HAL
open science

Actes des rencontres lyonnaises des jeunes chercheurs en linguistique historique

Timothée Premat, Ariane Pinche, Afra Pujol I Campeny, Iris Fabry, Prunelle Deleville, Vincent Surrel, Jérémy Rouquier, Enzo Pongio

To cite this version:

Timothée Premat, Ariane Pinche (Dir.). Actes des rencontres lyonnaises des jeunes chercheurs en linguistique historique. Timothée Premat; Ariane Pinche. Association Diachronies Contemporaines, 2019, 10.5281/zenodo.3462309 . halshs-02317116v2

HAL Id: halshs-02317116

https://shs.hal.science/halshs-02317116v2

Submitted on 26 Oct 2019

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.

Open licence - etalab

Actes des

Rencontres lyonnaises des jeunes chercheurs en linguistique historique

Édités par

Timothée Premat
Ariane Pinche

Diachronies Contemporaines

PREMAT, Timothée & PINCHE, Ariane (dir.) (2019). *Actes des rencontres lyonnaises des jeunes chercheurs en linguistique historique*. Lyon : Diachronies contemporaines, 70 p.

DOI : 10.5281/zenodo.3462309

Préambule

Le 6 juin 2019, se tenait à l'Université Jean Moulin Lyon 3 la première édition des *Rencontres lyonnaises des jeunes chercheurs en linguistique historique*. Cet événement, organisé par l'association *Diachronies contemporaines*, a regroupé de jeunes chercheurs et chercheuses qui, par leurs affiliations, leurs nationalités et leurs objets d'étude, ont apporté à ces rencontres une richesse à l'image de celle de la communauté linguistique gallo-romane. En effet, cette journée a été l'occasion de réunir des chercheurs et chercheuses de toute l'Europe de l'Ouest, de l'université la plus septentrionale du monde (Tromsø) à l'Université de Vérone, mais aussi de révéler la diversité de notre communauté scientifique, notamment à travers le large spectre des approches mobilisées, tant d'inspiration générative que plus traditionnelles. La diversité des communications présentées illustre aussi l'ampleur spatio-temporelle de la Gallo-Romania, prise dans son extension la plus large. En ce sens, les langues étudiées allaient du catalan médiéval au ladin contemporain, en passant par le gascon béarnais, l'occitan vellave et le français, ce dernier étant embrassé dans une diachronie particulièrement large, de la fin de l'antiquité à la période moderne.

Nous sommes fiers de reproduire ici le contenu de certaines des présentations de ces *Rencontres lyonnaises*[1]. Celles-ci vont de la phonologie à la morphologie, de la syntaxe à la dialectologie, de la lexicologie et de l'exercice de la philologie traditionnelle à la mise en place d'outils novateurs en philologie et humanités numériques. Ainsi, dans la continuité de la richesse de cette journée, ces travaux couvrent les champs les plus habituels de la discipline, mais s'ouvrent également à des outils numériques en cours de développement dans un champ précurseur qui, dès les années soixante, avait compris tout l'intérêt de l'émergence de cette nouvelle « transdiscipline »[2] que nous appelons aujourd'hui les *Humanités numériques* et qui semble enfin acquérir ses lettres de noblesse[3]. Avoir réussi à regrouper dans le même événement linguistes, littéraires travaillant sur la variation de la langue et des textes, éditeurs et acteurs de l'ingénierie des humanités numériques s'est avéré garant d'échanges passionnants — échanges que nous espérons pouvoir transmettre et diffuser par la publication de ces actes.

Enfin, nous sommes également très heureux de compter parmi nos auteur·trices deux étudiant·es de master, qui bénéficient ainsi de l'occasion de présenter pour la première fois leurs travaux. Accompagner les étudiant·es dans leur parcours de recherche, au-delà du cadre didactique mis en place par les universités, est aussi l'une des missions fondamentales de l'association *Diachronies contemporaines*.

Nous espérons que ce volume n'est que le premier d'une longue série, qui reproduira d'année en année la tribune que nous avons voulu donner aux jeunes chercheurs et chercheuses de notre domaine.

[1] Le programme de l'ensemble de l'événement est disponible sur le site de l'association, à l'adresse : https://diachro.hypotheses.org/1302.

[2] Marin Dacos, « Manifeste des *digital humanities* » , *That Camp Paris 2010 : Non-Conférence sur les Digital Humanities,* url : http://tcp.hypotheses.org/318, consulté le 15 septembre 2019.

[3] « One may affirm that the Digital Humanities has won its struggle to obtain official recognition », Tino Orlandi, *Reflections on the Development of Digital Humanities (Busa Lecture*), DH 2019, 12 juillet 2019, Utrecht.

L'association Diachronies contemporaines

L'association *Diachronies contemporaines* est née d'un double constat. D'une part, à l'heure de l'éclosion des 'labos juniors', aucune structure de ce type ne s'intéressait spécifiquement à la linguistique gallo-romane, pourtant constituée en champ scientifique cohérent par ailleurs. D'autre part, nous étions quelques-uns sur le site universitaire de Lyon à travailler sur ce domaine, mais nous étions dispersés entre différentes universités, laboratoires et écoles doctorales. C'est pour ces raisons qu'en février 2018 l'association *Diachronies contemporaines* a été fondée par Piotr Płocharz, Timothée Premat, Vincent Surrel et Fabian Zuk, très vite rejoints par Ariane Pinche. Après avoir co-organisé avec le département de Linguistique de l'Université Jean Moulin Lyon 3 des journées doctorales le 4 avril 2018, l'association s'est mobilisée pour organiser en juin 2019 les premières *Rencontres lyonnaises*, sous la coordination de son président, Timothée Premat.

Les *Rencontres lyonnaises des jeunes chercheurs en linguistique historique* ont été soutenues par les laboratoires CIHAM (UMR 5648), IRHIM (UMR 5317) et SFL (UMR 7023), par l'école doctorale 3LA de l'Université de Lyon, par le FSDIE et le département de Linguistique de l'Université Jean Moulin Lyon 3 ainsi que par la Métropole de Lyon.

Timothée PREMAT & Ariane PINCHE

Sommaire

Le mot *sí* en catalan ancien : une particule de polarité positive 4
Afra PUJOL I CAMPENY

La possession externe : trouve-t-on un emploi plus large de la construction ? Recherches sur un français classique non littéraire 15
Iris FABRY

Entre ancien et moyen français : le traitement de l'*Ovide moralisé* 22
Prunelle DELEVILLE

Variation textuelle et variation linguistique dans les textes occitans d'Antoine Clet 31
Vincent SURREL

Étude diachronique du gascon béarnais au sein des textes officiels Quelle(s) variation(s) par rapport au gascon ? 40
Jérémy ROUQUIER

Annoter facilement un corpus complexe, L'exemple de Pyrrha, interface de post correction, et Pie, lemmatiseur et tagueur morphosyntaxique, pour l'ancien français 48
Ariane PINCHE

Le PAM, un Programme d'Analyse Métrique pour le français médiéval 59
Enzo POGGIO & Timothée PREMAT

Le mot *sí* en catalan ancien : une particule de polarité positive

Afra PUJOL I CAMPENY
Batista i Roca Fellow, Fitzwilliam College
University of Cambridge

Résumé

Dans cet article nous explorons la distribution syntaxique et la fonction du mot *sí* en catalan ancien, en démontrant que déjà au XIIIe siècle, il avait été grammaticalisé en particule de polarité emphatique et présentait une syntaxe innovante, parallèle à celle de l'espagnol moderne, contrairement aux analyses de *sí* proposées pour l'italien ou le français ancien.

Mots clés : catalan ancien ; polarité positive ; *sí*.

DOI : 10.5281/zenodo.3464460

1. Introduction

Une des caractéristiques qui éloigne les langues romanes de leur source, le latin, est la façon dont la polarité positive est exprimée : alors que le latin n'avait pas de particule de polarité positive emphatique, les langues romanes en ont développées plusieurs. Le catalan et l'occitan ancien ont grammaticalisé le démonstratif latin HOC en *òc* 'oui' ; en français le syntagme HOC ILLUD a été grammaticalisé en *oui*, tandis que *si*, qui provient de l'adverbe latin *SIC* 'ainsi', a été grammaticalisé comme *sí* 'oui' en catalan, espagnol et italien, et comme *sim* 'oui' en portugais.

La grammaticalisation de l'adverbe latin SIC, un adverbe de manière situé dans le Syntagme Verbal (SV), en Particule de Polarité Positive Emphatique (PPPE) en catalan a été étudiée par Batllori & Hernanz (2008 ; 2013). Selon ces auteures, pendant la période médiévale, *sí*, encore un adverbe généré dans SV, était simplement utilisé pour renforcer l'expression de quelques stratégies pour exprimer la polarité positive emphatique (Batllori & Hernanz, 2008 : 14), comme dans (1), alors que dans la langue moderne et après avoir subi un processus de grammaticalisation, *sí* est généré dans la périphérie gauche, dans une projection qui héberge des traits liés à la polarité en tant que Particule de Polarité Positive Emphatique ou PPPE, comme dans (2) :

(1) Old Catalan & Old Spanish

$[_{\text{ForceP}}\ [_{\text{TopP}}\ [_{\text{FocP}}\ \text{sí}_i\ [_{\text{PolP}}\ t_i\ [_{\text{FinP}}\ [_{\text{TP}}\ [_{\text{VP}}\ t_i \ldots$

(2) Modern Spanish

$[_{\text{ForceP}}\ [_{\text{TopP}}\ [_{\text{FocP}}\ \text{sí}_i\ [_{\text{PolP}}\ t_i\ [_{\text{FinP}}\ [_{\text{TP}}\ [_{\text{VP}} \ldots$

L'analyse de la grammaticalisation de *sí* de Batllori & Hernaz (2008 ; 2013), laisse néanmoins quelques questions en suspens : à quel moment *sí* cesse d'être un adverbe de manière pour devenir une PPPE ? Quelle était sa distribution syntaxique exacte ? Dans cet article, je répondrai à ces questions à l'aide de données extraites de *El Llibre dels Fets*, une chronique historique écrite en catalan ancien au XIIIe siècle, dans le cadre du programme cartographique (Rizzi, 1997). La confirmation de l'analyse de Batllori & Hernanz réaffirme que *sí* en catalan ancien

ne peut pas recevoir l'étiquette d'élément explétif, une proposition assez fréquente pour le statut de *sí* dans des langues romanes qui ont été analysées comme V2 (Poletto, 2005 ; Ledgeway, 2008 ; Wolfe, 2015 ; Wolfe, 2018).

2. Polarité positive emphatique en latin

Le latin ne présentait pas une proforme positive de phrase comme *oui* en français ou *sí* en catalan. Néanmoins, cette langue possédait de nombreuses expressions qui pouvaient exprimer la polarité positive, comme ENIM (Spevak, 2012 : 201), QUIDEM (Danckaert, 2014) et SIC (Batllori & Hernanz, 2008) :

(3) Dissimilis **quidem** Chares horum et factis
différent certes Chares ses et faits.ABL

et moribus sed tamen Athenis et
et habitudes.ABL mais néanmoins Athènes.ABL and

honoratus et potens.
honorable et puissant

Certainement Chares se distinguait d'eux par ses actions ainsi que par son caractère, mais, néanmoins, il était un homme honorable et puissant.

Apud Danckaert (2014 : 98)

(4) Aliud quid ? **Etiam** : quando te proficisci
autre quoi oui quand toi aller.INF.PRES.PASS

istinc putes fac ut sciam.
d'ici penser.2SG fais.2SG.PRES.IMP.ACT comme sache.1SG.PRS.SBJV

Quoi d'autre ? Oui : quand tu penses à répondre à cette lettre, fais-le-moi savoir.

Cicero, *Epistulae ad Atticum*, 2.6.2.15

(5) {IVPP} Heia autem inimicos ?
inter néanmoins ennemis.ACC

{ALC.} **Sic** est, vera praedico.
ainsi est choses vraies.ACC proclame

Jupiter : Ha ! Des ennemis, vraiment ?
Alcmène : C'est ainsi, je dis la vérité.

Plautus, *Amphytruo*, 901

Outre ces particules, le latin possédait aussi des stratégies syntaxiques pour exprimer la polarité positive emphatique : les phrases à verbe initial (Devine and Stephens, 2006) et les réponses à des questions directes avec écho du verbe, comme illustré en (6) et (7) respectivement.

(6) **Contempsisti** L. Murenae genus, **extulisti** tuum
méprises.2sg L. Murena.GEN clan as loué le tien

Tu as certainement *méprisé le clan Murena, tu as* certainement *loué le tien.*

Cicero, *Pro Murena*, 7.15

(7) Clodius insidias fecit Miloni ? **Fecit**.
Clodius complot.ACC fait.3SG Milo.DAT fait

Clodius a-t-il fomenté un complot contre Milo ? Oui, il l'a fait.

Cicero, *Pro Milone*, 60.3

L'exemple (6) contient deux cas de polarité positive emphatique exprimée par V1. Dans la première phrase, le verbe CONTEMPTISTI apparait en position initiale, comme celui de la phrase suivante, EXTULISTI, pour fortement affirmer les deux actions et les contraster. Selon Devine & Stephens (2006 : 147), le mouvement du verbe à la périphérie gauche de la phrase, qui résulte en un ordre à verbe initial, est déclenché par la présence d'un opérateur associé à l'expression de la polarité positive qu'ils nomment POSPOL. L'activation de POSPOL implique la focalisation de la valeur véridique de la phrase tout en en excluant le partenaire négatif, autrement dit *Verum* Focus (Leonetti & Escandell-Vidal, 2009 ; Krifka, 2008). L'exemple (7) contient une question directe dont la réponse est un cas d'écho du verbe, qui apparaît toujours en l'absence de ses compléments. Cette particularité peut s'expliquer aussi par la montée du verbe à la périphérie gauche, comme nous le verrons dans les sections qui suivent.

3. La grammaticalisation de *sí*

Dans la langue latine, SIC 'ainsi', l'adverbe qui sera ré-analysé comme *si*, présente la distribution et le comportement d'un adverbe : contrairement aux particules polaires ENIM ou QUIDEM, qui obéissent à la loi de Wachernagel, SIC peut apparaître au début de la phrase et ne doit pas être adjacent au verbe.

D'après Batllori & Hernanz (2008), le contexte de ré-analyse de *sí* en catalan et en espagnol anciens correspond à des réponses à l'écho du verbe et avec verbes de support (similaires au *do*-support anglais) à des questions directes et indirectes, comme l'illustre l'exemple (8) :

(8)	E	dixo :	¿**es**	este	el	vuestro	hermano	el	menor
	et	dit	est	ce	le	votre	frère	le	mineur
	que·m			dixiestes ?		E	dixieron :	<u>Sí</u>	**es**.
	qui à moi			dîtes		et	dirent	oui	est

Et il dit : c'est lui, votre frère mineur dont vous me parlâtes ? Et ils répondirent : Oui, c'est lui.

Fazienda, fol. 8v, *apud* Rodríguez Molina (2014 : 874)

Selon Batllori & Hernanz (2013 : 14), les réponses à *do*-support et à l'écho du verbe impliquent la montée de l'adverbe *sí* à SFoc dans la périphérie gauche, après avoir été générées dans le domaine verbal. C'est dans ce contexte que l'adverbe devient ré-analysé comme PPPE.

Cette hypothèse implique l'existence de deux têtes fonctionnelles liées à l'expression de polarité dans les phrases principales : Syntagme de Polarité (SPol) et Syntagme Σ (terminologie de Laka (1990)). D'une part, SPol, dans le domaine la périphérie gauche, contient des traits de polarité relative. Les traits de polarité relative expriment l'accord ou la négation d'une assertion par rapport à une assertion précédente ([*reverse*] et [*same*] dans la terminologie de Farkas & Bruce (2010)). D'autre part, SΣ, la tête fonctionnelle plus élevée dans la projection temporelle, contient des traits de polarité absolue du prédicat de la phrase : [+affirmation] et [+négation]. L'activation de SPol pour exprimer la polarité positive emphatique peut prendre plusieurs formes. En espagnol ancien, ainsi qu'en galicien et en portugais modernes, cette tête attire le verbe vers elle pour exprimer la force assertive des phrases déclaratives non-marquées (Martins, 2013 ; Rodríguez Molina, 2014 ; Sitaridou, 2019).

Syntaxiquement, la montée du verbe en SPol implique la possibilité de générer des réponses à l'écho du verbe et au *do*-support. Dans ces réponses, la polarité de la phrase n'est pas simplement positive, mais également emphatique : le verbe monte de SPol à SFoc pour activer une lecture de *Verum* Focus sur la phrase qui supprime les lectures négatives potentielles de celle-ci. En outre, Martins (2013) constate que les langues qui permettent la montée du verbe à SPol dans des phrases déclaratives non-marquées ne possèdent pas de proforme de polarité positive capable de répondre à une question directe ou indirecte fermée, comme c'est le cas pour l'espagnol ancien.

En espagnol ancien, les réponses à écho du verbe et à *do*-support peuvent être accompagnées par l'adverbe *sí*, qui contribue à l'expression emphatique de la polarité positive de la phrase. Selon Rodríguez Molina (2014), *sí* commence effectivement comme un adverbe qui renforce l'effet emphatique de la montée du verbe, et qui sera finalement ré-analysé comme PPPE, générée directement dans la périphérie gauche à la suite d'un processus de focalisation qui implique sa montée en SFoc avec le verbe :

(9) a. $[_{CP} \ldots [_{FocP} \ldots [_{PolP} \ldots [_{TP} \ldots [_{VP}$ fago así$]]]]]$

b. $[_{CP} \ldots [_{FocP} \ldots [_{PolP} \ldots [_{TP}$ así/sí$_i$ $[_{VP}$ fago $t_i]]]]$

c. $[_{CP} \ldots [_{FocP}$ así/sí$_i$ $[_{PolP}$ t_i $[_{TP} \ldots [_{VP}$ fago $t_i]]]]]$

d. $[_{CP} \ldots [_{FocP}$ sí $[_{PolP} [_{TP} [_{VP}]]]]]$

Rodríguez Molina (2014 : 897)

Dans le cas de réponses à l'écho du verbe et de réponses à *do*-support[1], la présence de *sí* peut accompagner le verbe dans sa montée en SFoc. La montée du verbe ainsi que de *sí* en SFoc permettrait l'élision du complément de SFoc et empêcherait l'apparition de pronoms clitiques dans les réponses à l'écho du verbe et à *do*-support, comme l'illustre (10) :

(10) ¿Conosces esta tierra donde estamos ?
connais-tu cette terre où nous sommes

$[_{ForceP} [_{TopP} [_{FocP}$ Sí$_j$ $[_{Foc'}$ conozco$_i$ $[_{PolP}$ t_i $[_{Pol'}$ t_i $[_{FinP} [_{TP} [_{T'}$ t_i
oui je connais

$[_{VP}$ t_i ~~esta tierra~~ $t_j]]]]]]]]]$
cette terre

Est-ce que tu connais cette terre où nous sommes ? Oui, je la connais.

Amadís, 1.781, *apud* Rodríguez Molina (2014 : 891)

Selon Rodríguez Molina (2014), qui suit Batllori & Hernanz (2008), la grammaticalisation de *sí* en PPPE en espagnol ancien a eu lieu vers la fin du XIVe siècle. Avant cette date, les cas où *sí* apparaissait étaient limités aux réponses à l'écho du verbe et à *do*-support.

[1] Les réponses à *do*-support ne sont pas des phénomènes uniquement ibéro-romans : elles se trouvent aussi dans dialectes italiens du nord (Benincà & Poletto, 2004) et dans le français ancien (Roberts, 1993).

4. *Sí* dans *El Llibre dels Fets* : une particule de polarité positive

La distribution de *sí* en catalan ancien diffère substantiellement de celle du français ancien ou de l'espagnol ancien. Dans la suite de cet article, nous allons montrer que dès le XIIIe siècle, le catalan ancien présente déjà un *sí* grammaticalisé comme PPPE et dont la distribution syntaxique est significativement plus novatrice qu'en espagnol ancien.

Au XIIIe siècle, le catalan ancien ne présente pas de mouvement du verbe dans les phrases déclaratives non-marquées (Pujol i Campeny, 2018). Au contraire, comme le catalan moderne, il s'agit d'une langue SVO dont le verbe reste dans le domaine temporel dans des phrases non-marquées. Le catalan ancien, contrairement aux langues voisines et, encore une fois comme le catalan moderne, possédait aussi une proforme de phrase de polarité positive, partagée avec l'occitan : *òc*, dérivé du pronom démonstratif latin HOC. Cette proforme pouvait uniquement agir comme proforme de phrase qui répond à des questions directes ou indirectes fermées : elle n'apparait jamais comme renfort de réponses à l'écho du verbe ou à *do*-support. Néanmoins, elle ne peut pas prendre une phrase avec *que* en tête comme complément :

(11) E dixem -li : "Eres tu ab lo comenador?"
et dîmes=lui es tu avec le commandant

e él dix : "seyor, <u>och</u>".
et il dit monsieur oui

Et nous lui dîmes « es-tu avec le commandant ? », et il dit « oui, monsieur ».

Fol. 103r, l. 16

En catalan ancien, comme dans les autres langues romanes anciennes, la particule *sí* était homophone avec d'autres éléments : la conjonction subordonnante *si* « si », le pronom réflexif de 3e personne du singulier *si* « soi », qui, comme en français, apparait fréquemment accompagné de *mateix* « même » et après des prépositions, et à l'occasion, avec l'adverbe *si* « ainsi ». Après avoir écarté ces éléments homophones de *sí* de *El Llibre dels Fets*, il reste 19 cas de *sí* comme PPPE dans des phrases complètes (sans élision du SV). Syntaxiquement, ces phrases présentent trois caractéristiques : (i) *sí* est toujours adjacent au verbe, (ii) les pronoms clitiques objet et adverbiaux sont les seuls éléments qui peuvent apparaitre entre *sí* et le verbe, et (iii) *sí* peut être précédé par des topiques suspendus. Ces caractéristiques sont illustrées dans les exemples (12-15) :

(12) E metem mà a ·iii· guaytes fer.
et mettons main à 3 gardes faire

[La una guayta] **sí** era als genys, e a les cledes
la un garde si était aux machines et à les palissades

la altra, **sí** era contra la porta de Barbelec.
la autre si était contre la porte de Barbalec

Et nous décidions de mettre en place 3 sentinelles : l'une était effectivement aux catapultes et aux palissades, l'autre était effectivement contre la porte de Barbalec.

Fol. 48r, l. 19

(13) Si homes del món àn mala
si hommes du monde ont mauvaise

fama, [nós] **sí** la havem bona.
réputation nous si la avons bonne

S'il y a des hommes dans le monde ayant une mauvaise réputation, la nôtre est certainement bonne.

Fol. 29r, l. 10

(14) E **sí** hi ach altres paraules que començaven
et si y a autres discours que commençaient

ja de dir mas nós ho vedam.
déjà de dire mais nous cela interdîmes

Et il y avait effectivement d'autres discours qui commençaient déjà à apparaître, mais nous interdîmes cela.

Fol. 10v, l. 29

(15) E anch per aquesta paraula nengú no·s
et jamais pour cette intervention personne non se

moch, e **sí** la hoïren tots.
bougeait et si la entendirent tous

Et personne ne bougeait à cause de cette intervention, et ils l'avaient pourtant tous entendue.

Fol. 49r, l. 13

Dans (12), *sí* apparait entre le sujet topique *la una guayta* « un des gardes » et le verbe *era* « était ». Dans (13), *sí* est une fois de plus précédé par le sujet de la phrase, *nós* « nous ». Cependant, dans (14) *sí* apparait en première place dans la phrase, précédé uniquement par la conjonction *e* « et », et est suivi par le pronom clitique adverbial *hi* « y », qui se trouve entre *sí* et le verbe. Finalement, dans (15) le pronom clitique objet *la* apparait entre *sí* et le verbe, et le sujet de la phrase *tots* « tous » se trouve en position postverbale. Ces exemples montrent clairement que *sí* peut apparaitre dans de phrases complètes sans élision du SV, étant donné que *sí* coïncide avec des éléments dont la présence requiert la projection de SV comme les pronoms clitiques ou les sujets postverbaux.

Puisque la distribution de *sí* ne se limite pas aux environnements décrits pour l'espagnol ancien (elle n'est pas limitée aux réponses de questions fermées) ni ne peut être associée à une grammaire V2 (Pujol i Campeny, 2018), il faut identifier la raison de l'apparition de cette particule. Toutes les occurrences de *sí* dans *El Llibre dels Fets* ont un trait en commun : elles apparaissent dans un contexte non-véridique (Giannakidou, 1998 ; 2008) qui autorise la présence d'éléments polaires, définis de la façon suivante :

(16) ÉLEMENT POLAIRE

a. Un élément polaire α est une expression dont la distribution est limitée par la sensibilité à une propriété sémantique β du contexte où il apparait.

b. β est la non-véridicité.

Un opérateur non-véridique ne requiert pas qu'une proposition soit vraie dans un des modèles (ensemble de mondes) d'un contexte spécifique, et crucialement, l'opérateur ne comporte pas la fausseté d'une proposition.

Les éléments polaires appartiennent à deux catégories : celle des éléments polaires négatifs, et celle des éléments polaires positifs. Les éléments polaires positifs n'apparaissent pas que dans les contextes assertifs : ils peuvent aussi apparaitre dans les contextes non-véridiques (Ernst, 2009). Giannakidou (1998) identifie huit contextes non-véridiques :

(17)

a. Quantifieurs monotones ;

b. Verbes modaux ;

c. Verbes d'intention ;

d. Phrases non-déclaratives : interrogatives, impératives et exclamatives ;

e. Protases des phrases conditionnelles, phrases conditionnelles, phrases avec *avant que* ;

f. Temps verbaux habituels ;

g. Temps verbaux futurs ;

h. Implicatures non-véridiques ;

Tous les cas de *sí* dans des phrases complètes de *El Llibre dels Fets* se trouvent sans exception dans un contexte non-véridique, qu'il soit explicite ou implicite. Nous allons illustrer cette affirmation avec les exemples (18-21) présentés ci-dessous (pour une revue complète des contextes non-véridiques dans lesquels *si* apparait voir Pujol i Campeny (à paraître)).

(18) Exemple (12) :
Contexte non-véridique : quantifieurs monotoniques « trois » et « une »

(19) Exemple (13) :
Contexte non-véridique : protase précédente.

(20) Exemple (14) :
Contexte non-véridique : implicature dérivée du contexte.

Résumé : Le Roi convoque la cour dans la ville de Montsó. Il interdit aux chevaliers et aux grands hommes qui y assistent de porter une arme. Le chevalier Guillem de Muntcada, qui avait planifié un coup contre le Roi Jacques I, se dirige vers Montsó avec tous ses hommes, mais l'entrée de la ville leur est refusée. Malgré l'arrivée de Guillem de Muntcada avec ses hommes, les interventions commencent, mais le Roi l'interdit.

(21) Exemple (15) :
Contexte non-véridique : implicature dérivée de la phrase précédente. Étant donné que personne n'a bougé après le discours, il est possible d'inférer que le public n'a pas entendu le discours, bien que ce ne soit pas le cas.

Il n'y a, alors, pas de doute que la présence de *sí* dans des phrases complètes n'est avérée que dans la portée d'un opérateur non-véridique. Sa fonction est celle d'affirmer la vérité de la phrase déclarative contre le contexte non-véridique qui pourrait suggérer le contraire. Par conséquent, on peut établir que *sí* est un élément de polarité positive.

En ce qui concerne la structure syntaxique des phrases qui présentent *sí*, elle doit expliquer (i) son adjacence au verbe à l'exception des pronoms clitiques, (ii) la possibilité que *sí* soit précédé par des topiques et des éléments scéniques, (iii) la non-élision de SV.

Le catalan ancien, comme le catalan moderne, possède un genre de focalisation lié à l'assertion emphatique d'une phrase contre ses possibles lectures négatives dans des contextes non-véridiques. Cette structure porte le nom de *Verum* Focus (Krifka, 2008 ; Escandell Vidal, 2009 : 182). Dans Pujol i Campeny (2018), nous proposons que la structure syntaxique des phrases qui présentent *sí* est parallèle à celle des phrases qui présentent un *Verum* Focus, qui englobe : (i) le mouvement en SpecSFoc d'un constituant ou une tête ; (ii) le mouvement du verbe en SPol pour lui donner la valeur [+]. Le résultat de cette proposition est l'absence d'éléments autres que des pronoms clitiques entre *sí* et le verbe, et la possibilité que *sí* soit précédé par des éléments qui se trouveraient dans le champ scénique. (22) et (23) montrent comment cette structure s'applique autant au *Verum* Focus qu'aux phrases avec *sí* :

(22) Implicature négative qui évoque une version négative de la phrase, sous-entendue :

Vous n'allez pas me dire ce que vous pensez de ce sujet.

Réponse :

a.

'AÇO	us	**diré**	yo: (…)'
cela	vous=	dirai	JE

Je vous dirai cela : (…)

Fol. 29r, l. 8

b. $[_{CP}$ ... $[_{FocP}$ $[_{SpecFocP}$ AÇO $[_{F'}$ $[_{Foc}$ us diré$_i]$ $[_{PolP}$ $[_{Pol'}$ $[_{Pol}$ Pol + $t_i]]$ $[_{IP}$ $[_{SpecIP}$ yo$_j$ $[_{T'}$ $[_{T}$ $t_i]$ $[_{vP}$ $[_{Spec,vP}$ pro$_j$ $[_{V'}$ $[_{V}$ t_i]]]]]]]]]]]]]

(23)

a.

E	anch	per	aquesta	paraula	nengú	no·s
et	jamais	pour	cette	intervention	personne	non se

moch,	e	**sí**	la	hoïren	tots.
bougeait	et	si	la	entendirent	tous

Et personne ne bougeait à cause de cette intervention, et ils l'avaient pourtant tous entendue.

Fol. 49r, l. 13

b. $[_{CP}$ ... $[_{FocP}$ $[_{SpecFocP}$ SI $[_{F'}$ $[_{Foc}$ la hoïren$]$ $[_{PolP}$ $[_{Pol'}$ $[_{Pol}$ Pol + $t_i]]$ $[_{IP}$ $[_{SpecIP}$ tots$_j$ $[_{T'}$ $[_{T}$ $t_i]$ $[_{vP}$ $[_{Spec,vP}$ pro$_j$ $[_{V'}$ $[_{V}$ t_i]]]]]]]]]]]]]

Comme illustré en (23.b), nous postulons que dès le XIIIe siècle *sí* a perdu son sens adverbial, et est donc généré directement en SPol. La présence de *sí* affirme la « positivité » de la phrase en question contre un contexte non-véridique et bloque la possibilité de génération de *no* « non » dans SΣ, puisque SPol et SΣ doivent s'accorder pour exprimer une polarité positive emphatique.

Par conséquent, il peut être établi qu'au XIIIe siècle *sí* était déjà une PPPE, deux siècles avant que cette nouveauté n'apparaisse en espagnol comme l'indique Rodríguez Molina (2014).

Le catalan ancien, comme les autres langues romanes, présente aussi des cas de réponses à l'écho du verbe et à *do*-support. Néanmoins, dans *El Llibre dels Fets*, elles apparaissent toujours accompagnées de la PPPE *sí*, sans exception :

(24) E nós volguem saber dels altres si eren
et nous voulûmes savoir.INF des autres si étaient

en aquel consel, e atorgaren tots que **sí** eren.
dans cette réunion et accordèrent tous que si étaient

Et nous voulûmes savoir si les autres étaient dans cette réunion, et ils affirmèrent tous qu'ils y étaient certainement.

Fol. 104v, l. 26

(25) E nós resposem que : "Ço que nós n' avem feyt, que
et nous répondîmes que ce que nous en avons fait que

u havem feyt ab dret", e que neguna esmena no li ·n
ce avons fait avec droit et que aucun amendement ne lui=en

faríem. E éls dixeren que **sí** faéssem, (…)
ferions et ils dirent que sí fissions.

Et nous répondîmes : « ce que nous en avions fait, nous l'avions fait légalement », et que nous ne ferions aucun amendement [au texte]. Et ils dirent que certainement nous allions le faire.

Fol. 13r, l. 21

Le fait que *sí* soit nécessaire même dans des structures qui en latin et dans d'autres langues contemporaines suffisent pour exprimer la polarité positive emphatique est une preuve de plus que *sí* était déjà grammaticalisé comme PPPE et était toujours présent dans l'expression de polarité positive emphatique. Néanmoins, les réponses à l'écho du verbe et à *do*-support partagent les caractéristiques syntaxiques de ce genre de réponses en espagnol ancien : la montée du verbe à SFoc et l'élision du complément de SFoc.

En conséquence, il est possible d'établir que si *sí* était déjà grammaticalisé en catalan ancien et que sa distribution n'était pas limitée aux réponses à l'écho du verbe et à *do*-support, il était un élément nécessaire pour l'expression de la polarité positive emphatique. Les réponses à l'écho du verbe et à *do*-support n'étaient plus le seul moyen de répondre positivement et emphatiquement, mais elles étaient encore utilisées dans le discours direct, où les structures archaïques sont favorisées.

5. Conclusion

Dans cet article nous avons montré que la grammaticalisation du mot *sí* en catalan ancien date du XIIIe siècle, puis que sa distribution s'était déjà étendue hors du contexte de grammaticalisation, les réponses à l'écho du verbe et à *do*-support, et qu'il était essentiel pour l'expression de la polarité positive emphatique.

Bibliographie

Batllori, M. & Hernanz, M. L. (2013). « Emphatic Polarity Particles in Spanish and Catalan ». *Lingua* 128, p. 9–30.

Batllori, M. & Hernanz, M. L. (2008). « Emphatic Polarity from Latin to Romance ». New York [ms].
Url : http://habilis.udg.edu/~info/Activitat_Docent_Recerca_Gestio/Montserrat_Batllori/BatlloriHernanzGIST4Gent.pdf.

Benincà, P. &, Poletto, C. (2004). « A Case of Do-Support in Romance ». *Natural Language & Linguistic Theory*, 22(1), p. 51–94.

Danckaert, L. (2014). « *Quidem* as a Marker of Emphatic Polarity ». *Transactions of the Philological Society*, 112(1), p. 97–138.

Devine, A. M. & Stephens, L. D. (2006). *Latin Word Order*. (1ère édition), Oxford : Oxford University Press.

Ernst, T. (2009). « Speaker-Oriented Adverbs ». *Natural Language & Linguistic Theory*, 27(3), p. 497–544.

Farkas, D. F. & Bruce, K. B. (2010). « On Reacting to Assertions and Polar Questions ». *Journal of Semantics*, 27(1), p. 81–118.

Giannakidou, A. (1998). *Polarity Sensitivity as (Non)Veridical Dependency*. Amsterdam : John Benjamins Publishing Company.

Giannakidou, A. (2008). « Negative and positive polarity items: Variation, licensing, and compositionality ». In C. Maienborn, K. von Heusinger & P. Portner (dir.), *Semantics: An International Handbook of Natural Language Meaning*, Berlin : de Gruyter Mouton, p. 1660-1712.

Krifka, M. (2008). « Basic Notions of Information Structure ». *Acta Linguistica Hungarica*, 55, p. 243–276.

Laka, M. I. (1990). *Negation in Syntax: On the Nature of Functional Categories and Projections*. Thèse de doctorat, Massachusetts Institute of Technology.

Ledgeway, A. (2008). « Satisfying V2 in early Romance: Merge vs. Move ». *Journal of Linguistics*, 44(2), p. 437–470.

Leonetti, M. & Vidal, V. E. (2009). « Fronting and Verum-Focus in Spanish ». In A. Dufter & D. Jacob (dir.), *Focus and Background in Romance Languages*, Amsterdam & Philadelphia : John Benjamins Publishing Company, p. 155–204.

Martins, A. M. (2013). « Emphatic Polarity in European Portuguese and Beyond ». *Lingua*, 128, p. 95–123.

Poletto, C. (2005). « *Sì* and *e* as CP Expletives in Old Italian ». In M. Batllori, M. L. Hernanz, C. Picallo & F. Roca (dir.), *Grammaticalization and Parametric Variation*, Oxford : Oxford University Press, p. 206–235.

Pujol i Campeny, A. (2018). *Word Order in Old Catalan*. Thèse de doctorat, University of Cambridge.

Pujol i Campeny, A. [à paraître]. « E sí la hoïren tots: *Sí* and emphatic positive polarity in Old Catalan ». *Isogloss*.

Rizzi, L. (1997). « The Fine Structure of the Left Periphery ». In L. Haegeman (dir.), *Elements of Grammar*, Dordrecht & London : Kluwer Academic Publishers, p. 281–337.

Roberts, I. (1993). Verbs and Diachronic Syntax. A Comparative History of English and French. Dordrecht : Kluwer Academic Publishers.

Rodríguez Molina, J. (2014). « La Gramática Oculta de la Polaridad Positiva en Español Antiguo ». *Revista de Filología Hispánica*, 30(3), p. 861–915.

Sitaridou, I. (2019). « Word Order in Old Spanish: V2 or Not V2? ». In A. Breitbarth, M. Bouzouita, M. Farasyn & L. Danckaert (dir.), *The Determinants of Diachronic Stability*, Amsterdam : John Benjamins Publishing Company, p. 131-156.

Spevak, O. (2012). « Les enclitiques *enim* et *autem* dans la diachronie du latin ». In A. Christol & O. Spevak (dir.), *Les Évolutions du latin*, Paris : Association Kubaba, p. 335–352.

Wolfe, S. (2015). *Microvariation in Medieval Romance Syntax: A Comparative Study*. Thèse de doctorat, University of Cambridge.

Wolfe, S. (2018). « Probing the syntax of a problematic particle: Old French "si" revisited ». *Transactions of the Philological Society*, 116(3), p. 332–362.

La possession externe : trouve-t-on un emploi plus large de la construction ?

Recherches sur un français classique non littéraire

Iris FABRY
Étudiante, master « Linguistique et dialectologie »,
Université Jean Moulin Lyon 3

Résumé

Cette recherche s'intéresse à la construction syntaxique de possession externe et son évolution en français à travers la question de son extension combinatoire. Pour cela, nous commençons par une description sémantico-syntaxique de la possession externe qui introduit la problématique de notre analyse : l'état de la productivité de la construction au XVIIe siècle. Nous essayons de répondre à cette question au travers d'une analyse de corpus (section 2). Ce corpus est constitué de textes de correspondance et les résultats établis laissent à penser que la possession externe était, à l'époque du français classique, dans un état déjà avancé de figement nominal. Le tout nous amènera à interroger le français contemporain, qui sera plus amplement étudié dans une recherche à venir.

Mots clefs : Possession externe ; productivité ; français classique ; analyse de corpus.

DOI : 10.5281/zenodo.3464464

1. La possession externe

La construction de possesseur externe (dorénavant PE) se distingue de la construction de possesseur interne (dorénavant PI) en tant que le possesseur du nom n'est pas exprimé par le même moyen syntaxique. Dans la construction de PI, exemple (1), le possesseur est marqué dans le même syntagme que celui de l'objet possédé, le *possessum*. En PE, exemple (2), le possesseur se retrouve en dehors du syntagme du *possessum* et est exprimé en tant qu'argument du verbe à l'aide d'un pronom.

(1) [SC… [SD **possesseur**… possessum]] (Scholten, 2018 : 1)
[SC Il a tatoué [SD **son** bras]]

(2) [SC… **possesseur**… [SD possessum]] (Scholten, 2018 : 1)
[SC Il **lui** a tatoué [SC le bras]]

Lors de l'utilisation de la construction de PE, la valeur de possession est encodée syntaxiquement comme un argument du verbe, mais est sémantiquement comprise comme possesseur du nom. Le possesseur se retrouve ainsi dépendant de deux instances différentes : la possession est syntaxiquement liée au verbe, mais, sémantiquement, elle dépend du syntagme déterminatif du *possessum* (Deal, 2013 ; 2017).

De même, la PE exprime une valeur d'inaliénabilité supplémentaire. Lors de son utilisation, il n'y a pas d'ambiguïté sur le fait que l'objet n'est pas disjoint du possesseur. Ainsi, en (2), le possesseur de *bras* sera forcément le possesseur exprimé par le pronom *lui*, alors qu'en (1), le possesseur peut être soit le sujet de

la phrase, soit une personne tierce. L'inaliénabilité est inhérente à la notion de possession externe en français (Guéron, 1983).

Le possesseur de la PE peut être exprimé à travers plusieurs formes en français. Les trois qui seront discutées ici sont les suivantes : le possesseur sujet (3), datif (4), et accusatif — avec un SP *possessum* adjoint (5). Le possesseur est ici marqué en gras et le *possessum* est souligné.

(3) **Les enfants** lèvent la main.

(4) Il **lui** a cassé le bras.

(5) On **l'**a blessé à la jambe.

La PE exprime plusieurs contraintes. Tout d'abord, la PE exprime des contraintes sur le verbe. Ainsi, pour la PE sujet, il est admis que le verbe doit exprimer un geste physique simple (Hatcher, 1944). L'exemple (6) atteste cela puisqu'en (6a), la valeur d'inaliénabilité est toujours comprise, ce qui n'est pas le cas en (6b).

(6) a. Je lève la main
b. *Je lave la main

Pour toutes les PE, on trouve également que le verbe ne peut pas être un verbe d'état. Cela est illustré dans les exemples de (7) à (9) auprès de chaque type de PE (Guéron, 2006).

(7) a. Je lève la main
b. *Je regarde la main

(8) a. Je lui prends/gratte/chatouille la main
b. *Je lui aime/admire/adore la main

(9) a. Je tape Bill sur la main
b. *J'admire Bill sur la main

La possession externe est également soumise à des contraintes sur la modification. En effet, la seule modification acceptée sur le nom du *possessum* est celle à valeur de restriction. Ce phénomène est illustré en (10), où l'adjectif restrictif *gelés* est accepté, et en (11) où l'utilisation d'un adjectif qualificatif n'est possible qu'avec la PI (Guéron, 2006).

(10) Il lui a bandé les doigts gelés.

(11) a. *Le coiffeur a peigné les cheveux soyeux.
b. Le coiffeur a peigné ses cheveux soyeux.

Ensuite, des contraintes peuvent s'appliquer sur le possesseur. En français, des études montrent que le possesseur datif se doit d'être affecté par le procès en cours, et ce à un degré particulièrement élevé par rapport à d'autres langues européennes (Peteghem, 2006). Le possesseur à valeur dative doit « recevoir un rôle sémantique de cible consciente ou d'expériencieur » (Peteghem, 2006 : 364). Ce fait est également celui qui implique les restrictions sur le type de prédicat utilisé. Ainsi, les verbes d'état, n'affectant pas le possesseur, ne peuvent pas fonctionner en PE, exception faite d'un verbe d'état utilisé en dehors des conventions sociales et donc pouvant affecter le possesseur de l'objet (12) (Guéron, 2006 : 617) :

(12) Il lui mate/regarde le cul.

Enfin, la PE exprime des contraintes sur le possessum (Guéron, 2006 ; Kleiber, 1999 ; Peteghem, 2006). Ainsi, en français, il est généralement admis que la PE

s'utilise seulement avec un paradigme restreint constitué des parties du corps, vêtements, objets tenus dans la main ou parties intentionnelles (13-15).

(13) Il lui a tatoué le bras

(14) Il lui a baissé la jupe

(15) Une idée étrange lui traverse l'esprit

Ce paradigme est considéré comme restreint lorsqu'on le compare à d'autres langues comme l'espagnol qui utilise également la PE avec les possédés non humains et les liens de parenté (Barra-Jover, 2002 ; Van de Velde & Lamiroy, 2017) :

(16) A este abrigo se le ve la mala calidad
*A ce manteau, on lui voit la mauvaise qualité

(17) Se les casa la hija mañana.
*La fille se leur marie demain.

Le paradigme de mots utilisés avec les constructions de PE est différent dans chaque langue. Nichols (1988) propose une hiérarchie de la fréquence d'apparition des paradigmes utilisés avec la valeur d'inaliénabilité d'après une étude sur plusieurs langues du monde. Celle-ci propose alors les parties du corps et les liens de parenté comme étant les paradigmes les plus fréquents, suivis par les partie-tout et relations spatiales et enfin des objets possédés selon des principes culturels. Barra-Jover (2002) indique également que l'idée de baser une analyse de la PE sur des traits encyclopédiques n'est pas suffisante à cause de leur forte variation.

Cette dernière contrainte soulève alors la question de la productivité de l'expression en français, ce qui est discuté par Van de Velde et Lamiroy (2017). Dans cet article, les auteurs montrent que la PE est moins productive en français en comparaison de ses langues sœurs, l'espagnol et l'italien. Ceci est démontré par le fait que le paradigme du *possessum* y est le plus restreint et que la construction se retrouve principalement dans des expressions figées telles que (18) et (19) :

(18) La moutarde lui monte au nez

(19) Ça m'a donné froid dans le dos

Pour simplifier ce travail, la perte de productivité de cette construction pourrait être la preuve d'une évolution des langues romanes en direction d'un pivot nominal plus fort, comme l'est l'émergence des déterminants dans ces langues. Le français pourrait alors être la langue la plus « évoluée » de ce groupe.

Ces quelques paragraphes nous ont permis de mettre en évidence le fonctionnement de la PE en français contemporain. Ce dernier point, à propos de la différence d'état de grammaticalisation de la construction en français par rapport aux langues qui lui sont génétiquement liées, nous fait nous interroger sur la période où cette transition a pu se développer. C'est pour cela que nous allons dorénavant nous questionner sur l'état de la productivité de la PE dans le français du XVIIe siècle, période durant laquelle l'accès à la littérature se développe et le français se norme.

2. Analyse

Sur cette problématique, nous partons de quelques exemples littéraires datant du XVIe et XVIIe siècles, tirés de Lamiroy (2003) et Spillebout (1985). Ces phrases ne sont plus acceptées en français contemporain, mais sont attestées dans les écrits de ces siècles.

(20) Le visage leur reluisoit. (Rabelais, *Pantagruel*, XVIe siècle)

(21) Hélas ! Notre pauvre Péronne, il faudra bien la renvoyer si le mal lui continue. (d'après Lamiroy (2003), XVIIe siècle)[1]

(22) La Dauphine ne put tenir plus longtemps les éclats de rire. (Mme de Sévigné, *Lettre*, 1689)

(23) On m'assassine dans le bien, on m'assassine dans l'honneur. (Molière, *L'Avare*, v. 5, 1668)

Ces quelques exemples fondent notre hypothèse initiale et nous incitent à vérifier la productivité de l'expression de PE au sein du français du XVIIe siècle.

2.1. Corpus

Pour répondre à cela, nous nous sommes penchés sur une analyse de corpus. Ce corpus se compose de cinq textes tirés de la base de données Frantext (table 1).

Id. frantext	Auteur	Intitulé frantext	Date	Nombre de mots
Q739	PATIN Guy	Lettres	1630-1649	123 432
Q823	VOITURE Vincent	Lettres	1648	140 625
Q845	MERSENNE le père Marin	Correspondances t.1	1617-1627	6 405
Q846	MERSENNE le père Marin	Correspondances t.2	1628-1630	8 168
Q866	PEIRSCE Nicolas de	Lettres à sa famille	1625	84 593

Table 1. Description du corpus.

Ces textes ont été choisis selon plusieurs critères. Tout d'abord, le texte doit avoir été composé entre 1600 et 1699. Nous avons également choisi de questionner des textes non littéraires issus de correspondances, car, suivant W. Ayres-Bennett (2000), ces textes font partie du groupe de documents les plus informels. Ils peuvent refléter certaines caractéristiques de l'oral puisque la distance entre norme écrite et norme parlée y est plus faible que dans des textes littéraires, sans pour autant effacer les effets de la variation diamésique. C'est donc avec des réserves que l'on peut prendre les textes de correspondance comme représentation de la langue orale. Ceux-ci ne tendent pas à imiter la langue orale, mais peuvent atténuer les différences de cette dernière avec la littérature.

2.2. Résultats

Les résultats actuels sont les suivants : 21 occurrences de PE ont été trouvées dont 16 en référence à une partie du corps et 5 en référence à une partie intentionnelle. La distribution est présentée dans la table 2.

Q739	Q823	Q845	Q846	Q866
5/6	6/7	1/3	1/2	3/3

Table 2. Répartition des occurrences de parties de corps par rapport aux autres types d'occurrences.

[1] Les exemples (20) à (23) sont tirés de plusieurs publications citées plus tôt, les références suivent alors les indications données par ceux-ci.

Les cinq énoncés en référence à une partie intentionnelle sont les suivants :

(24) mais je n'ay peu y donner remède, ne m'ayant déclaré le mal qu'apres qu'il a este fait (Q739)

(25) où vous ne songez qu'à vous délasser l'esprit (Q823)

(26) sans aucune replique qui leur en decouvre la laideur et l'infamie. (Q845)

(27) ce prince, pour qui tout l'univers etoit trop petit, fist plainte au grand Aristote de luy avoir ravi l'unique avantage et la gloire la plus grande (Q845)

(28) et ceux qui luy ont voulu rabattre le courage (Q846)

2.2.1. Occurrences problématiques

Certaines occurrences se sont retrouvées être plus problématiques à analyser et n'ont ainsi pas été comptées parmi les résultats. Nous en citerons ici deux afin d'illustrer les problèmes qui se sont présentés à nous. Ainsi, dans les phrases (29) et (30), nous trouvons une phrase dont la valeur de possession peut être plus difficile à établir.

(29) Mais lors qu'elle eut pris le masque, en mesme temps que les autres le prirent (Q823)

En (29), la phrase exprime-t-elle une PE sujet ou l'usage décrit est-il plutôt culturel ? Dans cette phrase, le masque appartient-il à *elle* et serait alors, comme le décrit Spillebout (1985), un usage de l'article défini avec un possesseur déduit selon le contexte ou bien un usage culturel où le masque qui est pris serait fourni par quelqu'un d'autre pour le bal ?

(30) je vous supplie de ne faire l'adresse directement à nous (Q866)

En (30), la construction de la phrase est ambiguë. L'auteur utilise le SP *à nous*, qui est une construction dative, avec un nom modifié à l'aide d'un article défini. La question se porte alors sur le sens de la phrase : est-ce 'je vous supplie de ne pas vous adresser directement à nous' ou bien 'je vous prie de ne pas utiliser directement notre adresse' ? L'un exprimant une forme verbale et l'autre une possession.

Une réponse n'étant pas directement définissable, nous avons préféré mettre ces exemples de côté dans le cadre de cette analyse.

2.3. Conclusions

En conséquence, selon l'avancée de nos résultats, nous pouvons suggérer que les occurrences de PE outrepassant le paradigme des parties du corps restent marginales. De même, il semble que les occurrences de PE soient marginales par rapport aux occurrences de PI. Au XVIIe siècle, la PI serait alors l'expression de possession la plus utilisée et le français se démarquerait déjà de par son utilisation de la PE par rapport à ses langues sœurs, l'italien et l'espagnol, puisqu'aucune occurrence avec les liens de parenté n'a pu être retrouvée, comme le suggère le travail de Van de Velde et Lamiroy (2017). Suivant leurs termes, nous pourrions alors dire qu'une évolution vers un pivot nominal plus fort est déjà bien avancée à cette période.

3. Pour la suite

Pour la suite de cette recherche, il serait intéressant de se pencher sur le français contemporain familier. La question de la productivité de la PE dans cette forme de la langue peut se poser puisque, dans plusieurs recherches (Leclère, 1995 ; Van de Velde & Lamiroy, 2017), nous retrouvons l'assertion que la phrase *La table, je lui ai astiqué la surface* est agrammaticale en français contemporain. En tant que locuteurs natifs du français, nous avons l'intuition que cette affirmation n'est pas correcte en ce qui concerne le français non standard, familier. Pour étudier cela, nous aimerions alors poser la problématique suivante : la possession externe dative est-elle acceptable avec un paradigme plus large que celui des parties du corps en français moderne non standard ?

Soutenant cette idée, nous avons trouvé plusieurs occurrences de PE élargie lors de recherches internet. En (31), nous retrouvons une co-référence de partie-tout. En (32), une voiture pourrait représenter un objet culturellement possédé. En (33), cet exemple montre que ce type de construction peut se retrouver dans des contextes plus officiels puisque celui-ci a été retrouvé dans un article journalistique en ligne.

(31) Depuis je lui [mon pistolet] ai poli les flancs pour l'adoucir.
(message de forum)

(32) Un retardataire qui l'a percuté et lui [il] a cassé la voiture.
(message de forum)

(33) 10h30, Macron, resté sur le perron de l'Élysée avec Brigitte, va voir partir Hollande, tel le gérant d'une maison d'hôte qui regarde se tailler le client qui lui [le gérant] a ruiné le buffet de chouquettes.
(rtl.fr, *Tanguy Pastureau*, le 14/05/2017)

Bibliographie

Ayres-Bennett, W. (2000). « Voices from the Past: Sources of Seventeenth-Century Spoken French ». *Romanische Forschungen*, 112, p. 323-348.

Barra-Jover, M. (2002). « Datif, possessif, article défini, ou comment se passer de la possession inaliénable ». *Recherches linguistiques de Vincennes*, 31(1), p. 43-60.

Deal, A. R. (2013). « Possessor Raising ». *Linguistic Inquiry*, 44(3), p. 391-432.

Deal, A. R. (2017). « External Possession and Possessor Raising ». In M. Everaert & H. C. van Riemsdijk (dir.), *The Wiley Blackwell Companion to Syntax, Second Edition*, Hoboken, NJ : John Wiley & Sons, p. 1509-1540.

Guéron, J. (1983). « L'emploi « possessif » de l'article défini en français ». *Langue française*, 58(1), p. 23-35.

Guéron, J. (2006). « Inalienable Possession, Primarily in French ». In M. Everaert & H. C. van Riemsdijk (dir.), *The Wiley Blackwell Companion to Syntax, Second Edition*, Hoboken, NJ : John Wiley & Sons, p. 589-638.

Hatcher, A. G. (1944). « Il tend les mains *vs*. Il tend ses mains ». *Studies in Philology*, 41(3), p. 457-481.

Kleiber, G. (1999). « Anaphore associative et relation partie-tout : Condition d'aliénation et principe de congruence ontologique ». *Langue française*, 122(1), p. 70-100.

Lamiroy, B. (2003). « Grammaticalization and external possessor structures in Romance and Germanic languages ». In M. Coene & Y. d'Hulst (dir.), *From NP to DP,* vol. 2 (*The Expression of possession in noun phrases*), Amsterdam & Philadelphia : J. Benjamins Pub., p. 257-280.

Leclère, C. (1995). « Sur une restructuration dative ». *Language Research*, 31(1), p. 179-198.

Nichols, J. (1988). « On alienable and inalienable possession ». In W. Shipley (dir.), *In honor of Mary Haas: From the Haas Festival Conference On Native American Linguistics*, Berlin & New York : Mouton de Gruyter, p. 557-610.

Peteghem, M. V. (2006). *Inaliénabilité et partitivité : Le datif possessif en français, en néerlandais et en roumain*. Paris : Peeters.

Scholten, J. (2018). *The ins and outs of external possession: A micro-comparative perspective*. Utrecht : LOT.

Spillebout, G. (1985). *Grammaire de la langue française du XVIIe siècle*. Paris, France : Picard.

Van de Velde, F. & Lamiroy, B. (2017). « External possessors in West Germanic and Romance: Differential speed in the drift towards NP configurationality ». In D. van Olmen, H. Cuyckens & L. Ghesquière (éd.), *Aspects of grammaticalization: (inter)subjectification and directionality*, Berlin & Boston : Mouton de Gruyter, p. 353-400.

Entre ancien et moyen français : le traitement de l'*Ovide moralisé*

Prunelle DELEVILLE
Université Lumière Lyon 2 (CIHAM UMR 5648)
Université de Genève (FNRS)

Résumé

L'*Ovide moralisé* en vers, dont le témoin le plus ancien date environ de 1315-1320, se situe au tournant de l'ancien et du moyen français. Au sein de cette tradition manuscrite, nous nous proposons d'étudier une famille de représentants plutôt tardive (toute fin du XIVe siècle, première moitié du XVe siècle). Cette rédaction porte les traces d'un état textuel ambivalent : elle rend compte de traits pourtant désuets en même temps qu'elle témoigne de l'incompréhension de certains mots ou de la déliquescence de certaines formes. Nous nous proposons alors de nous concentrer sur les innovations sémantiques et métriques que ce groupe de copies apporte à l'histoire de la langue ancienne.

Mots-clefs : Tradition textuelle ; innovation sémantique ; métrique.
DOI : 10.5281/zenodo.3464468

Introduction

L'*Ovide moralisé* constitue la première traduction en français des *Métamorphoses* d'Ovide. Elle voit le jour au début du XIVe siècle. Son auteur anonyme fait suivre le récit de chaque métamorphose d'une série d'interprétations allégoriques concrètes (physiques, historiques), morales et spirituelles. La tradition manuscrite de *l'Ovide moralisé* se compose d'une vingtaine de manuscrits et s'étend sur plus d'un siècle (Baker et al., 2018). Le premier témoin que nous ayons conservé date de 1315-1320 et le dernier de 1480. La copie la plus ancienne, nommée *A1* dans le stemma, est probablement la plus proche du sub-archétype. En raison de sa date d'écriture, ce texte se situe au tournant de l'ancien et du moyen français, et rend compte de deux états de la langue. Cette particularité linguistique a déjà été étudiée par Y. Greub et T. Städler (2018). Ils ont aussi montré combien le vocabulaire de cette première version est riche et parfois très technique[1]. Nous ne nous attacherons donc pas particulièrement à ce moment dans l'histoire de l'*Ovide moralisé*, puisqu'il a déjà fait l'objet d'un examen approfondi. Nous souhaitons nous placer plus en aval en examinant une famille tardive, appelée *Z* dans le stemma, et qui a la particularité d'être une réécriture du texte. Elle se compose de quatre témoins : Berne, Burgerbibliothek, 10 (*Z1*) écrit après 1456 ; Paris, BnF, français 374 (*Z2*) copié en 1456 ; Paris, BnF, français 870 (*Z3*) composé autour de 1400 pour le texte et 1450 pour le décor ; Paris, BnF, français 19121 (*Z4*) probablement réalisé entre 1390 et 1410 (Besseyre et Rouchon, 2018 : 13-14).

Le vocabulaire de ces copies nous intéresse pour deux raisons : il témoigne de la réécriture et des modifications qui en découlent, mais aussi du siècle qui s'est écoulé depuis la première version du texte. La réécriture est un acte de reprise d'un texte déjà existant, mais elle se distingue de la répétition pure et de la copie. Elle

[1] Il emprunte à des domaines comme la musique, l'astrologie, l'art de tisser ou celui d'interpréter les textes.

implique reproduction, mais aussi différenciation par rapport au texte premier. D'un point de vue diachronique, le texte de *Z* est comme celui du témoin *A1* tributaire de l'ancien et du moyen français, mais sa période de rédaction le fait tendre complètement vers le moyen français. Cette rédaction porte ainsi les traces d'un état textuel ambivalent : elle rend compte de traits désuets en même temps qu'elle témoigne de l'incompréhension de mots ou formes tombés en désuétude.

La distance temporelle qui sépare cette version du texte de la première introduit une difficulté supplémentaire, celle de l'analyse de graphies qui témoignent de la présence d'une ancienne diérèse. Les hiatus internes de l'ancien français tendent à se résorber à partir du XIVe siècle. Les premiers exemplaires de l'*Ovide moralisé*, datant du début du XIVe siècle, possèdent des formes graphiques qui prennent en compte ces hiatus. Ces formes sont tantôt recopiées dans *Z* et tantôt changées pour faire disparaître la marque d'un hiatus qui n'est plus prononcé. On oscille ainsi entre deux paradigmes, ce qui engendre une instabilité métrique. Les témoins *Z*, en raison du siècle qui les sépare de la version « originale », permettent aussi de mesurer la distance en même temps que la proximité linguistique entre le début du XIVe siècle et sa toute fin, voire la première moitié du XVe siècle. C'est ainsi que nous nous proposons d'interroger l'attitude du copiste face à une difficulté de sens ou de versification et les innovations qu'il propose.

1. Réactions face à des formes surannées ou rares

1.1. Traitement global dans *Z*

L'auteur de l'*Ovide moralisé* recourt souvent à l'adverbe « seviaus », qui signifie « au moins ». Comme l'indique Wartburg (1922-1967 : XIV 215b), l'adverbe s'emploie en ancien français, puis disparaît de la langue. Les attestations des autres dictionnaires vont aussi dans ce sens ; le terme est même absent du DMF (http://www.atilf.fr/dmf, consulté le 31 mai 2019). Il serait donc déjà archaïque à l'époque de rédaction du premier *Ovide moralisé*. Nous remarquons aussi qu'il ne figure plus du tout dans les copies *Z*, alors qu'on le retrouve encore dans d'autres, même si certains témoins tardifs attestent de difficultés dans la compréhension ou la transmission du mot. Le terme pose déjà problème plus en amont, au niveau de la branche la plus proche de *Z*, la branche *Y*, dont les représentants s'échelonnent du troisième quart du XIVe siècle au tout début du XVe siècle. Tel est notamment le cas au cinquième livre. Cérès apprend que sa fille a été enlevée par le roi des Enfers et prie Jupiter de délivrer son enfant. Dans les témoins du début du XIVe siècle, Cérès dit à Jupiter :

> Dois tu estre esmeüs seviaus,
> Se pour moi faire ne le viaus (De Boer, 1915-1938 : V, v. 2212-2213).

L'adverbe n'est pas rendu dans *Y* :

> Dois tu esmeüs estre et viaux
> Se pour moi faire ne le viaus.

Le passage est totalement réécrit dans *Z*. Le remanieur de cette version évacue tout problème de compréhension avec une variante très claire :

> Et se pour moy ne le veus faire,
> Fais le pour ton enffant rettraire (Deleville : V, v. 1223-1224).

D'autres témoins tardifs, apparentés ou non à *Y* et *Z*, attestent eux aussi d'une incompréhension de l'adverbe dans un autre passage encore. Leurs copistes

changent le mot pour un autre de sens différent ou donnent à lire des formes encore graphiquement proches de « seviaus », mais aberrantes[2]. Contrairement à ces copistes qui œuvrent plus en amont dans la tradition et qui tentent de remplacer le mot par un autre, le rédacteur de *Z* n'en tient pas compte et n'essaie pas d'en rendre le sens global qu'il semble ne pas comprendre. Pour sa première occurrence, le remanieur de la famille *Z* l'écarte tout simplement, quitte à rendre le vers hypométrique. On passe de « Voiz seviaus tout ton ciel nerci » (De Boer, 1915-1938 : II, v. 567), qui se conserve dans toute la tradition, à « Voy com le ciel est noirci ». Pour la deuxième occurrence du mot dans le vers de la tradition « Seviaux, se plus n'en puis porter » (De Boer, 1915-1938 : IV, v. 3503), le réviseur ajoute quelques termes pour pallier la suppression de « seviaux », mais le vers en devient hypermétrique : « Tel mal ay que ne le puis porter » (Deleville : IV, v. 2415). Le traitement de l'adverbe « seviaus » illustre ainsi la façon dont l'auteur original reprend des termes parfois surannés ou rares et la façon dont ils ont été transmis ou non.

L'auteur de la version « originale » utilise quelques fois le verbe « essabouir » sous la forme de participe passé « essaboï » : « interdit, hébété » (Godefroy, 1881-1902). Les exemples des dictionnaires (Godefroy : III 563 ; Wartburg, 1922-1967 : XXI 328a ; Tobler et Lommatzsch, 1925-1974 : III-2 1283 ; Matsumura, 2015 : 1411b-1412a) sont assez rares et concernent surtout le *Roman de la Rose* puis des exemples uniques tirés du *Roman de Thèbes*, de l'*Ovide moralisé* et de la *Bible de Macé de la Charité* (texte de *ca* 1300). Nous pouvons donc penser que le terme est déjà rare en ancien français ou est sorti d'usage à cette époque. Dans la famille *Z*, l'adjectif est en effet systématiquement remplacé par un quasi synonyme : « esbahi ».

1.2. Une attitude spécifique : le cas de Z^2

Le témoin Z^2 offre la particularité d'avoir été corrigé, après sa rédaction, par une main médiévale qui n'est pas celle du copiste. Il pourrait s'agir de la main du possesseur du manuscrit, Louis du Périer, ou plutôt d'un autre lecteur. Nous savons bien que son intervention ne nous apprend rien de la tradition manuscrite de l'*Ovide moralisé*, mais il nous renseigne néanmoins sur l'évolution de la langue et sur la perception qu'un lecteur du XVe siècle peut avoir de certains traits langagiers hérités d'une version plus ancienne. Ce lecteur assidu ne cesse de biffer des éléments qui ont pu lui paraître surannés. Par exemple, il rature systématiquement l'imparfait[3] « yert » pour le remplacer par le passé simple « fut ». Il souhaite aussi se débarrasser de la forme étymologique du futur « yert », qui a elle aussi disparu au XVe siècle. Il la biffe pour écrire « sera ». La troisième personne du singulier du verbe « estovoir » (« estuet ») est également souvent raturée et supplantée par « faut », ou par « convient ». Le verbe « estovoir » ne

[2] Au livre VI, dans le vers « Qu'ele a seviaus grant esperance » (éd. C. De Boer, VI, v. 3426), l'adverbe « seviaus » est remplacé par *moult* dans huit témoins apparentés et compris entre la fin du XIVe siècle et la fin du XVe siècle (A^2YZ). Pour ce même vers, on lit en outre, dans deux *codices* de la famille *G* (XVe siècle), une version aberrante ou éloignée du texte : dans G^1 « sur yaulx », dans G^3 « en soy ». L'adverbe est également mal transmis dans deux manuscrits de la fin du XIVe siècle et un troisième du début du XVe siècle : « se veulz » (D^4E^1) ou « se vault » (E^2).

[3] C. Buridant (2000 : 276) rappelle que la progression vers l'élimination totale de l'imparfait et du futur « ert » est variable. « Dans les textes en vers, du XIIe au XIVe siècle, s'observe une lente progression vers la disparition totale », peut-être due à l'utilité de recourir à deux paradigmes pour faire varier le mètre ou la rime.

disparaît pas en moyen français ni au XVe siècle. On le retrouve notamment chez Christine de Pizan (*DMF*, www.atilf.fr/dmf, consulté le 31 mai 2019). En revanche, on ne le compte plus parmi les mots de la langue du XVIe siècle (Huguet, 1925-1973). Nous pensons donc que le lecteur interventionniste de la copie Z^2 a pu le considérer comme suranné.

2. Créations lexicales

2.1. Première ou unique attestation

Le rédacteur à l'origine de la famille *Z* crée de nouvelles formes ou propose des emplois qu'on ne trouve pas dans les dictionnaires ou seulement pour des exemples plus tardifs. Nous ne saurions dire si certains de ces emplois constituent vraiment la première attestation d'un mot. Nous en proposons tout de même un : l'adjectif « aficheux », commun aux témoins Z^3, Z^4 et à ceux de la famille *Y*. Cet adjectif, qui signifie « obstiné », apparaît seulement dans le dictionnaire de Wartburg (1922-1967 : III 509a), pour un ouvrage plus tardif de François de Sales (XVIe siècle). Notre texte pourrait donc en offrir la première occurrence.

En outre, dans les passages que développe le rédacteur de la famille *Z* nous trouvons parfois des mots absents des dictionnaires. On pense notamment à l'extrait de la fable de Mars et Vénus. Vulcain surprend en plein acte sexuel son épouse avec le dieu Mars. Le mari révèle cette scène d'infidélité aux autres dieux qui ne manquent pas de rire de cette révélation. Dans la version du début du XIVe siècle, nous lisons :

> Li damedieu tiennent grant fable,
> Grant bourde et grant murmurement
> De Venus et de son amant. (De Boer, 1915-1938 : IV, v. 1321-1337)

Le passage est amplifié dans *Z* :

> Li dieu demainent grant gaudisse,
> Grant feste, grant bourde, grant risse,
> Grant parler, grant murmurement
> De Venus et de son amant. (Deleville : IV, v. 1060-1063)

Le terme « gaudisse », utilisé par le remanieur pour développer le thème de la moquerie, n'apparaît pas dans les dictionnaires. On y relève seulement les substantifs « gaudisserie » au sens de « réjouissance, moquerie » (Godefroy, 1881-1902 ; Wartburg, 1922-1967 ; DMF, www.atilf.fr/dmf, consulté le 31 mai 2019 ; Matsumura, 2015), « goderie » ou « gaudie » (Wartburg, 1922-1967 ; DEAF : G 400). La forme a pu être influencée par l'existence de variantes du moyen français comme « gaudissant », « gaudisseur » (Wartburg, 1922-1967 : IV 78b). D'autre part, le vocable « risse » ne fait pas non plus partie des dictionnaires ; il a pu être été appelé par la rime. La forme « risse » se retrouve uniquement dans des variantes dialectales modernes du nom « ris », conservé jusqu'en français moderne : « risse » à Tournai, « rise » au sens de « sujet de rire » en franc-comtois, « rises » au pluriel « éclats de rire » en Suisse (Wartburg, 1922-1967 : X 419b).

Le rédacteur de *Z* innove aussi dans l'un des lexiques clef de notre texte : le vocabulaire de la métamorphose. Au moment du combat entre les Lapithes et les Centaures, le narrateur de la version la plus répandue évoque la force de Cénée. Le réviseur de *Z* complète l'évocation en rappelant la métamorphose du personnage :

> Bien approva son vaisselage
> Celui qui feme avoit esté
> Que Neptnus ot translaté. (Deleville : XII, v. 2014-2016)

Dans les exemples des dictionnaires, le verbe « translater » ne s'applique pas au phénomène de métamorphose. Il renvoie au mouvement d'un endroit vers un autre endroit, ou plus métaphoriquement au passage d'une langue vers une autre, mais non d'un corps vers un autre corps. Le remanieur emploie également deux fois le substantif « muage ». On le retrouve dans la fable de Céyx et Alcyon. Céyx et son équipage partent en mer, mais font naufrage. Junon souhaite faire connaître le funeste accident à Alcyon, l'épouse de Céyx. La déesse charge Morphée d'apparaître en songe à Alcyon sous les traits de son mari :

> En tel semblance, en tel muage
> A fait Morpheüs son message,
> Se qu'il sembloit appertement
> Que ce fust il vissablement
> Qui mort yere et en mer peris
> Et soulloit estre ses maris. (Deleville : XI, v. 2522-2527)

L'autre emploi du nom « muage » s'applique à la faculté de Périclymène de changer de forme quand bon lui semble :

> Celui ot ·i· bon aventage :
> Il se muoit en tel muage
> Qu'il vouloit, et quant bon li ere
> Reprenoit sa fourme premiere. (Deleville : XII, v. 2157-2160)

Dans les deux cas, le substantif renvoie bien au produit de la métamorphose, comme l'indique sa coordination avec « semblance » ou la référence à la « fourme premiere » du personnage. Pourtant, le vocable « muage » est seulement attesté, depuis la moitié du XIVe siècle, en droit pour désigner la redevance due lors d'un changement de propriétaire. Godefroy (1881-1902), que reprend Wartburg (1922-1967), relève un terme de fauconnerie tiré d'un texte du XVIIe siècle. Le « muage » désigne l'action de muer d'un oiseau. Il est donc probable que notre attestation soit la première au sens de « changement de corps, métamorphose ».

3. Innovations métriques

3.1. Réactions devant une morphologie d'ancien français

Nous évoquions en introduction la difficulté que soulève l'analyse de graphies marquant la présence d'une ancienne diérèse. Les hiatus internes caractéristiques de l'ancien français sont révolus à la période de rédaction de la branche *Z*. Pourtant, en tant que reprise d'un texte qui maintient souvent ces hiatus, on retrouve encore dans *Z* ces formes anciennes. Sur ce point, les manuscrits *Z* affichent un état qu'on pourrait qualifier de schizophrénique. Par exemple, dans un même couplet, un participe passé, qui se prononce avec diérèse, rime avec un autre où elle ne se prononce plus :

> Quant ot Philomena veü
> Par maintes fois et apperceu (Deleville : VI, v. 2223-2224)

Le second vers marquait pourtant bien une diérèse dans le manuscrit A^1 : « Maintes fois et apperceü ».

Cet état est comme exacerbé dans le sous-ensemble plus récent de ce groupe, Z^1 et Z^2, tous deux composés dans la seconde moitié du XVe siècle. Le rédacteur du modèle de cet ensemble ajoute fréquemment un monosyllabe, alors que le vers présente un ancien hiatus. Par exemple le rédacteur du modèle de Z^{12} ajoute au vers « Plusieurs s'en sont apperceüs » (De Boer, 1915-1938 : II, v. 2456 et Z^{34}, II, v. 1798), transmis dans toute la tradition, le monosyllabe « si » : « Plusieurs si s'en sont aperceus ».

Le correcteur de Z^2 n'hésite pas non plus à ajouter un monosyllabique qu'il suscrit, car il ne considère plus le hiatus qu'envisageait l'auteur « original ». Ainsi, un vers tel que « Les Musses furent esleües » devient « Les musses furent donc esleues » (Deleville : V, v. 818).

Enfin, un autre trait morphologique caractéristique, qui unit Z^3 et Z^4, atteste d'un traitement spécifique de l'octosyllabe. On relève la généralisation de la forme « celui » pour le cas où l'on trouve dans la tradition la forme « cil » du démonstratif. Ce changement ajoute apparemment une syllabe au vers et le rend irrégulier. Pourtant, « celui » est presque systématiquement écrit à la place de « cil » sans refonte du vers, comme dans les vers suivants : « En celui temps ou tout bien habonde » (Z^{34}, I, v. 220), « Celuy controuva les folles lois, / Celui rois se faisoit hounorer » (Z^{34}, I, v. 223-224)...

3.2. Innovations lexicales motivées par la métrique

Lorsque le remanieur remplace, au début du vers, un terme par un autre qui contient une syllabe en plus, il lui arrive d'abréger un autre mot du vers[4]. Nous pensons notamment à un passage à propos de Saturne. Dans la version « originale », Saturne est présenté de la façon suivante :

> Cil controuva, par sa mestrie,
> Premerains l'art de fabrerie (De Boer, 1915-1938 : I, v. 713-714)

Le deuxième vers subit quelques modifications dans Z :

> Celui controuva par sa maistrie
> Premierement l'art de fabrie (Deleville : I, v. 411-412)

Le passage de « premerain » à « premierement » nous semble être à l'origine du changement de « fabrerie » en « fabrie ». Le dernier terme a en effet le mérite de corriger l'hypermétrie. En revanche, rien ne l'atteste dans la langue française. On peut supposer qu'il est influencé par le verbe « fabrier » (« travailler le fer »), mais cette variante pour « favrier », qui est déjà rare, ne se rencontre que postérieurement chez Colard Mansion[5] (1493). Un autre vocable, peu attesté, est lui aussi abrégé pour des raisons qui nous paraissent métriques. Callisto, violée par Jupiter, tombe enceinte et craint que Diane ne s'en aperçoive. L'auteur original en profite pour glisser le commentaire suivant :

> Tous jours se crient li homs mesfais
> Dou fet dont il se sent coupable
> Qu'il soit a tous aparissable (De Boer, 1915-1938 : II, v. 1520-1522)

[4] Le remanieur propose ce type de changement quand l'autre mot qu'il abrège est difficile à identifier, rare. Les termes « fabrerie » et « apparissable », qui ont fait l'objet des modifications que nous traitons, posent en effet problème dans un large pan de la tradition.
[5] Le passage correspond à celui de notre texte et que reprend ici Colard Mansion : « Vulcan trouva premierement l'art de fabrier et de forger » (www.atilf.fr/dmf, consulté le 31 mai 2019).

Le dernier vers diffère dans *Z* : « Qu'il ne soit a touz apparable » (Deleville : II, v. 941). L'adjectif « aparissable » figure dans tous les autres témoins et se trouvait probablement aussi dans le modèle de *Z*. Le remanieur aurait changé le mot après avoir ajouté la particule de négation, ici explétive, qui rend le vers hypermétrique. Le DMF (www.atilf.fr/dmf, consulté le 31 mai 2019) et Godefroy (1881-1902) distinguent deux entrées pour « apparable ». Aucune des deux ne convient au contexte : sens de « comparable » (Wartburg, 1922-1967 : VII, 598a ; Tobler et Lommatzsch, 1925-1974) et sens de « clair, brillant » (Godefroy, 1881-1902), qui s'applique à un texte tardif (1413), et qui n'apparaît pas chez Wartburg (1922-1967).

Conclusion

La distance temporelle de près d'un siècle qui sépare l'*Ovide moralisé* « original » de la réécriture qu'offre la famille *Z* fait de ce groupe un bon exemple de l'évolution de la langue médiévale. Elle rend également compte de la capacité d'innovation du remanieur qui propose parfois ce qui nous semble être un nouveau vocable. Ses interventions nous renseignent sur la permanence ou au contraire sur l'obsolescence d'une forme, d'un terme et de sa signification. Le traitement dans les copies *Z* des formes qui marquent un ancien hiatus est également instructif. Si ces derniers ne se prononcent plus dans la langue que pratiquent les copistes des témoins *Z*, une ambivalence demeure liée à la reprise d'un texte qui les contenait.

Annexe : liste des témoins

Sigle	Cote	Datation approximative[6]
A[1]	Rouen, Bibliothèque municipale, ms. O. 4	*ca* 1325 (f. 3-15) et *ca* 1315-1320
A[2]	Rouen, Bibliothèque municipale, ms. O. 11bis	troisième tiers du XVe siècle (*ca* 1475 ?)
B	Lyon, Bibliothèque municipale, ms. 742	*ca* 1390
D[1]	Bruxelles, Bibliothèque royale de Belgique, ms. 9639	*ca* 1430-1440
D[2]	Cambrai, Bibliothèque municipale, ms. 973	troisième tiers du XVe siècle (*ca* 1470 ?)
D[3]	Paris, Bibliothèque nationale de France, fr. 24306	troisième quart du XIVe siècle
D[4]	New York, Pierpont Morgan Library, M. 443	*ca* 1410
D[5]	Paris, Bibliothèque nationale de France, fr. 24305	1356
E[1]	Genève, Bibliothèque de Genève, ms. fr. 176	*ca* 1380
E[2]	Vatican, Biblioteca Apostolica Vaticana, Reg. Lat. 1480	*ca* 1380
G[1]	Paris, Bibliothèque nationale de France, fr. 373	*ca* 1400
G[2]	Paris, Bibliothèque de l'Arsenal, ms. 5069	*ca* 1330-1335
G[3]	Copenhague, Kongelige Bibliotek, ms. Thott 399	*ca* 1480
Y[1]	Paris, Bibliothèque nationale de France, fr. 871	troisième tiers du XIVe siècle (*ca* 1380 ?)
Y[2]	Paris, Bibliothèque nationale de France, fr. 872	troisième tiers du XIVe siècle (*ca* 1370-1380 ?)
Y[3]	Londres, British Library, Add. MS 10324	*ca* 1400
Z[1]	Berne, Burgerbibliothek, ms. 10	après 1456
Z[2]	Paris, Bibliothèque nationale de France, fr. 374	1456
Z[3]	Paris, Bibliothèque nationale de France, fr. 870	*ca* 1400, décor *ca* 1450
Z[4]	Paris, Bibliothèque nationale de France, fr. 19121	*ca* 1390-1410 ?

[6] Selon Besseyre & Rouchon (2018) et Jung (1996).

Bibliographie

Baker, C., Besseyre, M., Cavagna, M., Cerrito, S., Collet, O., Gaggero, M., Greub, Y., Guillaumin, J.-B., Possamaï-Pérez, M., Rouchon Mouilleron, V., Salvo García, I., Städtler, T. & Trachsler, R. (2018). *Ovide Moralisé*, Livre I. Paris : Société des anciens textes français.

Besseyre, M. & Rouchon-Mouilleron, V. (2018). « Description des manuscrits ». In C. Baker et al., *Ovide Moralisé*, Livre I, Paris : Société des anciens textes français, p. 16-89.

Buridant, C. (2000). *Grammaire nouvelle de l'ancien français*. Paris : SEDES.

Cavagna, M., Gaggero, M. & Greub, Y. (2014). « Prolégomènes à une nouvelle édition de l'*Ovide moralisé* ». *Romania*, 132, p. 176-213.

De Boer, C. (1915-1938). *Ovide moralisé, poème du commencement du quatorzième siècle, publié d'après tous les manuscrits connus.* Amsterdam : Müller.

Deleville, P. (2019). *Métamorphose des Métamorphoses : édition critique et étude littéraire des manuscrits Z de l'*Ovide moralisé. Thèse de doctorat, Université Lyon 2 Lumière et Université de Genève.

Dictionnaire étymologique de l'ancien français (DEAF) (1974...). Tübingen, Québec, & Paris : Niemeyer, Presses de l'Université de Laval & Klincksieck ; continuation Berlin : De Gruyter.

Godefroy, F. (1881-1902). *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IXe au XVe siècle*, Paris : Vieweg-Bouillon.

Greub, Y. & Städler, T. (2018). « Le lexique de l'*Ovide moralisé* ». *Ovide Moralisé*, Livre I, Paris : Société des anciens textes français, p. 169-181.

Huguet, E. (1925-1973). *Dictionnaire de la langue française du seizième siècle*. Paris : Champion.

Matsumura, T. (2015). *Dictionnaire du français médiéval*, Paris : Les Belles Lettres.

Jung, M.-R., (1996). *La légende de Troie en France au Moyen Âge*. Tübingen & Basel : Francke.

Tobler, A. & Lommatzsch, E. (1925-1974). *Tobler-Lommatzsch Altfranzösisches Wörterbuch*. Berlin : Weidmannsche Buchhandlung.

Wartburg, W. von. (1922-1967). *Französisches Etymologisches Wörterbuch, eine Darstellung des galloromanischen Sprachschatzes*. Basel : Zbinden.

Variation textuelle et variation linguistique dans les textes occitans d'Antoine Clet

Vincent SURREL
Université Paris 8 (SFL UMR 7023)
École nationale des chartes

Résumé

Nous proposons d'analyser les phénomènes de corrélation entre variation textuelle et variation linguistique dans la pièce d'Antoine Clet intitulée *Le Sermon manqué* (*ca* 1750), connue par cinq témoins. La confrontation des données diasystématiques fournies par la critique philologique avec les données linguistiques, principalement au niveau grapho-phonologique, doit nous permettre d'interpréter les comportements sociolinguistiques des scripteurs.

Mots-clés : nord-occitan (vellave) ; dialectologie historique ; géolinguistique ; graphématique ; philologie linguistique.

DOI : 10.5281/zenodo.3464471

Introduction

La philologie et la linguistique ont trouvé, depuis la fondation même des études romanes, un point de rencontre privilégié dans la recherche de corrélations entre *variation textuelle* et *variation linguistique*. Ces deux niveaux de variation, en principe intrinsèquement liés, sont recherchés d'une part dans les textes littéraires — les textes documentaires étant a priori peu ou pas sujets à variation —, et, d'autre part, dans les textes médiévaux. On estime que les transformations opérées à l'aube de l'Époque moderne en ce qui concerne des conditions de transmission des textes, mais aussi le statut social et auctorial des producteurs limitent la possibilité de telles recherches pour les périodes postmédiévales. Le principe d'un fixisme textuel et linguistique des textes littéraires modernes et contemporains a cependant été battu en brèche : si depuis plusieurs décennies la critique génétique interroge de façon radicale la nature des textes littéraires, il faut constater par ailleurs que l'absence de pensée philologique dans certains champs de la recherche littéraire n'a pas souvent permis de développer une connaissance correcte des traditions textuelles[1], laissant de larges possibilités d'exploration aux linguistes et aux philologues.

L'élaboration du corpus des textes occitans de l'ancien Velay (auj. département de la Haute-Loire) nous donne l'occasion d'interroger les phénomènes de variation dans les pièces bilingues franco-occitanes d'Antoine Clet (Le Puy-en-Velay, 1705 — Le Puy-en-Velay, 1785) pour lesquels nous avons proposé une première interprétation en termes d'histoire des textes (2016). Nous examinerons ici seulement *Le Sermon manqué*[2], dont la tradition textuelle est constituée de cinq

[1] Pour le XVIIe siècle, voir Duval (2015).

[2] Le texte est désigné par l'abréviation *CletSM* dans notre corpus. *Le Sermon manqué* est une pièce de circonstance faisant allusion à des personnes réelles du Puy. Elle peut être résumée de la façon suivante : le chanoine Vivier, du chapitre cathédral du Puy, doit faire l'éloge

témoins manuscrits et imprimés, produits sur près d'un siècle, et qui présentent un haut degré de variation textuelle et linguistique dans les parties occitanes. En confrontant les données diasystématiques obtenues par la critique philologique avec les phénomènes de variation grapho-phonologique les plus saillants, nous souhaitons proposer une interprétation des systèmes graphiques de ces textes modernes comme un lieu de tension entre contraintes dialectales et marquages sociolinguistiques, et ainsi affiner le processus de 'patoisisation' de l'écrit occitan en cours depuis le XVIe siècle[3].

1. *Le Sermon manqué* : présentation de la tradition textuelle

1.1. Caractéristiques génériques et sociopoétiques

Les textes d'Antoine Clet peuvent être interprétés d'un point de vue générique comme le point de rencontre entre deux traditions littéraires vivaces au Puy, le théâtre et le portrait satirique, avec un infléchissement de la matière narrative dans le sens indiqué par des formes littéraires largement diffusées comme la dancourade. Si les pièces de Clet ont été comprises dès l'époque romantique comme le signe d'une contestation du pouvoir, dénotant une pensée en lien avec les Lumières et qui préfigurerait la Révolution française, elles témoignent davantage d'une prise de distance de l'élite ponote avec l'idéologie municipale d'essence médiévale. Elles ne s'inscrivent pas dans une culture de salon et on ne peut aucunement présumer d'un succès populaire. La mise au jour d'une tradition textuelle, surtout manuscrite, atteste en tout cas de formes de sociabilité encore mal connues, qui impliquent des cercles restreints de l'élite municipale, au sein desquels les pièces de Clet ont circulé et ont été réécrites[4], et dont le rôle pouvait être d'entretenir les rapports de connivence.

1.2. Données diasystématiques

Il ne s'agira pas ici de reprendre les détails de l'analyse philologique, et il nous suffit de présenter dans la table ci-contre les données diasystématiques disponibles, qui ressortissent de critères internes et externes, pour assigner un objectif à l'étude linguistique.

du patron des augustines de Vals dans leur propre couvent. Sa mère le lui déconseille, car son précédent prêche a été un échec retentissant. Il se rend tout de même chez les moniales qui l'accueillent avec beaucoup de considération. Il revoit son sermon avant de monter en chaire et se flatte d'être un prêcheur hors pair, mais, soudainement pris de panique, il s'enfuit. Envoyé à sa recherche, le clerc du couvent apprend d'un paysan qu'en fuyant il a renversé les murets des paysans. L'un d'entre eux vient s'en plaindre auprès de la supérieure, puis auprès de la mère du chanoine, qui le renvoient tour à tour. Réfugié chez sa fermière, Vivier cherche à savoir ce qu'on dit de lui en ville, puis finit, plusieurs mois après, par revenir auprès de sa mère. Le syndic du chapitre lui annonce à ce moment qu'il n'a plus le droit de monter en chaire.

[3] Pour l'occitan, nous renvoyons à la synthèse de Martel (2013).

[4] Sur les phénomènes de littérature manuscrite à l'Époque moderne, voir Weil (1993).

Témoins	**Manuscrits**				**Imprimé**
	A	*B*	*C*	*D*	*E*
Lieux de conservation	Le Puy-en-Velay, Bib. municipale, fonds Cortial, ms. 147	Le Puy-en-Velay, Bib. municipale, fonds Cortial, ms. 166	Le Puy-en-Velay, Archives dptales de la Haute-Loire, fonds Pomarat, 149 J 885	Le Puy-en-Velay, Bib. municipale, fonds local, ms. 26	Exemplaire consulté : Le Puy-en-Velay, Bib. municipale, fonds Cortial, n° 322
Titre	Sermon manqué. Le 28 aoust 1749. Comedie en trois actes	Sermon manqué. Le 22e juillet 1798. Comédie en 3 actes	Sermon manqué. Comedie en (3) actes	Le sermon manqué. Le 28 août 1749	Sermon manqué, comédie en trois actes, en vers, par Antoine CLET, poète vellavien au 17me (sic) siècle.
Espace	Velay central (Le Puy-en-Velay ?)	Velay central (Le Puy-en-Velay ?)	Velay central (Le Puy-en-Velay ?)	Velay central (Le Puy-en-Velay) ; qq. traits du Velay méridional	Velay central (Le Puy-en-Velay)
Temps	Témoin le plus ancien, probt ca 1750-1760	1798	Probt postérieur à *B*	1836	Impression non datée, mais postérieure à 1845
Lieu d'écriture, scribe, position sociale	Ø	Ø	Copie signée J.G. Bernard (non identifié)	Copie réalisée pour la Bibliothèque historique dptale de la Haute-Loire	Réalisé au Puy par le libraire-imprimeur François-Marie Clet, arrière-petit-fils d'Antoine
Contenu	3 actes 425 vers	3 actes 453 vers	3 actes 452 vers	5 actes 719 vers	3 actes 428 vers

Table 1. *Le Sermon manqué* d'Antoine Clet,
tradition textuelle et données diasystématiques

Il ressort des données ci-dessus que l'axe linguistique le plus exploitable est d'ordre diachronique. En effet, une chronologie absolue ou relative des textes peut être établie avec quelque assurance et permet une interrogation sur un siècle environ. En revanche, les témoins sont diatopiquement homogènes et ne posent pas de problème spécifique à ce sujet. D'un point de vue diastratique et diaphasique, les informations disponibles sont a priori peu exploitables ; nous

essaierons en dernier lieu de faire des rapprochements entre les choix graphiques des scripteurs et les contextes de copie/réécriture.

2. Variation textuelle et variation linguistique : analyse des données

2.1. Approche macroscopique

La collation des témoins du texte *Le sermon manqué* nous permet d'inférer les premiers résultats suivants :

1) Nous ne disposons pas de témoin autographe.
2) *A* constitue le témoin le plus ancien et il présente, notamment dans les parties françaises, des graphies plus archaïques. Sa place dans la tradition est singulière, car il présente de nombreuses leçons individuelles, qui correspondent souvent à des vers hypo- ou hypermétriques et à des aberrations, dénotant peut-être une plus grande proximité avec un original dont on n'attend pas une élaboration littéraire achevée.
3) *BCE* forment un groupe homogène du point de vue des leçons et corrigent fréquemment *A* ; *BC* sont les plus proches, y compris pour les choix graphiques ; *E*, témoin le plus récent, montre quelques innovations textuelles, mais est très indépendant pour les graphies occitanes.
4) *D* est une réécriture profonde de la pièce, opérée dans une optique de consécration de l'auteur comme figure littéraire et intellectuelle dominante du Puy au XVIIIe siècle, dans une période de formation des études historiques. Il se caractérise notamment par de nombreux ajouts. Il peut être cependant rapproché de *A* avec qui il a de nombreuses leçons communes, plus rarement avec *BCE*.
5) On ne peut aucunement inférer de la tradition des phénomènes de censure ou d'expurgation comme il a été avancé[5], mais davantage de réécritures, d'extensions et, pour les parties occitanes, de mises à jour des faits dialectaux.

2.2. Variation textuelle et évolutions dialectales

Il est attendu que, contrairement aux parties françaises, les parties occitanes, en l'absence d'une graphie de référence, sont sujettes à variation pour noter la langue vernaculaire. Sur la centaine d'années que permet de couvrir la tradition textuelle, on peut percevoir une évolution des systèmes grapho-phonologiques, corroborée par notre corpus des textes de l'ancien Velay et par la dialectologie moderne[6]. Cependant, il convient de prendre en compte le fait que dans la perspective de la dialectologie historique, les choix graphiques ne relèvent pas exclusivement de la transcription de phénomènes dialectaux en cours ou achevés, et qu'ils ressortissent également de choix sociolinguistiquement marqués. Nous présentons dans un

[5] Cf. Roux (2015 : 112).
[6] Nous nous appuyons principalement sur Nauton (1957-1963 ; 1974). Les indications de l'ALMC renvoient systématiquement au pt 21, situé à Polignac, à trois km au nord du Puy-en-Velay. Autres points de référence : pt 9 = Chamalières (nord), pt 22 = Saint-Privat-d'Allier (sud-ouest), pt 23 = Le Monastier (sud-est).

premier temps les évolutions phonologiques les plus saillantes mettant les différents témoins en opposition :

2.2.1. Vocalisme

Fermeture de /ˈe/ en syllabe finale tonique en pré-fermée /ɪ/ (parfois jusqu'à [i]), notée <i> (Nauton, 1974 : 37-38) : 'plaisir' *plazi* E (*plasé* A, *plaze* B, *plazé* C) ; 'seul' *souli* E (*soulet* ABCD), ALMC 1650 [suvi]. De même pour la fermeture de la nasale /ˈẽ/ en pré-fermée [ˈĩ] notée <in> : 'prendre.INF' *prindre* D (*prendre* ABCDE), ALMC 58* [prˈĩdre].

Fermeture (parfois jusqu'à la réduction) de la diphtongue /eɨ/ en /iɨ/ (parfois jusqu'à [i]) (Nauton, 1974 : 99) : formes contractes *a+lo* : *i y* D (*ey* AB, *ei* BCDE), ALMC 73 et 1795 [iɨ], et *de+lo* : *di dy* D (*dey* ABDE, *dei* ABCD) ; 'connaître.PRS.2SG' *counisses* D (*couneisses* D), PRS.2PL *counissé* DE (*couneisses* A, *couneisse* B, *couneissé* CD), mais ALMC 1897 pt 9 et 21 [kunˈeɨse], contre pt 22 [kunˈiɨse], pt 23 [kunˈɪɨse] ; 'presser.PTCP' *prissa* E (*preyssa* A, *preissa* BCD, *preïssa* BC, *preiça* E).

Évolution et dialectalement surévolution de la diphtongue /ˈuo̯/ en /ˈiw/ puis /ˈɪw/ (Nauton, 1974 : 194), la pré-fermée étant notée par le graphème <e> : 'mulet' *mieou* E (*miou* ABCDE), ALMC 480* pt 21 SE et NW [mjiuw], 21NE [mjɪw] ; 'cul' *quieou* E (*tiou* AD, *quiou* BC), ALMC 1314* [ʧiu̯].

2.2.2. Consonantisme

Évolution de la sifflante sourde /s/ en postalvéolaire /ʃ/ devant voyelle haute, notée par le digramme <ch> (Nauton, 1974 : 170). Composé ĔCCE+HĪC : 'ici' *ychi* C, *eichi* E (*eici* ABCDE, *icy* A, *eïci* B, *eicy* D, *ici* ABDE), ALMC 1847 pt 21 [iɨʃi] ; 'bénédiction' *bénédichiou* E (*benedictiou* A, *beneditiou* B, *bénédictiou* C) ; 'cinq' *chinq* CE (*cinq* ABD), ALMC 1877 [ʃẽ] [ʃĩk].

Évolution de la sifflante sonore /z/ en postalvéolaire /ʒ/ devant voyelle haute, notée par <g>/<j> (Nauton, 1974 : *ibid.*) : 'entendre.INF' (< AUDĪRE) *ougi* D, ALMC 1897* [uʒi], PST.3PL *oujion* E (*ousion* AD, *ousiont* C), FUT.2PL *ougiré* DE (*ousiri* A, *ausiret* B, *ouziret* C, *ouziré* D), PTCP *ougi* E (*ousy* A, *ousi* BC) ; 'fusil' *fugi* D (*fusi* BC), ALMC 378 [fyʒi].

Évolution en /ts/ de l'affriquée sourde /tʃ/, notée <ts> (Nauton, 1974 : 158) : 'chaire' *tsadeire* E (*chadeyre* ABDE, *chadeïre* B, *chadeire* BCDE), ALMC 1685 [tsadˈɛɨra] ; 'chaloir.PST.3SG' *tsau* E (*chau* ABCD, *chaut* A, *chou* B, *tchau* D), ALMC 738* et 1552* [tsu] [tsou̯], PST.3SG *tsougué* E (*chougué* ABC, *chugué* D) ; 'chemin' *tsami* E (*chami* ABCDE), ALMC 114* [tsamji] ; 'chercher.INF' *tsartsa* E (*charcha* ABCDE), ALMC 168* [tsertsa] ; 'prêcher.INF' *pretsa* E (*precha* ABCDE, *prêcha* CDE), ALMC 1693 [prɪtsa], FUT.3SG *pretsaro* E (*prechero* A, *precharo* BC, *prêcharo* D), PTCP *pretsa* E (*precha* AB, *prêcha* CD).

Évolution en /dz/ de l'affriquée sonore /dʒ/, notée <dz> (Nauton, 1974 : *ibid*). 'coucher.INF' *coidza coudza* E (*coigea* A, *couigea* BCD, *couigëa* D), ALMC 1453 [kwiɨdza].

Palatalisations consonantiques devant une voyelle haute ou un yod (Nauton, 1974 : 164-187). **/dɨ/ noté <gu>** : 'badiner.INF' *baguina* E (*badina* ABCD) ; 'courir.PTCP' *coureguiu* E (*courrediu* A, *courediu* BC) ; 'diable' *guiable* E (*diable* ABCDE). **/kɨ/ noté <ti> :** 'cul' *tiou* AD (*quiou* BC, *quieou* E), ALMC 474 [ʧiu̯] ; 'toux.PL' *tiaux* D (*quiaux* B, *quiaus* C, *quiaou* E). **/nɨ/ noté <gn>** : 'animal' *agnimau* DE (*animau* ABC) ; 'finir.INF' *figni* D, *fignir* E (*finir* ABC) ; 'tenir.INF' *tegni* D (*tenir* ABC, *teny* E). **/ʧ/ noté <qu>** : 'matin' *maqui* CE (*mati* ABCD), ALMC 1432 et 1441 [maʧi] ; 'sortir.INF' *sourqui* E (*sourti* ABCDE, *sorty* A, *sorti*

C), ALMC 706 [surtᵗi], FUT.2SG *sourquiras* C (*sortiras* A, *sourtiras* BDE), FUT.3SG *sourquiro* E (*sortiro* A, *sourtiro* BCD), PTCP *sourqui* D (*sourti* ABCDE, *sourty* AD). **/kʲ/ noté <ti>** est cependant davantage marqué dans *A* : 'ici' *ati* AD, *aty* A, *ty* A (*aqui* BCDE, *àqui* CD, *oqui* C, *qui* BDE), ALMC 1847 [atᵗi] ; 'coquin' *couti* A (*couqui* BCDE) ; 'quelque chose' *tiquon* A (*quiquon* AC, *quicon* BCDE), ALMC 1843 pt 21 [kikɔ̃], 22 23 [tᵗikɔ̃] ; **/gʲ/** est généralement noté <di> (*sediur* ABCDE, *sédiur* E), sauf *segur* D, qui est un des traits méridionaux de ce témoin (cf. Nauton 1974 : 331, carte 46).

La palatalisation des labiales /m/ et /v/ devant voyelle haute entraîne le dégagement d'un [e] noté <e> (cf. Nauton, 1974 : 174 et par ex. ALMC 1625 'fille' [fjˈeja]) : 'cependant' *pamien* E (*pas mins* A, *pas min* A, *pamin* BCE) ; 'rapidement' *viestament* C (*vistamen* B, *vistament* E).

Simplification de la latérale palatale /ʎ/ en /j/ (Nauton, 1974 : 172-174), notée <i>/<ill>. 'rapidement' *iau* E (*liau* ABCD), ALMC 1067, 1453 et 1866 [jau̯], et le composé *beiau beîau* E (*beliau* ABCD, *béliau* C) 'beaucoup', ALMC 1864 [bejau̯].

Amuïssement de /-v-/ dans les formes de l'imparfait de *aver* (Ronjat, 1930-1941 : § 642). 1SG *ayot oyo oyot* D (*aviot* ACE, *avio* B) ; 2SG *ayas* D ; 3SG *oyo* ACD, *ayo* BDE (*avio* B, *aviot* C) ; 2PL *aia* E (*avias* A, *avia* BCD), ALMC 1889 1SG [ajo], 2SG [ajas], 3SG [ajo], 2PL [aja].

Les exemples relevés ci-dessus paraissent recouper la chronologie obtenue par l'étude philologique des témoins. En effet, les manuscrits considérés comme les plus récents attestent davantage des évolutions phonologiques dialectales, dont on peut présumer qu'elles se sont pour beaucoup produites à la charnière de la seconde modernité et du premier XIXe siècle. Aussi le témoin *E* apparaît-il de loin comme le plus novateur, dans une moindre mesure *D* et *C*, ce qui confirmerait pour celui-ci sa postériorité à *A* et *B*.

2.3. Variation linguistique et marquage sociolinguistique

Il serait illusoire de considérer que, malgré l'absence de norme graphique pour l'occitan, les graphies utilisées par les scribes sont transparentes. L'analyse de leurs réflexes graphiques, mis en perspective avec l'ensemble du corpus des textes occitans de l'ancien Velay, doit nous permettre de mieux comprendre le processus de 'patoisisation' entamé au XVIe siècle et qui ne peut être conçu comme une simple adaptation pour l'occitan du système grapho-phonologique du français, langue ressource des scribes une fois les *scriptae* éteintes.

2.3.1. Vocalisme

Notation de /u/. Le digramme <ou> apparaît au XVe s. et devient majoritaire au siècle suivant. L'emploi de <o>/<u> relève d'une imitation du système graphique fr. ou latin (idem pour le phonème nasal équivalent) : 'insolent' *insolent* A, mais *insoulen* B, *insoulent* CDE ; 'monter.INF' *monta* AB, mais *mounta* BCDE (ALMC 79 [mũta]) ; 'tomber.PTCP' *tomba* AB, *tumba* C, mais *toumba* BCDE ; 'voleur' *voleur* ABC, *vouleur* DE (emprunt au fr. avec adaptation phonologique).

Voyelle finale atone /a/. Après 1500, le morphème désinentiel de 3SG (classe A) et les morphèmes féminins des catégories nominales sont notés <e>, malgré un timbre ouvert conservé jusque dans les dialectes modernes d'après l'ALMC : 'payer.PST.3SG' *paye* ABCDE, mais *paya* D (cf. ALMC 1882 /-a/) ; 'nonne portière d'un couvent' *touriere* AB, *tourriere* BC, *tourrière* CD, *tourière*

D, mais *tourriera* E ; 'trente' *trente* ABCD, mais *trenta* E (ALMC 1880 [trˈĩta]) ; 'un.ART.INDF.F' *une* ABCDE, mais *una* DE, *ena* E.

Fermeture de la voyelle nasale tonique /ã/ en /õ/ (parfois d'un degré supplémentaire). Phénomène attesté depuis le XVIIe s., avec la graphie <on>/<om> : 'chambre' *chombre* CD, mais *chambre* BE (ALMC 719 [tsˈõbra]) ; 'demander.PRS.1SG' *demonde* CD, *damonde* E, mais *demande* AB (ALMC 1828* IMP.1SG [damˈõda]).

Notation de la diphtongue tonique /ˈaw/ (< AU, A+L, ĪL). Le digramme traditionnel <au> est encore bien employé, concurrencé par le trigramme <aou>, rarement par <ao> : 'mal' *mau* ABCDE, mais *maou* E (ALMC 78 et 1245 [maṷ]) ; 'pauvre' *paure* ABCDE, mais *paoure* CE (ALMC 1538* [pˈaṷr(e)]) ; 'toux.PL' *quiaux* B, *quiaus* C, *tiaux* D, mais *quiaou* E ; 'sauter.PRS.3SG' *saute* ABCD, mais *saoute* E.

Notation de la diphtongue atone /ow/ issue de /aw/. La réalisation [ow], affaiblie aujourd'hui en [u_w], est attestée depuis le XVIe s. par le digramme <ou> : 'sauter.PTCP' *souta* BCDE, mais *sauta* A (ALMC 1514* INF [su_wta]) ; 'sauver.INF' *souva* BCDE, mais *sauva* A, PST.3SG *souvave* BCDE, mais *sauvave* A.

Ouverture de /e/ atone devant rhotique /r/ (Nauton, 1974 : 55-56). Le phénomène est attesté depuis le XVe s. avec le graphème <a> : 'sermon' *sarmou* DE, mais *sermon* ABDE, *sermoun* CDE, *sermou* ABCDE ; 'terrible' *tarrible* BE, mais *terrible* ACD.

Calques du fr. (i) niveau phonologique : 'raisonner.INF' *resouna* A, mais *rasouna* BCE, PRS.2SG *raisounes* A, mais *rasounes* BCDE (cf. *razo* CensChomB MedPrivada, *raso raison* MedPrivada, *rasou* NoëlsCordat SocAmisPuy) ; 'oser.PTCP' (< AUSARE) *osa* A, mais *ouza* BC, *oudza* E ; 'traiter.PTCP' *treta* A, mais *trata* BCDE. **(ii) niveau morphologique** : notation de la désinence 3SG du parfait des verbes de classe A (/-e,ɛ/) avec le morphème fr. de l'indicatif imparfait : *donnoit* A (*douné* BC, *donne* E) ; *arribait* D (*arribé* ABCE).

2.3.2. Consonantisme

Notation de /dz/ (< /dʒ/). Le graphème <j> est traditionnel, relayé par <dz> : 'jamais' *jamais* AC, *jamay* A, *jamaï* B, *jamai* BCDE, mais *dzamay* C, *dzamai* E (ALMC 1341 [dzamai̯]) ; 'jour' *jour* ABCD, mais *dzour* E (ALMC 1431 [dzur]).

Amuïssement des consonnes finales (Nauton, 1974 : 238-241). De date variable selon la nature des consonnes, il paraît avancé dans les dialectes vellaves centraux dès le XVIIe siècle. **Finale dentale** : 'enfant' *effon* ABD, *efon* E, mais *efont* A, *effont* C (ALMC 1626 [ɪfõ]) ; 'moment' *moumen* BD, mais *moument* ACE ; 'pied.PL' *pés* BE, *pes* CD, mais *peds* A ; 'souvent' *souven* BE, mais *souvent* ACD. **Finale labiale** : 'coup' *co* DE (*cop* ABCD) ; 'trop' *tro* DE, mais *trop* ABCDE. **Finale vibrante** : l'amuïssement de /-r/ des morphèmes désinentiels de l'infinitif est général pour les verbes de classe A (par ex. *vida* D 'vider', *vira* ABC 'virer'), mais les infinitifs de la classe E/I présentent l'allographie *veni* BCE, *vegni* DE, contre *venir* AB, *vegnir* D. **Finale sifflante**, notamment morphème du pluriel {-s} : 'mot.PL' *mou* E (*mous* A, *mouts* BC, *muots* D) ; 'passage.PL' *passadou* D (*passadoux* A, *passadous* ABCDE) ; 'planter.PTCP.PL' *planta* DE (*plantas* ABC).

Calques du fr. (i) niveau graphique : 'serrer.INF' *sarra* ABCD, mais *sara* E. **(ii) niveau morphologique** : 'parler.PRS.3SG' *parlent* C, mais *parlon* AE, *parloun* BD (ALMC 1882-1886 /-ũ/). **(iii) niveau lexical** : lexème savant fr. *prédicateur*

ACDE, *predicateur* B, mais forme occitane avec suffixe agentif nominatif *préchaire prechaire* dans les ajouts de D.

Les données relevées ci-dessus dessinent deux tendances antithétiques qui individualisent certains témoins et qui paraissent s'accorder avec les remarques faites en 2.2. : (i) les scribes des témoins les plus anciens, surtout *A*, ont des réflexes graphiques qui les amènent à rapprocher les formes occitanes des normes (ortho)graphiques, principalement du français moderne et incidemment du latin, mais sans ignorer totalement des principes graphiques d'origine médiévale. (ii) les scribes des témoins les plus récents, surtout *E*, opèrent un recentrement vers une plus grande transparence phonétique des formes occitanes, notamment par le recours à des digrammes ou trigrammes généralement empruntés au système grapho-phonologique du français. Dans les deux cas, la graphie de l'occitan reste dépendante du français, mais cette divergence traduit des statuts sociolinguistiques différents des témoins.

3. Conclusion

En l'absence de norme graphique de l'occitan, les témoins de la pièce *Le Sermon manqué* montrent un haut degré de polymorphisme linguistique, à la fois au sein de la tradition et à l'intérieur de chaque témoin. Cette variation est liée à la variation textuelle dans la mesure où les systèmes graphiques des scribes marquent des évolutions assez caractéristiques. Quoique l'écriture de l'occitan reste dans un rapport de soumission au français, langue ressource depuis la fin du Moyen Âge, nous faisons l'hypothèse que la recherche de solutions graphiques, allant de l'imitation des normes orthographiques françaises à une forme plus marquée d'oralité, va de pair avec un changement de point de vue sur la langue vernaculaire. En effet, on peut estimer que ce changement a partie liée avec le contexte socioculturel du premier XIXe siècle, engagé dans la patrimonialisation et la valorisation des sources considérées comme populaires. Aussi, le processus de patoisisation non seulement s'inscrit dans la longue durée, mais connait des phases dont les témoins de la pièce d'Antoine Clet donnent un exemple caractéristique.

Bibliographie

ALMC = Nauton, P. (1957-1963).

Duval, F. (2015). « Les éditions de textes du XVIIe siècle ». In D. Trotter (dir.), *Manuel de la philologie de l'édition*, Berlin & Boston : De Gruyter, p. 369-393.

Martel, P. (2013). « L'occitan ». In G. Kremnitz (dir.), *Histoire sociale des langues de France*, Rennes : PUR, p. 511-532.

Nauton, P. (1957-1963). *Atlas linguistique et ethnographique du Massif Central*. Paris : CNRS (4 volumes).

Nauton, P. (1974). *Géographie phonétique de la Haute-Loire*. Paris : Les Belles Lettres.

Ronjat, J. (1930-1941). *Grammaire istorique* [sic] *des parlers provençaux modernes*. Montpellier : Société des langues romanes (4 volumes).

Roux, J. (2015). *Huit siècles de littérature occitane en Auvergne et Velay. Morceaux choisis. Édition bilingue (occitan/français).* Lyon : EMCC.

Surrel, V. (2016). « La trajectoire d'une littérature bilingue au XVIIIe siècle : le cas d'Antoine Clet du Puy-en-Velay ». In W. Pfeffer et J. Thomas (dir.), *Nouvelles recherches en domaine occitan : approches interdisciplinaires. Colloque de l'Association internationale d'études occitanes, Albi, 11 & 12 juin 2009*, Turnhout : Brepols, p. 137-148.

Weil, F. (1993). « La fonction du manuscrit par rapport à l'imprimé ». In F. Moureau (dir.), *De bonne main. La communication manuscrite au XVIIIe siècle*, Oxford & Paris : Voltaire Foundation & Universitas, p. 17-28.

Étude diachronique du gascon béarnais au sein des textes officiels

Quelle(s) variation(s) par rapport au gascon ?

Jérémy ROUQUIER
Étudiant, master « Linguistique et dialectologie »,
Université Jean Moulin Lyon 3

Résumé

S'il existe bien une « scripta béarnaise », variété écrite possédant des traits spécifiques, peut-on tracer une chronologie de la naissance et de l'évolution desdits traits ? Pour tenter de répondre à cette question, nous étudions un corpus composé de textes béarnais et gascons, écrits entre le XIIIe et le XVIIe siècle afin d'étudier les traits par lesquels Jean Laffite définit la scripta béarnaise, ainsi que les traits spécifiques au béarnais sélectionnés dans les ouvrages de Massoure consacrés au Gascon.

Mots-clefs : gascon ; béarnais ; dialectologie ; corpus diachronique ; scripta béarnaise.

DOI : 10.5281/zenodo.3464453

1. Introduction

Peut-on parler de « scripta béarnaise » pour les textes écrits en gascon dans la région béarnaise ? Si c'est le cas, peut-on tracer une chronologie de la naissance et de l'évolution des traits spécifiques de cette variété ? Trouve-t-on des traits spécifiques qui sont apparus plus tardivement que d'autres ou, au contraire, des traits spécifiques présents dans les textes les plus anciens, mais qu'on ne constate plus ultérieurement ?

Si on se réfère à l'*Atlas Linguistique de Gascogne* (dorénavant ALG), le gascon béarnais possède des traits spécifiques, mais n'est pas aussi singulier que le « gascon noir » des Landes. Il est surtout moins bien délimité : certaines formes distinctes sont présentes au-delà du Béarn, d'autres uniquement dans certaines vallées pyrénéennes. Malgré cela, l'indépendance politique du Béarn, puis son autonomie jusqu'à la Révolution française, ont-elles pu conduire à la création d'une scripta particulière ? C'est ce que considère Jean Lafitte, qui a défini une poignée de traits décrivant la scripta béarnaise. Cette scripta reprend-elle des traits linguistiques spécifiques de la variété de l'ancienne vicomté ?

Nous souhaitons, grâce à une approche diachronique, discerner des influences du gascon béarnais par rapport à son entourage linguistique, et vérifier si la présence de variations observées dans l'ALG se retrouve dans ces textes, puis auquel cas le profil de leur distribution. Nous analyserons également la présence des traits décrits par Laffitte.

Notre répondrons à cette question au travers de l'analyse de textes officiels émis dans le Vicomté du Béarn entre le XIIIe et le XVIIe siècle, et les comparerons aux données récoltées sur les cartes de l'*Atlas linguistique et ethnographique de la Gascogne* ainsi qu'à des textes gascons d'époques similaires.

2. Traits et corpus étudiés

2.1. Traits étudiés

Nous souhaitons porter notre attention en priorité sur huit traits. Quatre d'entre eux sont définis par Jean Lafitte, et quatre autres ont été distingués par les travaux que Jean-Louis Massoure a effectués en se fondant sur l'ALG.

2.1.1. Traits de la scripta béarnaise définie par Lafitte

Jean Lafitte a défini la scripta béarnaise comme comportant les traits suivants (Lafitte, 2005 : 103-104) :

a) Doublement d'une voyelle ou formation d'un digramme, transcrit pour compenser la disparition d'une consonne étymologique : *paa* (pain), *Morlaas* (Morlans), *pees* (pied), etc. Ce doublement affecte des mots où la finale *-r*, *-d* ou *-n* s'est amuïe. Parmi ces mots, nous nous concentrerons sur ceux qui se terminaient par une nasale en finale.

b) Bêtacisme : phonème /β/ transcrit indifféremment <b> ou <v> : *baque* ou *vaque* (vache), *vesin* ou *besin* (voisin), etc.

c) Utilisation du <x> pour transcrire le son /ʃ/ : *medix* (même), *aixi* (ainsi), etc.

d) Utilisation du <f> pour retranscrire un /h/ initial, tel que dans *foec* (feu) ou *filhas* (fille), alors que ces mots sont prononcés avec une initiale aspirée.

2.1.2. Traits spécifiques définis par Massoure et l'ALG

Dans l'ouvrage *Le gascon ''lengatge estranh''* (Massourre, 2015), Jean-Louis Massoure étudie les données de l'atlas linguistique de Gascogne. Au sein de ces données, nous avons distingué quatre traits spécifiques de tout ou d'une partie de la zone béarnaise :

a) L'effacement de la nasale en fin de mot (qui se traduit parfois par le report de la nasalité sur la voyelle) : *pan* devenant /paã/ ou /pa/.

b) L'article indéfini masculin singulier prononcé /y/ face aux réalisations /yn/ ou /yŋ/ ailleurs dans la zone gasconne.

c) La particule « énonciative » verbale exclamative /e/ (et non pas *bé* ailleurs en zone gasconne). Cette particule, en attaque d'une phrase, permet d'insister sur l'information qu'elle transmet, et souvent de valider des informations introduites intérieurement dans le dialogue (Fossat, 2006). Un exemple en est « *bè hèi calor* » (« qu'il fait chaud ! »).

d) L'épenthèse du /b/ sur les pronoms possessifs *sue/sua/tue* en fin de phrase.

2.2. Le corpus

2.2.1. Spécificité du corpus : des textes officiels

Les textes de ce corpus sont écrits par des érudits, et ne représentent que partiellement la langue orale qui avait alors cours dans le Béarn. La notion de scripta (que je définis ici comme le système graphique d'une région donnée à un moment donné) implique que certaines conventions régissent l'écriture des documents. Ce sont ces conventions que nous voulons étudier. Ces textes sont des

textes de loi, liés soit aux coutumes, soit à la gouvernance des territoires. Ces textes ont en commun d'avoir été édités en vertu de considérations historiques ou philologiques. Les éditeurs ont tâché de rester fidèles au texte qu'ils ont retranscrit et, si la possibilité d'erreurs n'est pas à exclure, la sélection de formes « meilleures » qui a pu être de mise dans l'édition de certains textes à portée littéraire n'est pas à craindre ici.

2.2.2. Charte des boucheries d'Orthez

Écrite le 2 novembre 1270, à Pau, sur ordre du Vicomte Gaston VII Moncade, cette charte est l'un des plus anciens textes conservés en béarnais, et le premier de cette longueur. Ce document traite de l'emplacement des boucheries (que le vicomte veut réunir en deux endroits), du commerce de la viande, des droits des bouchers. Ce document a été publié par Pierre Tucoo-Chala dans la revue *Annales du Midi* en 1957 (Tucoo-Chala, 1957). Le texte, intégralement analysé, compte 1201 mots.

2.2.3. Fors de Béarn

Ces fors, ici des coutumes issues de la tradition et acceptées par le seigneur, proches des *fueras* basques, sont un regroupement de textes légaux (privilèges, règlements, décisions de justice, tarifs, formulaires) accumulés dans la vicomté de Béarn entre les XIe et XIIe siècles. La version parvenue jusqu'à nous a été écrite au XIVe siècle, mais semble fidèle aux textes antérieurs selon les témoignages (Mazure & Hatoulet, 1841 : X). Le lieu d'écriture des textes n'est pas précisé, mais dans la mesure où ceux-ci ont été écrits sous ordre des Vicomtes, il est probable qu'ils aient été écrits dans les capitales du Béarn de l'époque, Morlaàs, puis Orthez. Selon la même logique, le texte parvenu jusqu'à nous aurait également été écrit à Orthez.

Ont été analysées les 35 premières pages du for général du Béarn, les 25 premières pages du for de Morlàas, ainsi que les fors d'Oloron, d'Ossau et d'Aspe. La somme de ces extraits compte approximativement 18 000 mots.

2.2.4. Le livre de syndic des États du Béarn

Écrits entre 1488 à 1521, ces documents datent presque tous du règne de Catherine de Navarre (1483-1517). À cette époque, dans la vicomté du Béarn, les États possèdent auprès de la Vicontesse un rôle majeur dans la gestion du domaine. En leur sein, les syndics étaient chargés de la préparation et de la gestion des séances des assemblées d'États. Ils distribuaient et géraient les dépenses ordonnées par les États, et servaient d'intermédiaires entre le Grand corps et le second corps ou Tiers-état qui ne pouvaient pas négocier entre eux directement. Représentants des états face au vicomte et à ses conseillers, ils rédigeaient les cahiers des doléances à destination des dirigeants, et étaient habiletés à se rendre au sein de toutes les institutions pour enquêter s'ils le jugeaient nécessaire.

Ces manuscrits ont été publiés par Léon Cadier en 1889 pour la première partie, puis par Henri Courtault en 1906. Ils sont rassemblés au sein du registre C680 des archives des Basses-Pyrénées. Il s'agit d'une sélection de documents, qui ont été transcrits dans ce registre. Il ne s'agit pas d'un recueil officiel, mais d'une compilation au jour le jour.

On y trouve des comptes rendus de sessions des États, des fragments de Cahiers, des extraits de délibérations, des commissions, instructions et mémoires aux lieutenants du souverain pour la tenue des États, des rôles de répartition, des comptes de levée d'impositions et de dépenses faites, soit par le Trésorier du Béarn, soit par les Syndics des États.

Concernant le texte lui-même, on peut distinguer quatre périodes, et trois auteurs différents, qui se sont succédés : Jean de Castagnède (syndic du Béarn puis conseiller du pays), Roger de Boeil et Pèes de Tisnès (syndics). Les documents originaux (recopiés par ces trois hommes) ont été écrits dans diverses localités. Parmi nos textes, 15 d'entre eux l'ont été à Pau, 1 à Orthez, 2 à Sauveterre (aujourd'hui Sauveterre-en-Béarn), 2 à Lescar et 2 à Pampelune. Bien que les auteurs des manuscrits (qui sont des copies) restent les mêmes quelle que soit la localité d'origine des manuscrits originaux, une influence locale n'est pas à exclure complétement.

Ont été analysés les textes I à X, XLV à L, et LXXX à LXXXIV, comprenant en tout approximativement 16 500 mots.

2.2.4. Los fors e costumas deu royaume de Navarra deça Ports

L'original de ce texte fut rédigé en 1611 à Pau. Il est le résultat de la volonté du roi Henri IV de donner, en 1608, à la Basse-Navarre (située au sein du Pays basque) un for spécifique, alors qu'elle était jusqu'ici régie par le for général de Navarre et les usages locaux (Lafourcade, 2003). La Basse-Navarre, bien que n'appartenant pas au Béarn, était alors contrôlée par le Vicomte du Béarn, ce qui explique pourquoi sa coutume a été rédigée à Pau et en occitan béarnais plutôt qu'en basque.

Ce texte a été écrit peu avant que le Parlement de Navarre ne soit forcé par Louis XIII à adopter le Français. Dans les autres cours, les élites locales, bien qu'elles parlaient le français, continuèrent à utiliser le béarnais, principalement pour mettre en avant leur autonomie face à la couronne de France (Joseph, Destafano, Jacobs & Lehiste, 2003 : 181). Cette situation illustre l'importance qu'avait pour l'élite béarnaise l'utilisation d'une langue locale et non pas de la langue du roi de France, choix qui supporte l'idée symbolique de l'exceptionnalité du Béarn et de ses coutumes.

La version du texte étudié fut imprimée à Pau par Jérôme Dupoux en 1722.

Les rubriques les plus spécifiquement étudiées sont les rubriques 1 à 9 (p. 1-33), rubriques 24 à 25 (p. 74–87) et rubriques 29 à 35 (p. 117–139), comptant approximativement 11 500 mots.

2.2.6. Textes gascons

Les deux textes gascons que nous avons sélectionnés comme corpus de contrôle sont le cartulaire du chapitre de l'église métropolitaine Sainte-Marie d'Auch, qui date de la seconde moitié du XIIIe siècle, ainsi que les comptes *consulates* de la ville de Riscle, écrits entre 1441 à 1507. Ces textes sont considérés comme typiquement gascons par le *Lexikon der Romanistischen Linguistik* (ou LRL) (Holtus, Metzeltin, & Schmitt, 1995 : 450). N'ayant été écrits ni en zone béarnaise ni dans les landes, où le « gascon noir » est connu pour ses particularités, ces deux textes se veulent représentatifs de l'occitan gascon non béarnais tel qu'il pouvait être écrit à partir du Moyen Âge classique et jusqu'au début de la Renaissance. Ces textes ont également été choisis pour avoir été écrits dans des régions ayant été gouvernées par le la Vicomté du Béarn, comme le Bigorre. L'étude de ces textes a pour objectif de distinguer les traits spécifiquement béarnais de traits qui auraient pu être utilisés de manière plus générale en Gascogne.

3. Observations provisoires

En table 1 sont indiqués le nombre de termes possédant les traits étudiés, ainsi que le nombre de termes qui possèdent un trait alternatif à celui étudié. Les deux premières colonnes concernent également les groupes nominaux.

	Voyelle double	**Nasale finale**	**Bêtacisme**	**Non bêtacisme**	**<x> pour /ʃ/**	**Non <x> pour /ʃ/**
Charte des boucheries	4 (1)	12 (5)	9	8	3	0
Fors Béarnais	55 (12)	22 (7)	32	18	17	0
Livre des syndics	26 (8)	31 (11)	9	14	18	1
For et coutumes de Navarre	16	28	5	12	3	6

Table 1. Termes possédant un trait spécifique par document.

Certains traits de la scripta béarnaise définis sont déjà présents dans la charte des bouchers d'Orthez. Concernant les noms communs, on compte 4 digrammes : *fee* (« foi » ; §17, l. 3), *maa* (« main » ; §10, l. 5), etc., pour 7 termes conservant leur <n> final : *algun* (« quelque », §4, l. 5), *bezin* (« voisin », §4, l. 5), *mession* (« dépense » ; §6, l. 1), etc.

L'ensemble des noms propres conservent leur <n> final, à l'exception de Morlàas (« Morlans » ; §21, l. 2). On retrouve également 9 cas de bêtacisme : *biele* (« ville » ; §2, l. 5), *bolem* (« voulons » ; §6, l. 5), etc., ainsi que quelques cas d'utilisation du <x>, tels que : *arreconexem* (« reconnaissons » ; §1, l. 7).

La présence limitée de certains traits, notamment du digramme en final, est peut-être lié à son ancienneté : le béarnais n'avait commencé à remplacer le latin que depuis quelques décennies, et il est possible que la scripta n'était pas encore formée. Nous devons noter que la brièveté du texte empêche toute conclusion définitive à ce sujet.

Ces traits sont présents de manière plus systématique au sein des fors Béarnais : sur les 90 pages analysées, on retrouve 55 digrammes au sein de noms communs : *camiis* (« chemins » ; p. 16, art. 35, l. 2), *motoo* (« mouton » ; p. 22, l. 7), etc., face à 22 termes possédant un < n > final : *mayson* (« maison », p. 17, art. 39, l. 9), *messions* (p. 27, l. 15), etc. Certains termes, tels que *messions*, sont tantôt écrits avec un digramme, tantôt avec une nasale.

Concernant les noms propres, 12 termes possèdent un digramme : *Avidoos* (« Abidos » ; p. 12, art. 28, l. 2), *Samsoos* (« Samsons » ; p. 123, art. 42, l. 21) et 7 un <n> final : *Balensun* (p. 13, art. 30, l. 5), *Bertran* (p. 26, art. 62, l. 8). À noter qu'on trouve également la dénasalisation complète *Balensu* (p. 36, l. 1).

On retrouve également 17 termes contenant un <x> : *aixi* (« ainsi » ; p. 21, l. 3), *medixs* (« mêmes » ; p. 21, art. 48, l. 14), etc., ainsi que 32 cas de bêtacisme : *abesque* (« évêque » ; p. 3, art.1, l. 5), *recebe* (« reçoit », p. 115, art. 14, l. 2), etc. Dans tous ces cas, le nombre d'occurrences du traitement béarnais dépasse le nombre de traitements alternatifs, ce qui ne sera plus le cas dans les textes plus récents.

La présence de <f> initial ne concerne qu'un seul terme, mais pour autant aucun autre ne possède de <h> en initiale.

La présence de ces traits se décline dans une certaine mesure dans les textes postérieurs et particulièrement dans les fors et coutumes de Navarre. Ils n'en restent pas moins présents. Sur les 70 pages étudiées du syndic des états du Béarn, on retrouve 18 <x> remplaçant des /ʃ/[1] et 9 cas de bêtacisme, tandis que la graphie de certains mots a évolué depuis l'époque de l'écriture des fors du Béarn : *volem* (p. 3), *vescomte* (p. 5), etc. On retrouve moins de digrammes par rapport aux termes comprenant un <n> final, que ce soit au niveau des noms communs (18, dont 7 représentent une nasale étymologique pour les premiers, 20 termes pour les seconds) ou des noms propres (8 pour les premiers, 11 pour les seconds).

Dans les ouvrages les plus anciens, on ne trouve pas de graphies alternatives au <x>, elles-mêmes peu nombreuses, car l'occitan n'a pas connu la dépalatalisation qu'on trouve dans les domaines oïlique et francoprovençal. Dans les « fors e costumas », le <x> côtoie les graphies <ch> ou <sh>, et on n'en retrouve plus que trois. 16 termes comportent un digramme, 4 d'entre eux remplaçant une nasale, face à 28 termes possédant un <n> final. Le bêtacisme est lui aussi presque absent : on ne retrouve que 5 cas, et l'utilisation du <v> se fait plus systématique.

Ainsi, la présence des traits définis par Lafitte augmente dans un premier temps pour refluer sans disparaître. De plus, l'influence française peut être devinée dans les livres des syndics, et surtout dans les fors et coutumes de Navarre.

Dans le premier, on remarque l'ajout d'un <h> au terme *chrestien* (p. 6, l. 16), l'expression *à cause de* (p. 10). On retrouve également un grand nombre de termes à terminaison en *-ion* tels que *deliveration* (T1, p. 12, l.11), *expedition* (T2, p. 74, l. 16), *remonstration* (T2, p. 78, l. 5), *questions* (T1, p. 78, l.7), *occasion* (T1, p. 146, l. 3). La plupart de ces termes sont présents sous une autre forme dans le dictionnaire de Lespy (notamment avec des terminaisons en *-ou*) ou non présents. Si certains termes occitans possèdent cette terminaison, il semble que la plupart de ceux qu'on retrouve dans le texte sont des emprunts français. On trouve également quelques groupes de mots qui semblent être des expressions directement empruntées au français, tel que *destruction de la cause publique* (T1, p. 13, l. 6).

On retrouve également dans les fors et coutumes de Navarre des termes à terminaison -y : *prefency* (p. 14), *notary* (p. 27, l. 8), etc., l'utilisation d'accent : *lô jorns* (p.19), *âadge* (p.7), *Confelheê* (p. 9), etc., ainsi que des emprunts tels que *diminution* (p. 26), *faux* (p. 18), etc., ou des expressions telles que *faute de* (p. 123), *et pareillement* (p. 5), etc.

Les traits typiquement béarnais ne sont présents que marginalement. On ne trouve qu'une poignée de mots où la dénasalisation est complète, tous dans les fors du béarn, et l'article indéfini masculin est écrit non pas <u>, mais <un> ou <ung>. Les autres traits ne sont pas présents : cela est dû au type même du corpus, qui comporte des tournures de phrases n'employant pas ces traits ni sous une forme béarnaise, ni sous une forme gasconne ou occitane. Concernant les textes gascons, leur étude est encore en cours. Les conclusions provisoires penchent vers l'utilisation de certains traits béarnais, mais pas de tous. Par exemple, on retrouve des occurrences de <f> initial et de bêtacisme, mais le <x> n'est pas utilisé. Ainsi, si on parle d'influence béarnaise, il faudrait la considérer comme limitée.

[1] Ce n'est pas le cas de tous : *exaction* (T1, p. 10, l. 16), par exemple, semble être un emprunt au français.

4. Perspectives

Cette étude n'étant pas achevée, de nombreux travaux restent à faire. En premier lieu, comparer l'ensemble des formes obtenues avec les cartes de l'ALG nous semble nécessaire pour déterminer et délimiter une possible influence de la langue orale sur la scripta.

Ensuite, seule une partie des textes béarnais a été analysée complétement, et les textes gascons ne l'ont pas encore été en profondeur. Leur analyse complète nous permettra de définir avec plus de finesse les spécificités de la scripta béarnaise. La question de la graphie des noms propres nous semble également digne d'intérêt. Nous souhaiterions, par une étude anthroponymique et toponymique, étudier l'hypothèse de l'emplacement des toponymes et la région d'origine des anthroponymes pourrait influer sur la graphie. Enfin, après avoir répertorié l'ensemble des termes possédant un digramme ou une consonne finale nasalisée, nous souhaiterions déterminer si, d'un point de vue morphologique, syntaxique ou étymologique, des schémas quant au choix de la graphie employée peuvent apparaitre.

Bibliographie

Alibert, L. (1966). *Dictionnaire occitan-français selon les parlers languedociens*. Toulouse : Institut d'études occitanes.

Bec, P. (2000). *Manuel pratique de philologie romane*, tome 1 : Italien, espagnol, portugais, occitan, catalan et gascon. Paris : Editions A. et J. Picard.

Fossat, J.-L. (2006). « Les « petits mots » énonciatifs gascons : Le cas de *bè* énonciatif ». *Cahiers de Grammaire*, (30), p. 159–174.

Holtus, G., Metzeltin, M., & Schmitt C. (1995). *Lexikon der romanistischen Linguistik (LRL). Band II, 2, Die einzelnen romanischen Sprachen und Sprachgebiete vom Mittelalter bis zur Renaissance*. Tübingen : Max Niemeyer Verlag.

Joseph, B., Destafano, J., Jacobs, N. & Lehiste, I. (2003). *When Languages Collide: Perspectives on Language Conflict, Competition, and Language Coexistence*. Columbus : Ohio State University Press.

Lafitte, J. (2005). *Situation sociolinguistique et écriture du gascon aujourd'hui.* Thèse de doctorat, Université Rennes 2.

Lafourcade, M. (2003). « Les fors basques et les droits de l'homme ». In *Lapurdum, Revue d'études basques*, (8), p. 329–348.

Massourre, J.-L. (2015). *Le gascon "lengatge estranh"*. Villeneuve-sur-Lot : Jean-Louis Massoure.

Seguy, J. (1966). *Atlas linguistique et ethnographique de la Gascogne*. Paris : Centre National de la Recherche Scientifique.

Tucoo-Chala, P. (1957). « La charte des boucheries d'Orthez (2 novembre 1270). » *Annales du Midi*, 69(40), p. 323–332.

Annoter facilement un corpus complexe

L'exemple de Pyrrha, interface de post correction, et Pie, lemmatiseur et tagueur morphosyntaxique, pour l'ancien français

Ariane PINCHE
Université Lyon 3 (CIHAM UMR 5648)
École nationale des chartes

Résumé

L'annotation morphosyntaxique est une tâche qui occupe activement la communauté scientifique depuis l'émergence des outils numériques dans les sciences humaines, car ces informations sont précieuses aussi bien aux yeux des linguistes que de chercheurs moins spécialisés en sciences du langage pour, entre autres, des études de stylométrie ou de lexicométrie. Toutefois, certaines langues, comme l'ancien français, représentent un véritable défi parce qu'il est très difficile d'annoter des langues sans normalisation graphique.

Pyrrha est une interface d'aide à l'annotation linguistique développée au sein de l'équipe numérique de l'École nationale des chartes. Cette dernière s'appuie sur le lemmatiseur et annotateur morphosyntaxique Pie qui ne possède pas de dictionnaire prédéfini. Pyrrha propose également une interface de relecture pour assurer à l'utilisateur une complète maîtrise de ses données. Ainsi, annoter un texte est plus rapide et plus simple. Il est alors facile de constituer à partir de son corpus un glossaire qui répertorie sous une entrée unique toutes les variations graphiques d'un même lemme ou d'analyser un corpus étendu pour observer, par exemple, l'état du système casuel ou encore détecter certains traits dialectaux.

Mots-clés : ancien français ; interface ; lemmatiseur ; morphosyntaxe ; tagueur.
DOI : 10.5281/zenodo.3464473

1. Introduction

Faire un commentaire linguistique sur un corpus donné est une tâche extrêmement fastidieuse qui demande de consacrer beaucoup de temps à la préparation de données qui bien souvent resteront dans l'ombre de la recherche. Se pose alors la question de l'exhaustivité du relevé, souvent écartée en regard du manque de rentabilité entre le temps consacré à la préparation du corpus et ses possibilités d'exploitation. Il semble alors souvent préférable de consacrer son travail à un phénomène délimité et aisément quantifiable offrant au chercheur une pleine maîtrise de son objet.

Pourtant, les corpus annotés peuvent se révéler très utiles dans de nombreux champs d'investigation en sciences humaines. Une annotation proposant des lemmes pour chaque mot peut aider à repérer des phénomènes phonétiques et à identifier certains traits linguistiques qui à terme pourront être des indices du lieu de composition. Une annotation morphosyntaxique est utile pour étudier des

phénomènes d'ampleur sur un corpus complet comme l'utilisation du système casuel ou encore l'utilisation ou non du pronom personnel sujet, phénomènes en ancien français qui permettent parfois de dater un texte. Elle permet aussi des études encore plus pointues comme les analyses de stylométrie, notamment pour les études d'attribution d'auteur (Mellet, 2002). L'étendue des possibilités fait de ces données cachées des éléments extrêmement précieux pour la recherche, même si l'exploitation immédiate d'une telle masse de données n'est pas toujours évidente.

L'annotation des corpus a souvent été automatisée par les linguistes notamment dans le champ des TAL (Traitement automatisé de la langue) pour ensuite réexploiter les données de manière quantitative. Des lemmatiseurs et annotateurs automatiques ont été développés depuis les années soixante[1]. Les outils existants sont relativement performants pour les langues normées, mais perdent en fiabilité dès lors qu'ils sont confrontés à une grande variété graphique et à une langue en transition où les marqueurs morphologiques ne sont pas toujours utilisés de manière homogène. En outre, bien souvent ces lemmatiseurs n'ont pas été pensés pour un usage par les novices en numérique et ne possèdent pas d'interface graphique qui permettrait un usage sans connaissance préalable en informatique, donnant ainsi parfois l'impression de priver le chercheur d'un complet contrôle sur son annotation.

En effet, la plupart des lemmatiseurs produisent des données brutes en CSV ou TSV sans proposer d'interface de correction de l'annotation. Pourtant cette étape peut s'avérer cruciale pour la qualité du travail scientifique, ne serait-ce que pour un contrôle de la qualité du corpus. Cette étape permet d'améliorer l'annotation du set de données et d'éliminer le « bruit » lié au traitement automatique pour garantir la fiabilité des résultats de l'exploitation future des données et laisser une pleine maîtrise de ces dernières à l'humain. Ces points sont d'autant plus importants quand la production des données dépasse le simple cadre d'un usage individuel pour constituer un corpus annoté pérenne qui pourra par la suite être réutilisé par la communauté scientifique.

Face à ce constat, l'interface d'annotation et de post-correction Pyrrha a été développée dans le cadre de la cellule humanité numérique de l'École nationale des Chartes (Clérice, Pilla & Camps, 2018) pour proposer une interface simple d'utilisation qui permette de contrôler tout le cycle d'annotation de son corpus. On peut alors lemmatiser un texte en ancien français ou en latin, contrôler la qualité des données produites, voire les corriger si nécessaire grâce à l'interface de post-correction, et enfin les télécharger en CSV ou en XML-TEI pour leur exploitation et leur conservation.

2. Lemmatisation d'un corpus

L'annotation morphosyntaxique est une tâche qui occupe activement la recherche depuis l'émergence des outils numériques dans les sciences humaines, toutefois, certaines langues, comme l'ancien français, représentent un véritable défi parce qu'il est très difficile d'annoter des langues sans normalisation graphique, les lemmatiseurs traditionnels fonctionnant à partir de règles prédéfinies et d'un dictionnaire établi.

[1] On peut citer l'entreprise de lemmatisation, démarrée dans les années soixante et toujours en cours, des textes latins de l'équipe de recherche du LASLA (Laboratoire d'analyse statistique des langues anciennes de l'université de Liège).

2.1. Présentation de l'annotateur Pie

Pyrrha s'appuie sur l'annotateur Pie (Manjavacas, Kestemont & Clérice, 2019) qui associe à chaque mot un lemme, une nature[2] et une analyse morphosyntaxique. Afin de s'affranchir des problématiques des lemmatiseurs traditionnels, Pie a été conçu à partir d'algorithmes d'intelligence artificielle qui ne s'appuient pas sur un dictionnaire (Manjavacas, Kádár & Kestemont, 2019), ce qui le rend indépendant du langage auquel il est appliqué et rend possible le traitement des langues à forte variation graphique ou dont les règles morphosyntaxiques ne sont pas toujours suivies de manière homogène[3]. L'algorithme est capable d'apprendre la langue des textes sur lesquels il est appliqué par comparaison entre les résultats qu'il parvient à obtenir par lui-même et les résultats du corpus d'entraînement qui a été préalablement annoté à la main. Les performances de Pie s'améliorent au fur et à mesure des entraînements et plus le corpus réunit des textes nombreux et variés, plus la fiabilité des résultats de l'annotation automatique augmente. Les données d'entraînement permettent de produire un modèle d'annotation qui pourra être utilisé pour des corpus similaires. Dans le cadre de Pyrrha, deux modèles ont été générés, un pour le latin et un autre pour l'ancien français[4].

2.2. Les modèles d'annotation de Pyrrha

Le premier modèle a été établi à partir d'un corpus annoté de textes latins de plus d'un million de mots produit par le LASLA (Clérice, 2019). Ce modèle propose une série de lemmes constitués à partir du dictionnaire Forcellini (Forcellini & Furlanetto, 1965) et d'un jeu d'étiquettes morphosyntaxiques créées pour les projets du LASLA et accessible via la liste de contrôle « latin LASLA » disponible sur Pyrrha.

Le deuxième modèle est un modèle pour l'ancien français (Clérice, 2019). Les lemmes ont été établis à partir du dictionnaire Tobler-Lommatzsch (Tobler & Lommatzsch, 1952) et adaptés dans les cas où les entrées n'étaient pas homogènes[5]. L'étiquetage morphosyntaxique se subdivise en deux champs *POS* pour la nature du mot et *Morph* pour indiquer : le cas, le genre, le nombre, le degré de l'adverbe ou de l'adjectif si nécessaire ou bien le mode, le temps, la personne et le nombre. Le nommage et la constitution de ces catégories s'appuient sur le référentiel *Cattex 2009* développé par l'équipe de la base de français médiévale à Lyon (Guillot, Prévost & Lavrentiev, 2013).

[2] POS (Part Of Speech).

[3] L'explication présente le fonctionnement de manière extrêmement simplifiée (cf. Manjavacas, Kádár & Kestemont, 2019).

[4] D'autres modèles pourraient être ajoutés si des chercheurs contribuent à constituer des données d'entraînement nécessaires à la constitution d'un modèle.

[5] Voir la documentation : https://github.com/Jean-Baptiste-Camps/Geste/wiki, (consulté le 28 mai 2019).

Form	Lemma	POS	Morph
conmence	comencier	VERcjg	MODE=ind|TEMPS=pst|PERS.=3|NOMB.=s

Table 1. Exemple d'annotation verbale

Form	Lemma	POS	Morph
signeur	seignor	NOMcom	NOMB.=s|GENRE=m|CAS=r

Table 2. Exemple d'annotation nominale

Le modèle pour l'ancien français a été entraîné sur un panel relativement large à partir de corpus constitués en grande partie à l'École nationale des chartes. Il comprend 100 000 mots issus du corpus « *Geste* » (Camps, Cochet, Ing & Albarran, 2019), 100 000 tokens[6] du *Corpus Juris Civilis* annotés par F. Duval et L. Ing, 200 000 tokens issus du corpus de Chrétien de Troyes fourni par P. Kunstmann, et alignés avec nos référentiels, 15 000 tokens du *Lancelot en prose* annotés par L. Ing et enfin 45 000 tokens issus du corpus hagiographique de Wauchier de Denain annoté par mes soins. Des corpus en occitan ont également été ajoutés comme le corpus Monferrand (30 000 tokens) annoté par J.-B. Camps, G. Couffignal et M. Mazars et le corpus Flamenca (20 000 tokens), d'après l'édition de P. Meyer et annoté par O. Scrivner. Malheureusement si, aujourd'hui l'intégralité du corpus comporte les informations au niveau du lemme et des POS, ce qui nous permet d'atteindre un taux de fiabilité de 95 %, seuls 30 % du corpus (environ 150 000 tokens) possèdent des informations morphosyntaxiques. Toutefois, nous arrivons à un taux de fiabilité avoisinant les 93 %.

Si les deux modèles proposés dans l'interface Pyrrha ne sont pas pertinents pour un projet ou demandent à être enrichis, l'utilisateur est libre d'utiliser ses propres lemmes et étiquettes de morphosyntaxe, et même de proposer ses propres modèles entraînés après demande aux administrateurs. Il est également possible d'importer et de soumettre un corpus déjà annoté au format TSV[7] comportant, associées à chaque mot les rubriques suivantes : *Form*, *Lemma*, *POS* et *Morph*.

2.3. Lemmatiser son corpus à l'aide de Pyrrha

Pyrrha offre à ses utilisateurs un accès aisé, sans ligne de commande, au lemmatiseur-annotateur Pie grâce à une interface en ligne[8] qui permet d'importer au format texte un corpus sans annotation. La seule condition préalable est de s'enregistrer avec une adresse mail pour créer un compte utilisateur qui permet de stocker son corpus et donc d'y avoir accès à chaque connexion au service.

Pour créer un nouveau corpus, il faut importer sur la page de création de corpus son texte via un simple copier-coller dans la première case de la section *data*. Grâce à la section *metadata*, on peut nommer le corpus et définir le nombre de mots à afficher avant et après le terme traité pour le contexte.

Grâce à la fonction de lemmatisation située en dessous de la case texte de la section data, on peut lemmatiser à l'aide de l'un des deux modèles le texte à

[6] Un *token* est un élément du texte obtenu suite à la tokenisation qui peut aussi bien être un mot qu'un signe de ponctuation.
[7] Format texte où les différents champs sont séparés par des tabulations.
[8] https://dh.chartes.psl.eu/pyrrha.

importer. Enfin, on peut ajouter des listes de contrôle[9] en choisissant parmi celles proposées ou bien en important la sienne[10]. Ces listes viendront aider à l'annotation en répertoriant les annotations qui dérogent aux valeurs de contrôle, mais aussi en proposant une aide à l'annotation depuis l'interface de post-correction grâce à un système d'autocomplétion. Une fois le corpus lemmatisé et les listes choisies, le corpus doit être enregistré afin de pouvoir utiliser l'interface de post-correction.

3. Correction et export des données

Même si le taux de réussite de Pie est proche des 95 % pour l'ancien français, il peut être intéressant pour le linguiste d'améliorer encore son corpus. Grâce à Pyrrha, une relecture peut être opérée via une interface de post-correction qui permet d'annoter aisément et plus rapidement un texte.

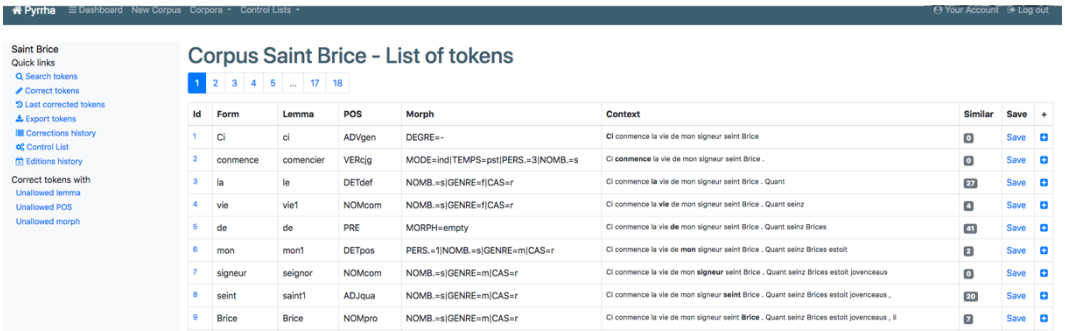

Figure 1. Interface de correction, capture d'écran
(source : Clérice, Pilla & Camps, 2018)

3.1. Fonctionnalités de base : relecture et édition des corrections

L'accès à l'interface de post-correction se fait via l'onglet *corpora* pour sélectionner le texte à traiter. L'interface affiche un tableau avec neuf catégories différentes.

1) *Id* : numéro attribué à chaque token pour l'identifier.
2) *Form* : terme tel qu'il apparaît dans le texte.
3) *Lemma* : lemme attribué à chaque terme permettant ainsi de l'associer à une forme normalisée.
4) *POS* : nature du mot.
5) *Morph* : annotation morphosyntaxique.
6) *Context* : contexte textuel.
7) *Similar* : nombre de termes dans une situation comparable.
8) *Save* : sauvegarde des modifications opérées sur l'annotation.
9) « + » : accès vers les options de modification du token : correction, suppression, ajout.

Il est possible d'intervenir directement dans les catégories *Lemma*, *POS* et *Morph* pour en changer le texte. On peut ainsi corriger la valeur d'un lemme à condition

[9] La liste standard pour l'ancien français est « Ancien Français – École des Chartes », pour le latin : « LASLA latin ».
[10] Pour créer ses propres listes de contrôle, il suffit de respecter les champs suivants : *Form*, *Lemma*, *POS* et *Morph*.

que la nouvelle valeur existe dans la liste de contrôle. L'annotation morphosyntaxique peut également être directement modifiée si les étiquettes utilisées correspondent à celles définies dans la liste de contrôle des éléments de morphosyntaxe. Une fois la modification opérée, elle peut être sauvegardée grâce au *Save* (8) présent en fin de ligne. Si l'une des catégories comporte des informations divergeant de celles des référentiels, une coloration rouge apparaît et la sauvegarde est empêchée.

Deux catégories ont été ajoutées pour simplifier les corrections. La première, *Similiar* (7) donne le nombre de cas similaires du corpus et la possibilité d'effectuer des corrections en chaîne grâce à un lien cliquable qui apparaît à la sauvegarde des modifications. La catégorie « + » (9) donne accès à trois opérations différentes : la modification d'une forme fautive, la suppression du token ou l'addition d'un nouveau token dans le cas où le texte importé serait fautif. Ces options sont à utiliser avec beaucoup de précautions au risque de ne plus pouvoir aligner le texte importé avec le texte résultant de l'annotation dans Pyrrha. Si le texte est exporté par la suite en XML-TEI, la liste des opérations effectuées sur les formes du texte sera répertoriée au début du fichier.

3.2. Fonctionnalités avancées : corrections en fonction de filtres de recherche

Pyrrha possède des fonctionnalités d'exploration du corpus accessibles via la liste de liens dans la barre de navigation de gauche.

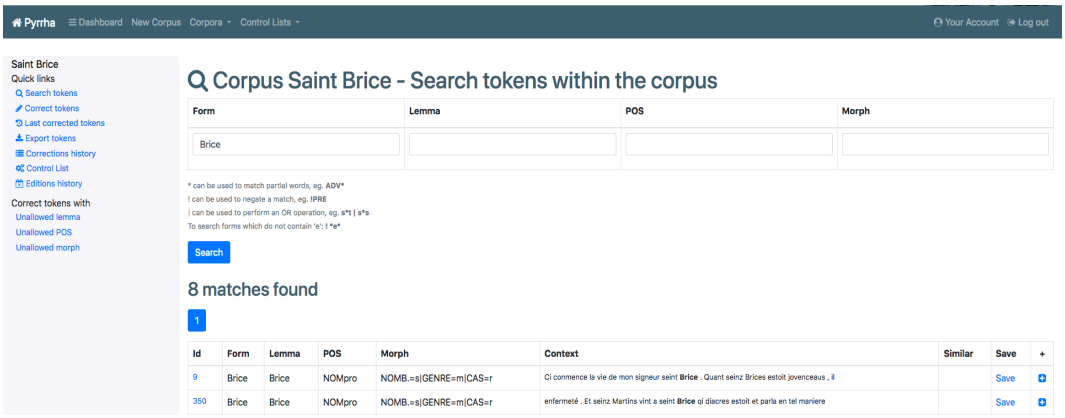

Figure 2. Interface de requêtes, capture d'écran
(source : Clérice, Pilla & Camps, 2018)

Il est possible de chercher une forme (*search token)* en particulier pour effectuer des corrections en série ou bien vérifier que l'annotation est homogène. La recherche peut porter sur le lemme, la nature et la morphologie.

On peut ne traiter que les données pour lesquelles les catégories *Lemma*, *POS* et *Morph*[11] possèdent des valeurs non valides afin d'effectuer un « nettoyage » rapide de son corpus grâce aux fonctionnalités proposées par la catégorie « *correct token with* ».

Afin d'assurer de la qualité des corrections apportées, un accès aux modifications permet à tout moment d'avoir un historique des actions effectuées grâce au lien *corrections history* avec en jaune l'ancienne version et en bleu la

[11] Dans l'ordre : Unallowed lemma, Unallowed POS et Unallowed morph.

nouvelle version. Les tokens sont classés dans l'ordre chronologique des changements, de la modification la plus ancienne à la plus récente.

Enfin, les listes de contrôle (*control list)* sont accessibles et ordonnées selon les trois catégories suivantes : *Lemma*, *POS*, *Morphologies*. Si les listes officielles sont soumises à modération avant de pouvoir être modifiées, on peut utiliser ses propres listes et les modifier en fonction de ses besoins. Les listes sont des fichiers CSV ou TSV basiques. Pour la liste des lemmes, chaque valeur est séparée par un retour à la ligne, pour les POS par une virgule. La liste des étiquettes morphosyntaxiques contient une colonne pour le label et une colonne de description de la signification du label (séparée par une tabulation). Chaque nouvelle étiquette occupe une nouvelle ligne.

3.3. Export des données

À l'issue des corrections, les données peuvent être intégralement sauvegardées et exportées dans un fichier CSV au format Pie ou en XML-TEI pour être interrogées. Le format CSV se prête aisément à des analyses statistiques, tandis que le fichier XML-TEI est davantage un fichier de conservation pour assurer la pérennité de l'annotation morphosyntaxique et à terme son partage. Le fichier TEI est très simple. Il possède un teiHeader vide qu'il conviendra de compléter. Le choix d'encodage est le suivant :

```
<w xml:id="t1" n="1" lemma="ci" type="POS=ADVgen|DEGRE=-" >Ci</w>.
```

Chaque token est englobé dans une balise <w> et toute l'annotation est contenue dans les attributs.

1) @*xml:id* pour l'identifiant du mot qui correspond à l'identifiant du mot dans l'interface Pyrrha, soit son numéro, précédé de « t ».
2) @*n* correspond à la position du mot dans le texte[12].
3) @*lemma* correspond au lemme.
4) @*type* correspond à la fois à la catégorie POS et à la catégorie Morph. Chaque subdivision de cette catégorie est séparée par « | »

4. Exploitation des données : deux cas pratiques

Les exemples proposés sont issus des Vies de saint Martin, saint Brice, saint Gilles, saint Jérôme et saint Alexis, ainsi que des *Dialogues sur les vertus de saint Martin*[13] de Wauchier de Denain. L'annotation a été générée à partir d'une lemmatisation automatique par Pie entraîné sur le modèle pour l'ancien français de l'École nationale des chartes qui était alors en cours de développement. Les corrections ont été opérées au moyen de l'interface Pyrrha. L'enjeu sera d'étudier certains

[12] Attention, la position correspond à la position du terme à l'export, elle peut être différente de celle de l'import en cas d'addition ou de suppression de tokens. Voir le modèle de données : https://github.com/hipster-philology/pyrrha/blob/dev/app/templates/tei/geste.xml. L'intégralité des modifications opérées sur les tokens est consignée dans le préambule du XML-TEI.

[13] Ces textes sont issus de mon corpus de thèse : Ariane Pinche, Édition nativement numérique du recueil hagiographique « Li Seint Confessor » de Wauchier de Denain d'après le manuscrit 412 de la Bibliothèque nationale de France, dir. C. Pierreville et B. Bureau, thèse en cours de rédaction. Les *Seint Confessor* de Wauchier de Denain sont un recueil de neuf textes composé au début de XIIIe siècle dans le contexte de la cour de Flandre pour Philippe de Namur (1174-1212), comte de Hainaut et de Namur.

traits de la scripta picarde, choisis parmi les traits caractéristiques décrits dans la *Petite Grammaire de l'ancien Picard* de Charles Théodore Gossen (Gossen, 1951), dont nous avons confronté les résultats avec ceux de l'étude d'A. Dees dans l'*Atlas des formes linguistiques des textes littéraires de l'ancien français* (Dees, Dekker, Huber & Van Reenen-Stein, 1987).

4.1. Réalisation de yod + ATA > pic. *-ie* (franc. *-iée*)

En Picard, la triphtongue descendante « -iee » s'est réduite en « -ie » (Gossen, 1951, p. 41). Ce trait est le trait dialectal qui connaît le moins d'exceptions dans les corpus et apparaît essentiellement dans la terminaison des participes passés au féminin. Il peut donc servir d'indicateur fiable pour déterminer l'appartenance dialectale d'un texte. En outre, ce phénomène touchant une catégorie de mots restreinte, il est aisé de déterminer les éléments à interroger. Nous avons limité notre étude au cas des participes féminins singuliers au cas sujet afin de s'assurer de l'accord.

À partir d'un export en XML-TEI de l'annotation[14], nous avons pu déterminer que le texte possède 169 participes passés au féminin employés avec l'auxiliaire être, dont 32 formes se terminant en « -ie » telles que « trenchie »[15] , tandis qu'une seule forme se termine par « -iee », à savoir la forme « sechiee ». Parmi toutes les formes en « -ie », seules les formes de participes passés du premier groupe sont touchées par le phénomène de réduction de la triphtongue, ainsi nous avons affiné la requête en ne sélectionnant que les verbes de ce groupe[16]. Sur les 32 participes passés en « -ie », tandis que 15 formes sont le résultat d'un radical en « -i » suivi du « -e » flexionnel, 17 termes présentent bien le picardisme à savoir :

> acouchie, aidie, amenuisie, apareillie, apeisie, apesie, aprochie, baillie, enforcie, esmaie, essaucie, mollie, negie, noncie, trenchie.

Ainsi, l'accord des participes féminins est respecté dans le texte et, pour les verbes concernés, dans 95 % des cas, la règle de simplification picarde de la triphtongue est effective.

4.2. Alternance graphique à l'initiale de <c> et <ch>

En ancien picard, C + e/i en position forte aboutit au son [ʃ] (Gossen, 1951 : 71) qui peut se noter à la fois <c> et <ch>. La proportion de la répartition des formes peut varier en fonction des régions (Dees *et al.*, 1987). De même, alors que C + A à l'initiale aboutit au son [k], la scripta picarde laisse parfois apparaître des variations en graphique en <ch> (Gossen, 1951 : 75-78). Nous avons donc essayé d'étudier ce phénomène complexe d'alternance sur notre corpus[17] afin de voir si

[14] Nous avons interrogé le corpus grâce à une feuille XSLT pour extraire toutes les formes de participes passés féminins au cas sujet, puis nous avons classé les formes en trois catégories différentes : les participes se terminant en « -iee », les participes se terminant en « -ie », et les autres.

[15] Liste de formes apparaissant dans le corpus : *acouchie, aemplie, aidie, amenuisie, apareillie, apeisie, apesie, aprochie, baillie, convertie, deguerpie, departie, enforcie, esmaie, essaucie, establie, florie, garantie, garie, guerie, mollie, negie, noncie, raemplie, raverdie, trenchie.*

[16] Le xpath comprend une expression régulière qui permet de trier en fonction de la terminaison du lemme en *-er* ou non.

[17] Nous avons interrogé le corpus grâce à une feuille XSLT pour comparer les graphies initiales des lemmes en *ce*, *ci*, *ca* ou *che*, *chi* ou *cha* et la graphie du terme tel qu'il apparaît dans le corpus afin de déterminer si une variation apparaissait pour certains termes et le nombre de fois où le phénomène pouvait être observé.

une région d'origine plus précise pouvait se dégager. L'étude se limite, ici, au phénomène en position initiale.

Sur les 1 432 fois où un terme possède un lemme avec C + e/i à l'initiale, seuls trois cas de graphie divergente en <ch> se distinguent : *chainte* (1) pour *ceindre*, *cheinture* (2) pour *ceinture.* Ces faibles statistiques invitent à situer le texte dans le nord de la Picardie : Nord, Hainaut, mais également en Wallonie (Dees *et al.*, 1987 : 4) ce qui est cohérent avec le corpus. Toutefois les occurrences en <ch> méritent d'être observées plus précisément et pourraient être moins marginales que les chiffres ne semblent le montrer, car toutes les occurrences des démonstratifs *ce/cel/cest* qui représentent tout de même près de 85 % des termes[18] sont graphiées en <c>, ce qui peut relever d'une habitude scribale sur un terme d'origine populaire extrêmement courant et dont la graphie se serait complètement normalisée. Autre fait intéressant, l'ensemble de ces graphies en <ch> n'apparaît que dans la *Vie de saint Alexis*. Il est difficile d'aboutir à des hypothèses fermes sur le corpus suite à cette enquête, mais cela peut nous amener à nous demander pourquoi la *Vie de saint Alexis* est différente du reste du corpus. Fait-elle bien partie du recueil ? Wauchier de Denain a-t-il été influencé par un substrat différent de celui des autres Vies pour sa translation ? Il faudrait poursuivre l'étude sur d'autres phénomènes linguistiques pour évaluer si ce phénomène est marginal ou non.

Si on regarde à l'inverse les termes qu'on attendrait avec la graphie <ch> à l'initiale, le texte ne possède aucune marque dialectale sur les 98 occurrences avec un vocalisme en « e ». L'alternance graphique en <c>/<ch> devant <i> n'apparaît que dans une unique occurrence de *cier* pour *chier*[19].

Pour C + A, la graphie en <ch> apparaît 4 fois sur 600 termes dont le lemme commence par <ca>, à savoir *chartage* (1) pour « Cartage », *Chaton* (2) pour « Caton », *chariole* (1) pour « cariole ». À l'inverse, seule une occurrence de *cascun*[20] pour « chacun » est à observer, ce qui nous éloigne des pourcentages attendus pour le terme dans le Hainaut ou le nord de la France d'après l'atlas d'A. Dees (Dees *et al.*, 1987 : 48) qui suggère au contraire une très grande domination de la graphie en <c>[21]. Enfin, le corpus présente également sept occurrences de *ceaille* pour « chaille[22] », montrant bien que la graphie <c> marque le son [ʃ].

Selon C. T. Gossen la scripta picarde présente le système suivant : c/ch + e, i = [ʃ], exceptionnellement [k], c + a = [k] et ch + a = [ʃ], exceptionnellement [k]. Cette théorie semble fonctionner sur l'ensemble de notre corpus. Toutefois cette courte étude ne nous ne permet pas encore de préciser à quel sous-ensemble de la scripta nos textes pourraient appartenir, car les marques dialectales sont trop rares pour ce phénomène et laissent penser que la langue littéraire est relativement normée et ne laisse apparaître que rarement des traits dialectaux.

[18] 1 195 occurrences.
[19] Sur 176 termes dont le lemme commence par « ci » et 33 termes dont le lemme commence par « chi ».
[20] Sur 28 occurrences du terme dans le corpus.
[21] L'observation est toutefois à nuancer par le fait que seuls trois textes ont été utilisés pour les statistiques sur ce phénomène dans l'atlas.
[22] Sur 183 termes dont le lemme commence par « cha ».

5. Conclusion

Nous espérons que ce court exposé permettra de montrer les avantages d'une annotation linguistique à travers ces quelques exemples et comment l'interface Pyrrha peut réellement faciliter la génération de l'annotation, mais aussi assurer le contrôle de la lemmatisation et des informations morphosyntaxiques pour permettre au chercheur de contrôler l'intégralité de la chaîne de génération de ses données. En outre, l'annotation, certes chronophage, mais automatisée en partie et vérifiée manuellement assure la qualité des informations et leur pérennité grâce à une sauvegarde en XML-TEI. L'étiquetage linguistique fin permet aussi de faire des analyses précises et variées sur un corpus donné qui ne relèvent plus de l'extraction d'un phénomène hors contexte, mais qui prennent en compte l'ensemble des occurrences. Enfin, la production en masse de tels corpus permettrait à terme de mettre en place des modèles de lemmatisation de plus en plus performants, mais aussi, un jour peut-être, de mettre à jour les grammaires et les atlas grâce à des études statistiques plus fiables à partir de masses de données suffisamment importantes pour que les phénomènes marginaux ne viennent plus fausser les résultats.

Bibliographie

Camps, J.-B., Cochet, A., Ing, L. & Albarran, E. (2019). *Jean-Baptiste-Camps/Geste : Geste : un corpus de chansons de geste, 2016-…*
Url : https://doi.org/10.5281/zenodo.2630574.

Clérice, T. (2019). *Deucalion Latin Lemmatizer*.
Url : https://doi.org/10.5281/zenodo.2707476.

Clérice, T., Pilla, J. & Camps, J.-B. (2018). *hipster-philology/pyrrha: 1.0.1*.
Url : https://doi.org/10.5281/zenodo.2325428.

Clérice, T. (2019). *chartes/deucalion-model-af: 0.2.0*.
Url : https://doi.org/10.5281/zenodo.3237455.

Dees, A., Dekker, M., Huber, O. & Van Reenen-Stein, K. (1987). *Atlas des formes linguistiques des textes littéraires de l'ancien français*, De Gruyter : Berlin (reprint 2014).

Forcellini, E., Furlanetto, G., Corradini, E. & Perin, J. (1965). *Lexicon totius latinitatis*. Gregoriana : Padoue.

Gossen, C. T. (1951). *Petite grammaire de l'ancien picard : phonétique, morphologie, syntaxe, anthologie et glossaire*. C. Klincksieck : Paris.

Guillot, C., Prévost, S., & Lavrentiev, A. (2013). *Manuel de référence du jeu Cattex09*.
Url : http://bfm.ens-lyon.fr/IMG/pdf/Cattex2009_manuel_2.0.pdf.

Manjavacas, E., Kádár, Á. & Kestemont, M. (2019). « Improving Lemmatization of Non-Standard Languages with Joint Learning ». In : *Proceddings of NAACL-HLT 2019*, Minneapolis : Association for Computational Linguistics, p. 1493-1503.

Manjavacas, E., Kestemont, M. & Clérice, T. (2019). *emanjavacas/pie v0.2.3*.
Url : https://doi.org/10.5281/zenodo.1637878.

Mellet, S. (2002). « La lemmatisation et l'encodage grammatical permettent-ils de reconnaître l'auteur d'un texte ? ». *Médiévales*, 21(42), p. 13-26.

Tobler, A. & Lommatzsch, E. (1952). *Altfranzösisches Wörterbuch*. E. Steiner : Wiesbaden.

Le PAM, un Programme d'Analyse Métrique pour le français médiéval

Enzo POGGIO* & Timothée PREMAT**
*Développeur pour les SHS, chercheur indépendant.
**Université Paris 8 (SFL UMR 7023)

Résumé

Les corpus de français médiéval existant ne proposent pas d'étiquetage des propriétés prosodiques et métriques des textes. Un tel étiquetage est pourtant nécessaire, non seulement pour celui qui étudie la métrique, mais aussi pour le phonologue. Avec le PAM, nous proposons un script Python capable de segmenter un texte en syllabes et d'associer à chaque syllabe un statut de proéminence prosodique à partir duquel on doit pouvoir déduire le gabarit métrique sous-jacent. Cet article détaille certains points du fonctionnement général du programme, toujours en cours de développement.

Mots-clefs : Linguistique de corpus ; annotation métrique et prosodique ; syllabation ; français médiéval.

DOI : 10.5281/zenodo.3464477

1. Introduction

Cet article présente les principes du fonctionnement général du PAM, le Programme d'Analyse Métrique. Le PAM est un programme d'annotation prosodique qui permet la déduction de la structure métrique du texte à partir de ses propriétés prosodiques. Il prend la forme d'un script Python exécutable dans une console, et propose différentes fonctions d'export des données annotées.

Après avoir brièvement exposé le panorama des pratiques numériques dans lequel s'inscrit le PAM (section 1), nous abordons les propriétés prosodiques qu'annote le PAM (2), la façon dont celles-ci peuvent être transposées en propriétés métriques (3), et enfin les perspectives dans le développement à venir du programme (4).

1.1. Panorama de la philologie numérique historique française

Depuis le début des années 80, des corpus informatiques de textes en français médiéval ont commencé à voir le jour. Les buts de ces corpus diffèrent : certains visent principalement à l'étude de la répartition dialectale des formes graphiques (Dees, 1980), d'autres s'intéressent à l'étude de la syntaxe diachronique (c'est le cas des fondements de la *Base de Français Médiéval* (BFM), dirigée à l'origine par Ch. Marcello-Nizia et aujourd'hui par C. Guillot-Barbance) ou à la constitution de bases de données pour des dictionnaires (*Dictionnaire du Moyen Français* (DMF), *Anglo-Norman Dictionnary* (AND)). Aujourd'hui, ces corpus tendent à utiliser les normes d'encodage de la *Text Encoding Initiative* (TEI), ce qui permet une interopérabilité optimale. Ainsi, le second corpus d'A. Dees (1987) a été ré-encodé aux normes de la TEI puis versé dans le *Nouveau Corpus d'Amsterdam* (NCA) dirigé par P. Kunstmann et A. Stein. Les outils d'interrogation des corpus sont eux aussi en voie de standardisation ; ainsi la dernière version du NCA est devenue

interrogeable dans la plateforme TXM (Heiden *et al.*, 2010), qui est aussi le moteur de consultation et d'interrogation privilégié par la BFM.

Aucun de ces projets ne vise à annoter l'ensemble des propriétés d'un texte, tâche qui confinerait à l'impossible. Ainsi, l'étiquetage principalement morphosyntaxique de la BFM correspond à l'orientation syntaxique de son origine, aujourd'hui renforcée par l'étiquetage en arborescences syntaxiques du sous-corpus externe *Syntactic Reference Corpus of Medieval French* (SRCMF) dirigé par S. Prévost et A. Stein. D'autres types d'annotations ont été ajoutés à la BFM par la suite : certains textes sont également lemmatisés, et les passages au discours direct sont également annotés. La tendance générale semble être à l'annotation des propriétés du discours et des modules 'élevés' de la grammaire. Sont donc naturellement privilégiées des approches lexicologiques, syntactiques et morphosyntaxiques.

Les chercheurs qui, au contraire, travaillent sur les modules 'bas' de la grammaire, ceux qui traitent des 'formes sonores' (prosodie, métrique, phonologie et phonétique) ne disposent pas d'outils idoines pour mener leurs recherches. Ils se retrouvent contraints d'utiliser des requêtes portant sur des graphèmes, ne pouvant que marginalement s'aider des propriétés encodées. Pourtant, les *guidelines* de la TEI proposent des outils pour annoter précisément ces propriétés des 'formes sonores'. On peut par exemple aller jusqu'à encoder l'ensemble des traits phonologiques des segments d'un texte grâce aux balises de *feature structure*[1]. Cependant, à notre connaissance, un tel encodage n'est mis en place par aucun corpus historique français. L'importante variation graphique du français médiéval et l'incertitude constitutive de la phonologie diachronique rendraient de toute manière une telle approche périlleuse. Enfin, il semble que l'objet encodé soit plus souvent appréhendé comme un texte écrit que comme un texte *oralisable*, transcription d'un texte oral et/ou destiné à une communication orale. En ce sens, il est naturel que les annotations se concentrent sur des aspects qui appartiennent autant à la linguistique textuelle qu'à la linguistique en général, délaissant par-là l'étiquetage des propriétés 'sonores'[2].

1.2. Annotation métrique et syllabique : projets existants

Dans le domaine de la métrique, on retrouve la même orientation, c'est-à-dire que les propriétés macrostructurales des textes sont régulièrement annotées. Par exemple, les textes versifiés de la BFM sont encodés en strophes et en vers. Un texte, les *Chroniques concernant la guerre d'Écosse* de Jordan Fantosme (éd. 1981), pousse l'annotation jusqu'à l'étiquetage des césures. Tous les autres textes, en revanche, n'annotent aucune propriété inférieure à l'échelle du vers : ni les césures ni les syllabes ne sont annotées.

Sans même parler des besoins des chercheurs travaillant sur les formes métriques, une telle annotation serait pourtant précieuse pour le phonologue. Il en va ainsi, par exemple, de l'analyse des rimes, tâche primordiale pour la phonologie diachronique. En l'absence d'une démarcation des syllabes à l'intérieur du texte, le phonologue se retrouve contraint d'extraire le dernier mot précédant la balise de fin de vers, ce qui complique considérablement le travail des données. Et encore, une telle interrogation n'est pas toujours possible ; par exemple, sur le corpus de la BFM, il n'est par défaut pas possible d'interroger la position d'un mot

[1] https://tei-c.org/release/doc/tei-p5-doc/en/html/FS.html
[2] L'on pourrait objecter qu'il existe pourtant une prosodie de l'écrit (Martin, 2011).

dans le vers[3]. Le phonologue ne dispose pas non plus d'outils lui permettant de différencier la syllabe tonique de la syllabe posttonique dans le cas d'une rime féminine.

Une des conséquences de cette focalisation de l'encodage sur les propriétés macrostructurales est qu'elle pousse les chercheurs qui travaillent sur les propriétés microstructurales à développer leurs propres outils. Ainsi, T. Rainsford, pour sa thèse de doctorat portant sur l'évolution entre accent de mot et accent de groupe en français médiéval (Rainsford, 2011b), a choisi de développer son propre outil lui permettant, sur 87 extraits d'environ 3000 mots chacun, de repérer la position des syllabes accentuables et des frontières de groupes prosodiques et métriques ainsi que leur interaction avec la structure syntaxique. L'auteur ne conseillant pas la réutilisation de son outil, nous avons été contraints d'en développer un nouveau. Par ailleurs, son programme tend à corriger les vers pour gérer les hiatus et diphtongues, or le PAM a été conçu, à l'origine, pour repérer les vers pouvant être mal formés, dans la cadre de la thèse de doctorat du second auteur de ce papier. Un tel mécanisme, qui simplifie considérablement l'architecture du programme et améliore son entropie, n'était donc pas approprié à cette recherche. Précisons néanmoins que certains des mécanismes du PAM sont inspirés, non directement du code de T. Rainsford, mais de la logique algorithmique qu'il a mise en place.

Un autre projet de T. Rainsford mérite d'être mentionné. En collaboration avec O. Scrivner (2014), T. Rainsford travaille à l'enrichissement d'arbres syntaxiques par des informations métriques. Le corpus qui leur sert à développer cette approche est constitué d'ancien occitan, ce qui leur impose de recourir à la lemmatisation pour placer l'accent. Nous n'avons pas adopté cette approche compte tenu de la facilité avec laquelle l'accent du français médiéval peut être prédit à partir de sa forme graphique.

Pour conclure cette introduction sur les corpus et les annotations prosodiques et métriques, mentionnons que d'autres corpus, comme le corpus historique des chansons néerlandaises *Nederlandse Liederenbank*, ont préféré exprimer les propriétés métriques dans les métadonnées du texte : la structure des strophes et des vers est détaillée sous forme d'étiquettes conventionnelles hors du texte. Une telle approche permet la mise en réseau des chansons entre elles en fonction de leur forme, mais se révèle peu appropriée pour analyser l'interface entre les énoncés linguistiques et les structures métriques. Enfin, citons le projet *Anamètre* dirigé par É. Delente et R. Renault qui annote et analyse en profondeur les propriétés 'sonores' dont nous parlions ci-dessus, mais qui le fait sur un corpus moderne du XVIIe au début du XXe siècle. Le fonctionnement général des algorithmes d'Anamètre n'est pas sans similarités avec celui utilisé par T. Rainsford ou par nous-mêmes, mais leur projet se trouve facilité par la stabilité des formes graphiques aux époques qu'ils étudient. Si nous ne nous sommes pas inspirés de ce projet pour la conception du PAM, il demeure néanmoins exemplaire pour le traitement des données textuelles versifiées dans le domaine français.

[3] Une telle interrogation n'est possible qu'à condition d'appliquer au corpus une préparation spécifique, comme celle qui a été mise en place pour le projet ANR *Oriflamms*. Une telle préparation n'étant pas livrée 'clefs en main' à l'utilisateur et ne faisant l'objet d'aucune publicité, sa diffusion demeure confidentielle.

2. Quelles propriétés annoter, et comment le faire ?

Dans cette partie, nous détaillons rapidement selon quelle approche de la prosodie et de la métrique a été développé le PAM, avant d'aborder de manière plus concrète la segmentation des unités que nous avons choisi d'utiliser et leur annotation.

2.1. Cadre théorique : métrique générative

Comme indiqué dans l'introduction, une démarche de constitution et d'annotation de corpus est toujours dépendante de l'intérêt scientifique de celles et ceux qui la mettent en place. Il faut également ajouter une autre forme de dépendance : elle dépend également des options théoriques favorisées par ces acteurs. En effet, le choix d'un appareil théorique peut orienter fondamentalement le traitement des données, que ce soit au niveau de leur préparation ou de leur interrogation. Il convient donc d'expliciter le cadre théorique dans lequel s'inscrit le PAM.

Ce cadre théorique est celui de la métrique générative ou post-générative, qui sous-entend une approche phonologique des unités linguistiques mises en jeu par la versification. Dès l'article séminal de Halle & Keyser (1966), cette approche s'inscrit en opposition à l'analyse traditionnelle, parfois qualifiée de scolaire, des spécificités formelles du discours versifié. En effet, la théorie traditionnelle est une théorie *surface-oriented*, ce qui signifie qu'elle ne postule aucune échelle de représentation sous-jacente et opère d'éventuelles transformations directement en surface. Ce faisant, la théorie traditionnelle ne peut pas répondre à l'objectif d'une grammaire générative : ses transformations ne sont pas suffisamment contraintes pour ne générer que des structures attestées (elles permettent des résultats non attestés) et, étant focalisée sur le niveau de la surface phonétique, cette approche dépend des structures accentuelles effectivement mises en place par un locuteur donné et des seules propriétés que ces structures expriment phonétiquement. Cette approche ne fait donc pas de différence claire entre *rythme* et *mètre*, alors même que la distinction entre ces deux objets n'est pas une idée nouvelle (cf. Platon, 1975 : 502c) et est capitale pour la métrique générative. Enfin, une telle approche dispose d'un pouvoir explicatif faible : elle se borne le plus souvent à décrire des faits de surface sans pouvoir rendre compte de la cohérence profonde qui unit ces faits de surface dans leur variété.

Il est à noter qu'une approche *de surface* présenterait pourtant un avantage pour le traitement automatique du langage : elle permettrait de faire une réelle annotation métrique. Au lieu de cela, ainsi que nous le verrons plus bas, le PAM ne fait pas d'annotation métrique : il produit une annotation prosodique qui permet, dans un second temps, d'effectuer une analyse métrique.

La méthode de la métrique générative consiste à observer des régularités prosodiques (nombre et disposition des frontières de groupe, nombre et disposition des syllabes en fonction de leur proéminence) pour en déduire une représentation sous-jacente appelée *gabarit* ou *patron métrique*. Chaque type de vers (octosyllabe, décasyllabe, etc.) dispose de son propre gabarit. L'association entre les unités (groupes et positions) du gabarit métrique et celles d'un vers donné est régie par une grammaire d'association, qui contient des règles ou contraintes exprimant la possibilité ou l'interdiction d'associer telle unité du gabarit à telle unité du vers.

Ces règles d'association doivent, pour le français, autoriser les configurations que sont la rime féminine et la coupe épique, la coupe lyrique et la coupe

enjambante. Elles autorisent également le libre placement des accents et frontières prosodiques à l'intérieur des vers et hémistiches, les contraintes d'association ne s'appliquant qu'en fin de constituant métrique.

2.2. Structure métrique et annotation de surface

Puisque le gabarit métrique correspond à un niveau de représentation sous-jacent, il ne peut être directement annoté dans le texte, qui est toujours un texte de surface. Le gabarit n'est pas nécessairement réalisé tel quel, les règles d'association permettant d'associer des structures prosodiques qui peuvent être en léger décalage avec la structure métrique. Étant par essence *absent* du texte, le gabarit métrique n'est immédiatement disponible ni pour le lecteur/auditeur, ni pour le chercheur, ni pour l'annotateur, que celui-ci soit humain ou machine.

C'est par une opération mentale que l'humain peut restituer le gabarit métrique qui sous-tend un ensemble de vers donnés. On suppose que les gabarits et les règles d'association sont acquis, sur le modèle de l'acquisition phonologique. Une fois cette acquisition réalisée, un sujet peut reconnaître qu'un énoncé est une instanciation bien-formée d'un gabarit métrique donné, lorsque le contexte sociolinguistique l'induit à faire cette analyse. D'autres approches, comme le travail de B. de Cornulier, insistent sur la comparaison contextuelle : pour lui, la structure métrique d'un vers ne peut être détectée que par la mise en relation avec la structure des autres vers qui l'entourent (de Cornulier, 1995 : 21). En comparant un vers avec les vers qui précèdent et suivent (ou qui se trouvent à la même position de la strophe dans le cas des strophes hétérométriques), le sujet isole les propriétés récurrentes et considère que ces propriétés récurrentes sont la réalisation des propriétés structurales du vers.

Cette option, qui s'inscrit dans le travail de B. de Cornulier dans une négation de l'existence autonome de gabarits métriques abstraits, ne correspond pas à notre ancrage théorique. Elle est en revanche très adaptée à une approche informatique et automatique des textes : il suffit d'annoter toutes les propriétés prosodiques des vers, de repérer celles qui sont récurrentes, et de considérer que ces propriétés récurrentes sont des propriétés métriques. Ensuite, le chercheur est libre de considérer que ces propriétés traduisent l'existence d'un gabarit métrique sous-jacent ou que, suivant B. de Cornulier, elles fondent le caractère métrique des vers dans une répétition de surface de certaines propriétés prosodiques.

Il est cependant à noter que, même dans l'approche de B. de Cornulier, il faut prendre en compte une certaine variation de surface comme n'étant pas structurale. Ainsi, par exemple, l'alternance entre vers masculins et féminins n'empêche pas de leur reconnaître une structure commune, dans la mesure où ce qu'il faut considérer, ce n'est pas le nombre de syllabes du vers, mais le nombre de ses syllabes *anatoniques* (i.e. situées avant ou sous l'accent de fin de vers) (Cornulier, 2010). Le même genre de mécanisme doit être mis en place pour l'alternance des coupes possibles en français médiéval. Ceci ne représente pas une difficulté majeure, dans la mesure où les principes de ces variations de surface sont bien connus pour le vers français, qu'ils soient exprimés dans l'anatomie du vers ou sous la forme des règles d'association de la métrique générative.

2.3. Unités prosodiques à annoter et étiquettes

C'est donc par l'annotation des unités prosodiques que le programme peut donner accès à la restitution des propriétés structurales et/ou gabaritiques des vers. Ces unités prosodiques sont les syllabes et les syntagmes phonologiques.

La première tâche du PAM est donc la segmentation du texte en syllabes. Cette segmentation est opérée par le module de fonction utilitaire `syllabise.py`, qui utilise la librairie NLTK (*Natural Language ToolKit* : Bird, Klein & Loper, 2009) pour le découpage en mots. Ce module est par ailleurs disponible de manière autonome sur GitHub. Le syllabeur a été pensé de manière à respecter au mieux la coupe syllabique (jonction des attaques branchantes et disjonction des groupes hétérosyllabiques, etc.), même si la syllabation d'un texte écrit est une tâche confinant à l'absurde, dans la mesure où le syllabeur traite des graphèmes alors que la syllabation humaine traite des phonèmes et que la correspondance graphèmes-phonème est loin d'être simplement bijective. Travaillant sur une base graphémique, le syllabeur respecte les élisions qui sont graphiquement notées (<qu'alors> → <qua.lors>), mais ignore celles qui ne sont pas notées (<amie et> → <a.mi.e.et>). En effet, l'élision du schwa dépend de paramètres suprasegmentaux qui ne peuvent pas être considérés à l'échelle strictement locale et interne au mot à laquelle opère le syllabeur. De plus, ces paramètres suprasegmentaux peuvent s'appliquer d'une manière différente en prose (écrite et/ou orale) et en vers, dans la mesure où la grammaire d'association est une paraphonologie (Kiparksy, 1977) apte à faire adopter à la phonologie des vers des comportements qui ne sont pas attestés en prose (orale et/ou écrite). La syllabation que propose `syllabise.py` est donc une syllabation sous-jacente, située *avant* l'application de l'élision des posttoniques.

Une fois cette syllabation effectuée, l'output du syllabeur sert d'input à l'annotateur du PAM. Celui-ci va associer une étiquette à chaque syllabe, qui représente son statut prosodique (dans l'optique de la paraphonologie du vers). Ces étiquettes sont les suivantes :

(1) Étiquettes prosodiques du PAM

a. **-1.** Est étiquetée « -1 » toute syllabe qui est non métrifiée (syllabes apocopées et rimes féminines) ;

b. **0.** Est étiquetée « 0 » toute syllabe métrifiée qui est posttonique ou qui appartient à un clitique :

pour, *-e* de *cunte* devant initiale consonantique, etc.,

c. **1.** Est étiquetée « 1 » toute syllabe métrifiée prétonique :

plai- dans *plaisir*, *par-* dans *parler*, etc.,

d. **2.** Est étiquetée « 2 » toute syllabe métrifiée qui peut porter un accent :

-sir dans *plaisir*, *-ler* dans *parler*, etc.

L'étape (1a) correspond à l'élision des schwas posttoniques présents dans la forme graphique. Elle est simplement calculée sur le principe suivant : chaque fois qu'un mot se termine par un <e> et n'appartient pas à une liste d'exception, celui-ci est étiqueté « -1 » si le mot suivant commence par une voyelle. Le traitement du statut disjonctif du <h> initial (qui n'autorise l'élision que lorsqu'il est dépourvu de représentation phonologique, ce qui n'est pas toujours le cas) est

provisoirement géré par l'édition de deux types de <h> dans le texte d'entrée. Ce processus sera amélioré ultérieurement.

Dans le terminal, la sortie de cette annotation (accompagnée par l'analyse métrique) prend l'apparence de la figure 1, ici sur les quatre premiers vers du *Chevalier au Lion* (Chrétien de Troyes, éd. 2009).

```
1 Analyse de ../../Desktop/YvainKu.txt:
1   2 |0 |  2  | 2  |0 | 1   2   -1  m:8[8]
ar·tus li boens rois dë bre·tain·gne num_l:1
0 | 0 | 1  2  0 | 0 |1   2  -1       m:8[8]
la cui pro·es·cë nos en·sei·gne      num_l:2
 0 | 0 |1   2  | 2  |0 | 1   2       m:8[8]
quë nos so·ιens preu et cor·tois     num_l:3
 2  | 2  |0 |2   0 |1  0 | 2         m:8[8]
tint cort si ri·chë co·më rois       num_l:4

1 ../../Desktop/YvainKu.txt  m:8
100.00% soit  4  /4 vers bien formés
00.00% soit  0  /4 vers mal formés
```

Figure 1. Exemple de sortie du PAM dans un terminal.

2.4. Insuffisance des formes graphiques

Le PAM étant conçu pour pouvoir fonctionner sur des corpus peu ou pas annotés en syntaxe et en partie du discours, il travaille essentiellement sur les graphèmes. Néanmoins, une telle approche est insuffisante : il faut par exemple que le PAM soit capable de différencier la finale *-ent* des troisièmes personnes du pluriel de la finale *-ment* adverbiale (← -MENTE). Dans la mesure où la finale verbale *-ent* peut être précédée d'une base se terminant par *-m-* (p. ex. *aiment*), une approche purement graphémique ne peut différencier ces deux morphèmes. Dans ce cas, le PAM utilise une liste de tous les adverbes attestés dans la partie annotée en parties du discours de la BFM : chaque fois qu'il rencontre une finale *–ment*, l'annotateur vérifie si l'ensemble de la forme graphique est attesté dans la liste des adverbes ; si oui il annote « 2 » la syllabe *-ment*, sinon il l'annote « 0 » et place un « 2 » sur la syllabe précédente. D'autres mots, en quantité négligeable, sont gérés dans une liste d'exception (p. ex. *argent*).

Dans le cas des adverbes et des finales verbales en *-ent*, la vérification manuelle des résultats du PAM n'a montré l'existence d'aucune paire homographique à même de provoquer l'identification d'un adverbe à la place d'une P6 verbale, ou vice-versa. En revanche, certaines formes graphiques d'adverbes ne sont pas représentées dans le corpus annoté en partie du discours, ce qui provoque leur identification en tant que P6 verbales. Il sera essentiel, lorsque le PAM sera à une étape plus avancée de son développement, de calculer la marge d'erreur générée par ce genre de résultats, de manière à pondérer les résultats des analyses statistiques. Notons que, lorsque l'adverbe mal identifié est situé en fin de vers, il provoque automatiquement une erreur métrique (le vers est détecté hypométrique, puisque sa phase anatonique s'arrête une syllabe avant là où est vraiment l'accent), ce qui permet de repérer très facilement ces erreurs et d'ajouter ces adverbes à la liste d'exceptions.

En revanche, la détection des mots qui, pour des critères grammaticaux, ne peuvent pas porter d'accent provoque un nombre plus élevé de mauvaises

analyses. En effet, le nombre d'homographes est ici beaucoup plus grand, notamment à cause de la très forte variation graphique médiévale. Ainsi <deus> peut être une graphie pour *deux* comme pour *Dieu*. Par ailleurs, si certains pronoms changent de forme lorsque, postposés, ils deviennent toniques (*me donne* ~ *donne-moi*), d'autres n'en changent pas (*nous donne* ~ *donne-nous*). En ce cas, en l'absence d'arborescence syntaxique, il n'est pas possible de savoir si une forme <nous> est atone ou tonique.

Le traitement de ces confusions peut s'opérer de deux manières, suivant le type de texte : sur les textes non annotés en parties du discours, une étiquette spéciale est prévue, qui permettra à l'analyste de repérer tous les vers contenant des occurrences incertaines soit pour les annoter lui-même, soit pour les exclure tout simplement de son corpus et ne pas biaiser ses analyses quantitatives. Sur les textes annotés en partie du discours, le PAM pourra, pour chaque occurrence douteuse, interroger l'étiquette du XML originel, et établir son annotation prosodique à partir de cette information. Ces deux fonctionnalités ne sont pas encore présentes dans le PAM au moment où nous écrivons ces lignes, mais le seront dans des versions ultérieures du programme.

2.5. Gestion des paramètres linguistiques : les configs

Comme toute langue naturelle, la langue médiévale française est sujette aux variations habituelles de la sociolinguistique. Parmi celles-ci, deux sont pertinentes pour le PAM : la variation diachronique et la variation diatopique. En effet, le rapport entre graphèmes et phonèmes peut varier suivant les dialectes ou suivant le moment de composition du texte. Il en est ainsi, par exemple, de la graphie <-ni-> devant voyelle qui, en wallon, sert à noter l'équivalent d'un <gn> dans les autres dialectes. Si le PAM n'est pas informé de cette différence, il va syllaber le mot <companie> (« compagne ») en quatre syllabes (<com.pa.ni.e>) alors que la représentation phonologique de ce mot n'en contient que trois, le <i> formant avec le <n> un digramme consonantique (<com.pa.nie>).

Les paramètres fondamentaux de syllabation sont présents dans un fichier `constants.yaml`, accompagné par les fichiers qui servent à gérer les exceptions et les adverbes (`special_syll.yaml` et `special_type.yaml`). Ces trois fichiers sont regroupés dans une configuration (*config*) générale. Lorsqu'il fait face à un texte dont la situation diachronique ou diatopique modifie les paramètres de syllabation, l'utilisateur peut dupliquer ces trois fichiers dans un autre répertoire et les modifier à sa guise, ce qui crée une nouvelle config, spécifique à un texte ou à un ensemble de textes. Lorsqu'il exécute le PAM, si l'utilisateur ne fournit aucune information relative à la config, c'est la config par défaut qui est exécutée ; s'il veut utiliser une config spécifique, il l'indique en ajoutant `-C [chemin de la config]` à la ligne de commande.

3. De la prosodie à la métrique

L'annotation syllabique pratiquée en 2.3 et modifiée telle qu'en 2.4 permet d'établir partiellement la structure prosodique du texte, puisqu'elle permet de repérer ce qui correspond aux groupes clitiques. Ceux-ci s'étendent d'une étiquette 2 (et l'éventuel -1 qui le suit) jusqu'à la précédente étiquette 2 (et l'éventuel -1 qui le suit). Les groupes clitiques ne sont cependant pas le constituant prosodique le plus intéressant pour l'analyse métrique. Mais, en l'absence de consensus sur l'extension du syntagme phonologique en français, et face à la

difficulté de reconstruire la structure prosodique d'une langue morte sans enregistrements audio, nous suivons une démarche analogue à celle de T. Rainsford (2011a, 270-272), en analysant de *possibles syntagmes phonologiques*. Ceux-ci sont constitués au minimum d'un seul groupe clitique[4] et au maximum de plusieurs groupes clitiques, la tête du syntagme phonologique étant toujours celle du dernier groupe clitique.

Notons par ailleurs que les règles d'associations étant capables d'invoquer une paraphonologie à même de contrecarrer la phonologie du langage naturel, il n'est pas nécessaire de s'interroger précisément sur l'extension du syntagme phonologique médiéval : on considère qu'une fin de constituant métrique (hémistiche ou vers) doit être alignée avec une fin de syntagme phonologique, mais sans avancer d'hypothèse sur le fait que l'ensemble du constituant corresponde à un seul syntagme phonologique. Dans notre analyse, à laquelle répond le fonctionnement du PAM, ce sont les frontières des constituants et non l'extension de ceux-ci qui sont ciblées par la grammaire d'association.

De fait, l'annotation du PAM propose des syllabes toujours atones (« -1, 0, 1 ») et des syllabes qui, si tel syntagme phonologique est réalisé par le locuteur, doivent en porter l'accent (« 2 »). Mais, si le locuteur choisit de réaliser un syntagme phonologique plus grand et qu'il y a un autre groupe clitique sur la droite du groupe clitique considéré, alors la tête du premier groupe clitique n'est pas une tête de syntagme phonologique et ne porte pas d'accent. Cela ne revient pas non plus à faire l'hypothèse que les têtes de syntagmes phonologiques associées avec des têtes de constituant métrique sont nécessairement accentuées lors de la diction du vers, la réalisation (surface) du vers pouvant être plus ou moins éloignée de sa structure métrique (profonde) (cf. Verluyten, 1989 : 51-53).

En revanche, selon l'approche de B. de Cornulier exposée en 2.2 et implémentée dans le fonctionnement du PAM, les seules têtes de groupes clitiques à être associées à des têtes de constituant métrique sont celles qui sont régulièrement disponibles de vers en vers. Pour la tête du vers, cela est assez simple à repérer. Pour la tête de l'hémistiche, il faut considérer la possibilité que la tête soit décalée d'une position vers la gauche (coupe lyrique). Il faut aussi considérer que la posttonique d'un premier hémistiche féminin peut être extramétrique (coupe épique) ou métrifiée dans l'hémistiche suivant (coupe enjambante). Ces paramètres sont gérés par le PAM qui propose, pour chaque ligne et en fonction du mètre de référence indiqué par l'utilisateur, différentes coupes possibles, et qui réunit ces possibilités dans les statistiques globales de l'extrait analysé. (2) représente les coupes analysées sur les trois premiers vers de *Li Ver del juïse* (Anonyme, 1982).

(2) Exemple de coupes analysées par le PAM

Sanior oiez raison glorïose et saintisme	6ma
Del ciel en est la voiz de paradis la vie	6ma
Deus la tremist en terre por amendeir noz vies	6épC

En (2), les deux premiers vers ont pour seule coupe possible une coupe masculine après la sixième position (6ma). En revanche, dans le troisième vers, la seule césure possible est une césure épique devant initiale consonantique (6épC). Les césures épiques devant initiale consonantique sont aisées à repérer, dans la

[4] Nous considérons que, dans le cas où un mot phonologique n'est pas précédé de clitiques, il forme à lui seul un groupe clitique de manière à respecter le caractère strict de la hiérarchie prosodique.

mesure où, avant l'analyse des césures, elles provoquent un vers extramétrique (si l'on métrifie le *-e* de *terre*, alors le vers fait 13 syllabes métriques). Devant initiale vocalique, le PAM propose toujours deux analyses : une césure épique devant voyelle ou une césure masculine par apocope de la posttonique. Par exemple, sur le vers (3) (chanson homonyme, dans l'édition de Rosenberg, Switten & le Vot, 1998), le PAM proposerait les deux coupes suivantes :

(3) Bele Doette as fenestres se siet 4ma/4épV/6na

Des configurations comme (3) sont traditionnellement analysées comme étant des cas d'apocope de la posttonique aboutissant à des césures masculines. C'est notamment l'opération couramment utilisée par les philologues qui cherchent à corriger toutes les coupes épiques résiduelles dans le corpus des trouvères. Mais, pour cette chanson comme pour quelques autres chansons, la notation musicale accompagnant ce vers semble indiquer que le *-e* de *Doette* est bien chanté (cf. Premat, 2017), ce qui n'invalide pas tout à fait l'hypothèse traditionnelle (mais demanderait alors à ce que la paraphonologie du vers considère comme apocopée une syllabe effectivement prononcée), mais permet également de soutenir qu'en fin de constituant métrique, une posttonique peut être maintenue même devant initiale vocalique. Le fait de proposer les deux analyses donne au PAM une neutralité, celui-ci laissant à l'analyste le dernier mot. Enfin, le PAM détecte également les césures lyriques et enjambantes, et indique aussi les coupes qui, compte tenu du mètre de référence indiqué par l'utilisateur, ne sont pas acceptables, comme à la sixième position en (3) (6na, pour *not applicable*), qui tomberait entre la prétonique et la tonique de *fenestres*.

4. Perspectives

4.1. Export : la perspective du XML

Au moment où nous écrivons ces lignes, le PAM produit son analyse dans la console où l'utilisateur l'exécute. Ayant produit l'annotation ligne par ligne, il produit également un résumé donnant accès aux statistiques générales de l'extrait interrogé : combien de vers bien formés vis-à-vis du mètre de référence indiqué par l'utilisateur, combien de vers mal formés, quantité des différents types et positions de coupes. Le PAM propose également un export en table .csv ou .xlsx, permettant à l'utilisateur d'appliquer des analyses statistiques ou simplement de sauvegarder l'annotation effectuée.

Ces exports ne permettent ni l'enrichissement des analyses du PAM par les informations supplémentaires éventuellement présentes dans le fichier d'entrée (tel que l'étiquetage en parties du discours) ni leur réutilisation ou leur croisement avec ces autres données. Le format de sortie du PAM, par ailleurs, ne répond pas aux normes de standardisation des corpus actuellement en vigueur. L'objectif, qui sera poursuivi dans les années à venir, est donc de réinjecter les informations du PAM dans les fichiers XML d'origine, en suivant les consignes de la TEI-P5. Ainsi, si les responsables des corpus en question s'y accordent, ces informations seront à terme disponibles dans les fichiers XML et pourront être interrogées dans des plateformes comme TXM, tant dans une perspective d'analyse métrique que phonologique. Enfin, comme indiqué plus haut, cela permettrait, sur les textes annotés en partie du discours, d'affiner l'annotation du PAM en cas d'homographie.

4.2. Accessibilité : sortir du terminal

Enfin, il est à considérer que l'installation et l'exécution du PAM dans la console de l'ordinateur réduit son accessibilité auprès des membres de la communauté scientifique qui ne seraient pas habitués à l'usage du terminal et/ou de commandes Python. Nous envisageons en ce sens la possibilité d'une interface HTML qui libérerait l'utilisateur du terminal et qui, si elle était disponible sur serveur, le dispenserait d'avoir à télécharger et installer le programme sur sa machine.

Conclusion

L'absence d'annotation phonologique, prosodique et métrique pousse les chercheurs travaillant sur ces domaines soit à travailler sans outils informatiques idoines soit à développer les leurs. Dans cette optique, le PAM est un Programme d'Analyse Métrique qui, après avoir effectué une syllabation et une annotation prosodique du texte, propose à l'utilisateur des analyses métriques. Il travaille sur une base essentiellement graphémique, augmentée de quelques dictionnaires de formes, de manière à pouvoir fonctionner sur des textes peu ou pas annotés. Toujours en développement, le projet vise à améliorer la précision des résultats fournis, à les réinjecter dans les fichiers XML-TEI d'origine pour l'amélioration de cette précision et pour diffuser l'information annotée auprès de la communauté, et éventuellement à sortir du terminal pour améliorer son accessibilité.

Bibliographie

Anonyme (1982). *Li ver del juïse* (éd. E. Rankka). Uppsala : Almqvist och Wiksell. Publié en ligne par la Base de français médiéval (dernière révision le 10/08/2010).
Url : http://catalog.bfm-corpus.org/Juise.

Bird, S., Klein, E. & Loper, E. (2009). *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*. Sebastopol, Beijing, Cambridge : O'reilly.

Chrétien de Troyes (2009). *Chevalier au Lion ou Yvain* (éd. P. Kunstamm). Ottawa, Nancy : Laboratoire de Français Ancien, ATILF.
Publié en ligne par la Base de français médiéval (dernière révision le 06/05/2009).
Url : http://catalog.bfm-corpus.org/YvainKu.

De Cornulier, B. (1995). *Art Poëtique, Notions et problèmes de métrique*. Lyon : Presses universitaires de Lyon.

De Cornulier, B. (2012). « Si le mètre m'était compté... Sur la notion fallacieuse de mesure du vers ». In L. de Saussure, A. Borillo & M. Vuillaume (dir.), *Grammaire, Lexique, Référence, Regards sur le sens, Mélanges offerts à Georges Kleiber pour ses quarante ans de carrière*, Bern : Peter Lang, p. 355-376.

Dees, A. (1980). *Atlas des formes et des constructions des chartes françaises du 13e siècle*. Tübingen : M. Neimeyer Verlag.

Dees, A. (1987). *Atlas des formes linguistiques des textes littéraires de l'ancien français*. Tübingen : M. Neimeyer Verlag.

Fantosme, J. (1981). *Chroniques concernant la guerre d'Écosse* (éd. R. C. Johnston). Oxford : Clarendon.
Publié en ligne par la Base de français médiéval (dernière révision 23/05/2009).
Url : http://catalog.bfm-corpus.org/Fantosme.

Halle, M. & Keyser, S. J. (1966). « Chaucer and the Study of Prosody ». *College English*, 28(3), p. 187–219.

Heiden, S., Magué, J.-P. & Pincemin, B. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement ». In S. Bolasco, I. Chiari & L. Giuliano (dir.), *10th International Conference on the Statistical Analysis of Textual Data – JADT 2010*, Rome : Edizioni Universitarie di Lettere Economia Diritto, p. 1021-1032.

Kiparksy, P. (1977). « The Rhythmic Structure of English Verse ». *Linguistic Inquiry*, 8(2), p. 189–247.

Martin, P. (2011). « Ponctuation et structure prosodique ». *Langue française*, 172, p. 99–114.

Platon (1975). *Lysis Symposium Gorgias* (éd. W. R. M. Lamb). London : W. Heinemann.

Premat, T. (2017). *Conflits et synergies des patrons structuraux dans la musique des trouvères*, mémoire de master.
Url : https://scd-resnum.univ-lyon3.fr/out/memoires/langues/2017_premat_t.pdf.

Rainsford, T. (2011a). « Dividing Lines: The changing syntax and prosody of mid-line Break in medieval France octosyllabic verse ». *Transactions of the Philological Society*, 109, p. 265–283.

Rainsford, T. (2011b). *The Emergence of Group Stress in Medieval French*. Thèse de doctorat, Cambridge University, St. John's College.

Rainsford, T. & Scrivner, O. (2014). « Metrical Annotation for a verse treebank ». In V. Henrich, E. Hinrichs, D. de Kok, P. Osenova & A. Przepiórkowski (dir.), *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13)*, Tübingen : University of Tübingen, p. 149–159.

Rosenberg, S. N., Switten, M. & le Vot, G. (1998). *Songs of the Troubadours and Trouveres: An anthology of poems and melodies*. New York & London : Garland.

TEI Consortium. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, version 3.6.0, modifiée le 16/07/2019.
Url : http://www.tei-c.org/Guidelines/P5/.

Verluyten, S. P. (1989). « L'analyse de l'alexandrin, mètre ou rythme ? ». In M. Dominicy (dir.), *Le souci des apparences : Neuf études de poétique et de métrique*, Bruxelles : Éditions de l'Université de Bruxelles, p. 31-74.