécois. *SHS Web of Conferences*, 27 (2016) 13003. <http://dx.doi.org/10.1051/shsconf/20162713003>.
- ÉDMONT (Edmond) & GILLIÉRON (Jules), *Atlas linguistique de la France*. Paris, Champion, 1902-1910.
- EMRIK (Robert), « La particule interrogative 'jou' », dans *Nos Patois du Nord*, n° 15 (1966/1969), pp. 2-5.
- LEGENDE (Géraldine) & CULBERTSON (Jenny), « Qu'en est-il des clitiques sujet en français oral contemporain ? », dans DURAND (Jacques), HABERT (Benoît) & LAKS (Bernard), éd., *Congrès Mondial de Linguistique Française - CMLF'08*, Paris, 2008, pp. : 2663-2674. (en ligne)
DOI: 10.1051/cmlf08308.
- MASSOT (Benjamin), « L'énigme des contraintes grammaticales sur les sujets poyaudins », dans *SHS Web of Conferences*, 27 (2016) 13005, 2016. (en ligne).
DOI: <https://doi.org/10.1051/shsconf/20162713005>
- POLETTA (Cecilia), *The Higher Functional Field. Evidence from Northern Italian Dialects*. New York, Oxford University Press, 2000.

SÉGUY (Jean), « L'accusatif prépositionnel en gascon et dans le français du sud-ouest », dans *Travaux de linguistique et de littérature* n° 11-1 (1973), Strasbourg, Centre de philologie et de littératures romanes de l'Université de Strasbourg, pp. 429-433.

¹ Le projet ANR SyMiLa est piloté par P. Sauzet (CLLE-UT2J). Il réunit deux partenaires institutionnels, le laboratoire CLLE (CNRS et Université de Toulouse 2 Jean-Jaurès) et l'institut Jean Nicod (CNRS et ENS Paris), pour lequel le pilote était initialement Dominique Sportiche, qui a entre-temps rejoint l'UCLA en Californie. Outre le PI et moi-même, ses membres titulaires à Toulouse sont N. Guilliot et J. Sibille pour CLLE. Ont également participé au projet Guylaine Brun-Trigaud, Michèle Oliviéri, Sylvain Casagrande, Eve Fumaroli et Philippe del Giudice pour l'UMR BCL (CNRS et Université de Nice) qui a conduit plusieurs enquêtes à Nice. A l'étranger, Benjamin Massot (U. Stuttgart/Munich) s'est également tout particulièrement impliqué dans le volet NES du projet dans le Morvan et les Vosges, auquel a aussi notablement contribué M.O. Hinzelin (U. Hambourg) pour le Val d'Aoste. Le financement a également permis de recruter des post-docs et personnels d'appui en CDD dédiés au projet : Heather Burnett, Rafèu Sichel-Bazin, Liudmila Smirnova (postdocs 'enquêtes et analyse'), Y. Champclaux (développement informatique), Karine Aura et Maguelone Sauzet (saisie ALF), A. Traisnel (enquêtes et coordination). Enfin, le projet a bénéficié de nombreux collaborateurs ponctuels pour la conduite de diverses enquêtes, que nous ne saurions tous remercier faute de place.

² Ces cartes sont généralement accompagnées de mentions marginales pour des points de la moitié nord, pour lesquels l'enquêteur a probablement pu recontacter les témoins pour poser les questions correspondantes a posteriori ; plusieurs sont en domaine picard, et au premier chef le point 284.

³ La base explicite les correspondances API / alphabet Rousselot-Gilliéron de manière en principe exhaustive, et permet au besoin un retour à l'original ; sur ce point, voir : http://symila.univ-tlse2.fr/alf/notation_phonetique.

⁴ Sauf exception, je glose ici les occurrences dialectales sous la forme orthographique française, afin de faire apparaître les cohérences morpho-syntaxiques parfois masquées par la variation phonétique et phonologique.