

# Traitements pour l'analyse du français préclassique

Sascha Diwersy<sup>1</sup>, Achille Falaise<sup>2</sup>, Marie-Hélène Lay<sup>3</sup>, Gilles Souvay<sup>4</sup>

(1) Université de Cologne, (2) ICAR, ENS de Lyon, (3) FoReLL, Université de Poitiers, (4) ATILF-CNRS, Université de Nancy

*sascha.diwery@uni-koeln.de, achille.falaise@ens-lyon.fr, marie-helene.lay@univ-poitiers.fr, gilles.souvay@atilf.fr*

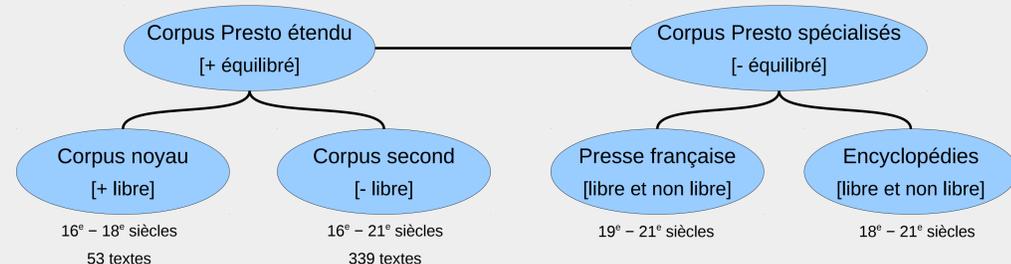
## Situation

Entre l'écrit médiéval et l'écrit moderne, les écrits en langue française de la période préclassique (ca. 1501–1650), qui correspond aux débuts de l'imprimerie, et dans une moindre mesure de la période classique (ca. 1651–1800), présentent un début de normalisation graphique, mais la graphie est encore loin d'être stabilisée.

Cet état de la langue écrite est encore peu traité, paradoxalement moins que celui de la période médiévale. À la différence de ce dernier, le français préclassique reste relativement intelligible pour un locuteur moderne ; c'est surtout la variabilité graphique qui pose problème pour un traitement automatique.

Nous présentons ici un travail d'adaptation de ressources et d'outils pour le traitement du français préclassique et classique.

## Le corpus Presto : un corpus diachronique du français



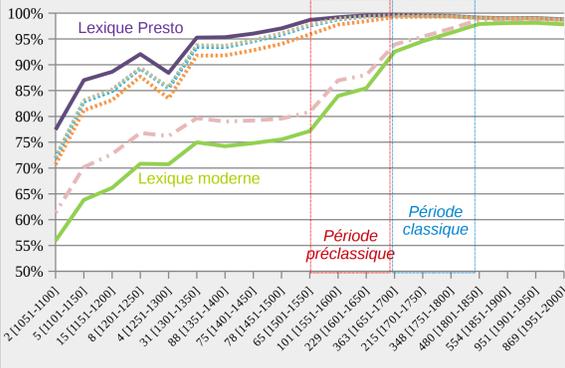
## Exemple

C'est assez dict pour ceste foys.  
 Quand sçavoir en vous s'associe,  
 Monsieur Rien, l'on vous remercie  
 Du bien qu'avons aprins de vous.  
 Bazochiens, entendez tous :  
 Je veulx en triumpant arroy  
 Eslire et faire ung nouveau roy,  
 Comme il est coustume de faire ;  
 Pourtant chacun pense a l'affaire,  
 Autant les grandz que les petitz,  
 Et faire les preparatifz ;  
 Car, ainsi comme liberalle,  
 Je tendz a monstre generalle  
 Qui, l'esté qui vient, sera faicte.  
 En honneur du triumphe et feste,  
 Ne faillez monstrez vos bons cueurs  
 Qui font de la vertu approche,  
 Tant que l'on dye par honneurs :  
 Vive l'excellente Bazoche !

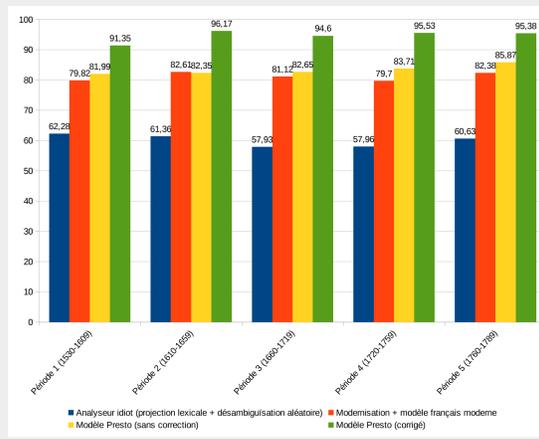
Extrait de *Sottie pour le cry de la bazoche*, Anonyme, 1549

## Évaluation

Couverture lexicale, mesurée sur le corpus Frantext, pour le lexique moderne (vert), les 3 itérations d'archaïsation (pointillés), et le lexique Presto final (violet).



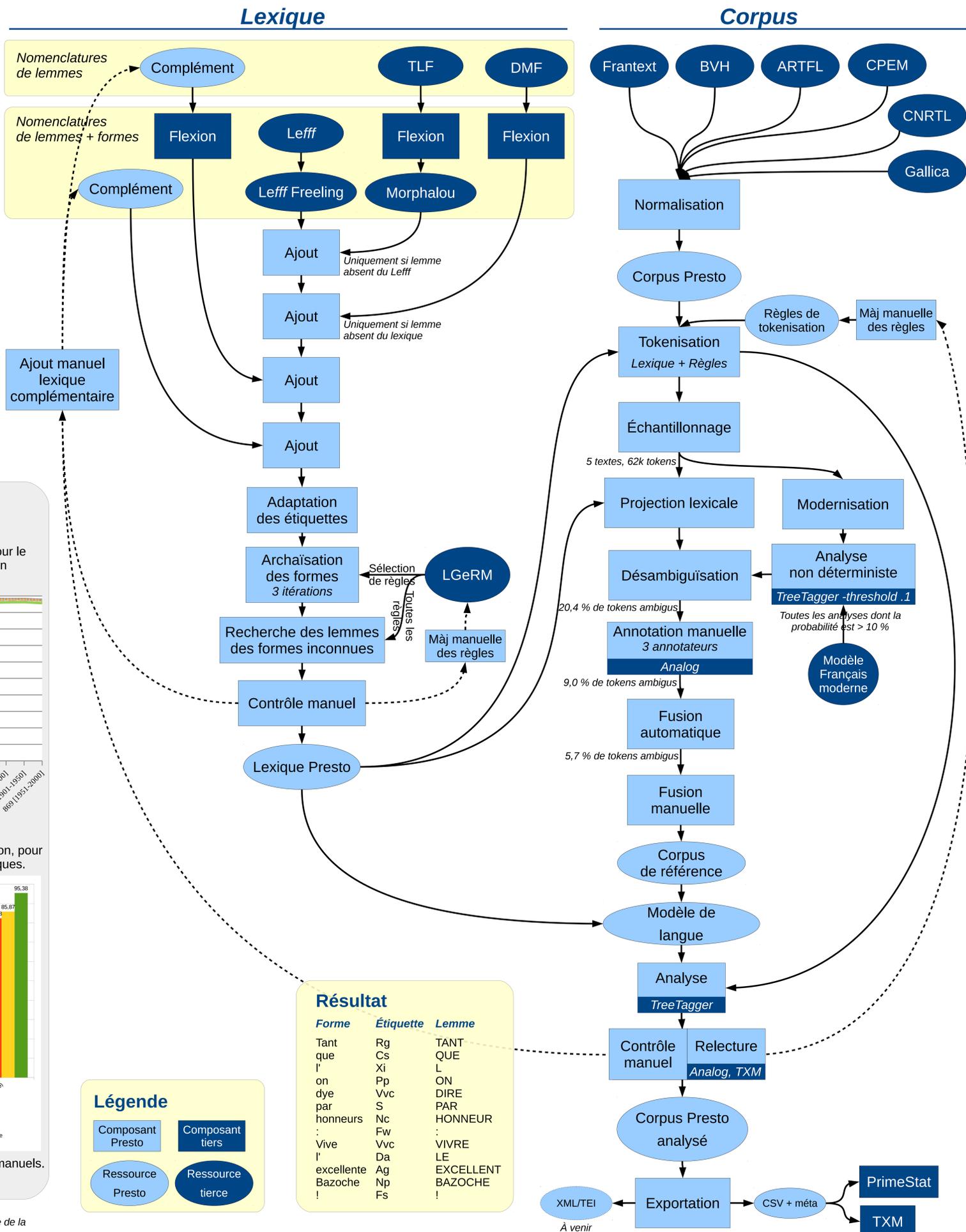
Exactitude obtenue en fonction de la méthode d'annotation, pour 5 sous-périodes des périodes préclassiques et classiques.



Seul le modèle Presto (corrigé) a bénéficié de contrôles manuels.

Ce travail est issu du projet Presto, cofinancé par l'Agence Nationale de la Recherche et la Deutsche Forschungsgemeinschaft.

## Chaîne de traitement Presto



### Résultat

Forme	Étiquette	Lemme
Tant	Rg	TANT
que	Cs	QUE
l'	Xi	L
on	Pp	ON
dye	Vvc	DIRE
par	S	PAR
honneurs	Nc	HONNEUR
:	Fw	:
Vive	Vvc	VIVRE
l'	Da	LE
excellente	Ag	EXCELLENT
Bazoche	Np	BAZOCHÉ
!	Fs	!

