



**HAL**  
open science

## Expression et position du sujet pronominal du 12e au 14e siècle : une approche quantitative et contrastive

Sophie Prévost

► **To cite this version:**

Sophie Prévost. Expression et position du sujet pronominal du 12e au 14e siècle : une approche quantitative et contrastive. 11èmes Journées Internationales d'Analyse des Données textuelles (JADT 2012), A.Dister D. Longrée et G. Purnelle Jun 2012, Liège, Belgique. pp.791-804. halshs-01423631

**HAL Id: halshs-01423631**

**<https://shs.hal.science/halshs-01423631>**

Submitted on 30 Dec 2016

**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.

# Expression et position du sujet pronominal du 12<sup>e</sup> au 14<sup>e</sup> siècle : une approche quantitative et contrastive

Sophie Prévost

Lattice-UMR 8094 (CNRS/ENS Paris) – sophie.prevost@ens.fr

## Abstract

This study is based on a corpus of texts from 12<sup>th</sup> to 14<sup>th</sup> century. It aims to provide a quantitative account of the evolution of the expression and of the position of pronominal subjects. We make a systematic distinction between 1<sup>st</sup> and 3<sup>rd</sup> persons, which allows to put forward (test of khi2) significant differences in distribution, especially as regards non expression. Besides, the coefficient of correlation proves that there is no statistic relation between the declines of inversion and non expression, whatever the person is. We conclude in insisting on the difficulty to determine general tendencies related to external characteristics of the texts (date, domain, form and dialect).

## Résumé

Fondée sur un corpus du 12<sup>e</sup> au 14<sup>e</sup> siècle, l'étude propose une approche quantitative de l'évolution de la position et de l'expression du sujet pronominal. Pour chacun des phénomènes on distingue les personnes 1 et 3, mettant ainsi au jour, à l'aide du test du khi2, des différences de distribution significatives, en particulier pour la non-expression. Le recours au coefficient de corrélation montre l'absence de liaison statistique entre le recul de l'inversion et celui de la non-expression, quelle que soit la personne. La conclusion insiste sur la difficulté à dégager des tendances nettes liées aux caractéristiques externes des textes (date, domaine, forme, dialecte).

**Mots-clés :** position et expression du sujet pronominal, français médiéval, diachronie, corrélation, khi2

## 1. Introduction

En français moderne, l'expression du sujet est presque toujours obligatoire et il occupe majoritairement une position préverbale, en particulier lorsqu'il s'agit du pronom personnel sujet (le seul considéré ici, désormais 'sujet pronominal'). La situation était fort différente en ancien français : la non-expression du sujet dominait, et la postposition était beaucoup plus fréquente. L'évolution vers le système moderne s'amorce à la fin de l'ancien français, puis s'accélère en moyen français. Cette étude vise à déterminer si les deux évolutions se sont faites parallèlement, et plus, si elles présentent une corrélation d'ordre statistique. Pour cela, on distinguera de manière systématique la 1<sup>ère</sup> et la 3<sup>e</sup> personne (*je* = P1 ; *il(s)/elle(s)* =

P3), distinction généralement écrasée dans les études sur l'évolution de la syntaxe du sujet pronominal, alors qu'il apparaît des divergences significatives et signifiantes<sup>1</sup>.

Je rappellerai tout d'abord très brièvement quelques éléments essentiels de la syntaxe du sujet pronominal en ancien français et de son évolution. Je présenterai ensuite les aspects méthodologiques de cette étude, puis l'analyse des données et les conclusions à en tirer.

## 2. Evolution de la syntaxe du sujet pronominal

Je résumerai ici de manière très schématique l'évolution de la syntaxe du sujet pronominal. L'ancien français connaît un ordre des mots beaucoup plus souple sur le plan syntaxique que le français moderne, mais il est régi par une contrainte forte, bien que non absolue, celle du verbe en seconde position, contrainte 'V2' qui le rapproche des langues germaniques. L'ancien français se caractérise par ailleurs par la fréquence de la non-expression du sujet (c'est une langue à sujet nul), propriété typique des langues romanes. Dans tous les textes, la non-expression du sujet est très fréquente, supérieure à 50% de l'ensemble des sujets, exprimés ou non. Il a souvent été proposé que le sujet non-exprimé correspondait à un pronom personnel postverbal (par exemple Foulet, 1930, Skårup, 1975, Vance 1997, Buridant, 2000), point contestable en ce qui concerne le présupposé positionnel (voir Prévost, 2010 et 2011b pour une discussion), en particulier au regard de la relative rareté de l'inversion pronominale et de la fréquence de la non-expression.

L'essor de l'ordre 'sujet pronominal-verbe' s'inscrit dans un mouvement général de fixation de l'ordre des mots. Le développement du sujet préverbal a été expliqué par la chute de la déclinaison, le recul de la contrainte du verbe en seconde position, et le passage à un principe syntaxique d'organisation de l'ordre des mots (Vennemann 1976, Combettes 1988). Pour le développement de l'expression du sujet, on a avancé l'explication de la perte de la richesse morphologique verbale (Foulet, 1930, Vance 1997). Mais une autre explication, non exclusive de la précédente, a été proposée : c'est à partir des effets de mise en relief liés à l'expression du pronom en ancien français que celle-ci se serait progressivement systématisée. Detges (2003) a ainsi suggéré que le développement de l'expression du sujet pronominal se serait fait à des fins de stratégie discursive, dans des contextes de prise de parole, et donc, prioritairement, avec la première personne. La hausse de la fréquence dans de tels contextes aurait eu un effet de dévaluation rhétorique, qui aurait conduit à une généralisation de l'emploi, et donc à l'affaiblissement des pronoms. Le mouvement se serait ensuite généralisé aux autres pronoms.

On a coutume de considérer que le recul de l'inversion et celui de l'omission du sujet se sont déroulées en partie conjointement. C'est vrai si l'on envisage une très large diachronie qui irait du 12<sup>e</sup> au 17<sup>e</sup> siècle. Il est légitime de s'interroger sur l'éventuel lien entre ces deux évolutions : l'expression accrue s'est-elle faite conjointement à celle du développement de la position préverbale ? L'une a-t-elle précédé l'autre, et l'a-t-elle du coup provoqué, ou au moins accéléré ?

---

1 Les personnes 2, 4 et 5 (*tu, nous, vous*) ont été écartées. Ces sujets pronominaux apparaissant comme quantitativement marginaux comparés à ceux des personnes 1 et 3, j'ai préféré me concentrer sur ces derniers, d'autant qu'il s'agit d'évaluer prioritairement les différences d'évolution entre le pronom désignant la personne du locuteur et les pronoms excluant l'implication du locuteur (indirectement « tu » et « vous » supposent un « je »). Outre les pronoms des personnes 1 et 3, j'ai retenu ceux de la personne 6 (groupés avec ceux de la personne 3), « ils » étant, référentiellement, un véritable pluriel de « il », alors que « nous » ne consiste pas en l'addition de plusieurs « je » (Benveniste parle de personne « amplifiée » pour « nous » et « vous »).

### 3. Méthodologie

#### 3.1. Présentation du corpus

Le corpus est composé de 14 textes qui s'étalent du début du 12<sup>e</sup> à la fin du 14<sup>e</sup> siècle. Ils ont été sélectionnés selon trois critères principaux : la date, le domaine<sup>2</sup> et la forme (vers/prose/mixte). Un quatrième critère a été pris en compte de manière secondaire : le dialecte (ou les traits dialectaux). Voici les textes finalement retenus<sup>3</sup> :

Titre (+ auteur )	Date	Dialecte	Forme	Domaine	Genre	Nbr. mots <sup>1</sup>
<i>Chanson de Roland</i>	vers 1100	anglo-normand	vers	littéraire	Epique	29 338
<i>Eneas</i>	vers 1155	Normand	vers	littéraire	Roman	34 958
<i>Tristan</i> , Beroul	entre 1165 et 1200	traits normands	vers	littéraire	Roman	27 257
<i>Ami et Amile</i>	vers 1200	non marqué	vers	littéraire	Epique	25 283
<i>Conquête de Constantinople</i> , Robert de Clari	après 1205	picard	prose	historique	chroniq.	33 994
<i>Aucassin et Nicolette</i>	1 <sup>ère</sup> m. 13 <sup>e</sup>	traits picards	mixte	littéraire	réécits brefs	10 009
<i>Miracles de Nostre Dame</i> , G. de Coinci	1218-1227	non marqué	vers	religieux	lyrique	22 500
<i>Queste del saint Graal</i>	vers 1230	non marqué	prose	littéraire	Roman	45 000
<i>Coutumes Beauvaisis</i> , P. de Beaumanoir	vers 1283	traits picards	prose	juridique	traité	22 500
<i>Mémoires ou Vie de saint Louis</i> , Joinville	entre 1305 et 1309	non marqué	prose	historique	mémoires	45 000
<i>Chroniques</i> , Froissart	entre 1369 et 1400	franco-picard	prose	historique	chronique	45 000
<i>Estoire de Griseldis</i>	1395	traits picards	vers	littéraire	Dramatique	16 243
<i>Manières de langage</i>	1396, 1399	non marqué	mixte	didactique	Manuel	20 282
<i>Quinze joies de mariage</i>	vers 1400	non marqué	prose	littéraire	Nouvelle	39 404
Total						457 736

Tableau 1 : Textes, descripteurs et taille

2 Les domaines sont ceux qui ont été proposés pour les textes de la *Base de Français Médiéval* (BFM) <<http://bfm.ens-lsh.fr>> dans le cadre du projet Corptef <<http://w7.ens-lsh.fr/corptef/spip.php?rubrique60>>

3 Les relevés ont été effectués en 2008 dans la *BFM* à l'aide du logiciel *Weblex* <<http://weblex.ens-lsh.fr/>>.

4 Le nombre de mots inclut l'ensemble des unités graphiques, signes de ponctuation compris. En raison des phénomènes de contraction 1 unité graphique peut correspondre à 2 unités linguistiques. Les textes dépassant 45 000 mots ont été réduits à 3 extraits (début, milieu, fin) de 15000 mots chacun.

Le 12<sup>e</sup> siècle est peu diversifié (prévalence du domaine littéraire, du vers, des textes anglo-normands ou normands) : cela tient largement à la moindre disponibilité de textes variés pour cette époque. De même, le 13<sup>e</sup> siècle contient un nombre proportionnellement important de textes plus ou moins picardisants : c'est une réalité historique. Par ailleurs, la fin du 12<sup>e</sup> siècle et le début du 13<sup>e</sup> siècle sont particulièrement bien représentés, ainsi que, dans une moindre mesure, la fin du 14<sup>e</sup> siècle. C'est un choix délibéré, qui a permis, sur ces créneaux chronologiques, de faire varier les autres paramètres plus systématiquement que sur les autres périodes, afin de tester, sur deux « échantillons », les effets d'une variabilité plus poussée.

### 3.2. *Les objets et les phénomènes à mesurer ; les outils de mesure*

Cette étude a pour objectif d'évaluer l'évolution de l'expression et de la position du sujet pronominal, en mettant l'accent sur les deux phénomènes qui se sont marginalisés – non expression et inversion. Il s'agit de les quantifier séparément, mais aussi de déterminer si ces évolutions sont liées entre elles. Par ailleurs, si l'hypothèse selon laquelle l'expression du sujet se serait d'abord développée avec P1, pour des raisons d'expressivité, est exacte, nous devrions observer, au moins dans les textes les plus anciens, une différence quantitative significative entre l'expression des personnes P1 et P3. Il est par ailleurs utile d'examiner si de telles différences apparaissent entre personnes 1 et 3 au regard de la position. Ce sont donc différents paramètres qui vont être croisés : expression, position, et personne.

Les fréquences respectives d'inversion et d'omission ont été comparées d'un texte à l'autre, afin de voir si une évolution le long de l'axe chronologique se dessine et si le rapport entre les deux est constant ou au contraire variable. Plus précisément, on a tenté d'établir s'il existe des convergences, voire des corrélations, d'une part entre l'évolution des personnes 1 et 3 au regard de chacun des phénomènes, et d'autre part entre expression et position au regard de chacune des personnes.

Les outils de mesure utilisés sont : les fréquences, le calcul du khi2 et l'indice de corrélation.

## 4. Analyse des données

### 4.1. *L'inversion du sujet*

Voici quelques exemples :

(1) Dist Oliver : « Or vos **oi jo** parler / Jo ne vos vei, veied vus Damnedeu ! ...» (*Roland*, v. 2003)

(2) La dame respondi : « Dieus i ait part ! » Adont **prist elle** congiet au conte de Hainnau et a la contesse, et les remerchia moult doucement de la bonne et honnorable requelloite que fait li avoient. (*Froissart*, p. 70)

(3) - Sire, fet Lancelot, qui fu cil qui tant a parlé a vos ? Son cors ne **poi je** veoir, mes sa parole **oï je** bien qui est si laide et si espoantable qu'il n'est nus qui grant poor n'en deust avoir. (*Queste*, p. 122)

(4) Jamés n'avra joye et est de mervoilles qu'il ne entre en desesperance, et si **feroit il** si n'estoit qu'il est sages homs. Si lui convient prendre en pacience, quar aultre remyde n'y **peut il** metre; (*Quinze Joyes*, p. 76)

#### 4.1.1. Les fréquences

Le tableau 2 ci-dessous montre que la proportion de P1 inversés sur l'ensemble P1 + P3 inversés connaît une forte variation d'un texte à l'autre : de 0% (*Clari*) à 71.4% (*Griseldis*), les fréquences plus ou moins élevées se répartissant sur toute la période considérée.

Texte + date	Dialecte	Forme	Domaine	Invers. P1+P3 <sup>2</sup>	P1		Invers. P3
					Invers. P1	P1/ P1+P3 inversés <sup>3</sup>	
<i>Roland</i> (1100)	anglo-normand	v	littéraire	21.1	17.8	47.4	25.3
<i>Eneas</i> (1155)	normand	v	littéraire	9.7	12.6	35.3	8.7
<i>Béroul</i> (1165-1200)	traits normands	v	littéraire	17.5	16.1	51	19.3
<i>Ami Amile</i> (1200)	non marqué	v	littéraire	23.8	25.9	52.7	21.8
<i>Clari</i> (1205)	picard	p	historique	42.6	0	0	51.1
<i>Aucassin</i> (début 13 <sup>e</sup> )	traits picards	m	littéraire	9.9	16.9	64.7	5.6
<i>Miracles</i> (1220)	non marqué	v	religieux	21.9	19.6	32.1	23.2
<i>Queste</i> (1230)	non marqué	p	littéraire	22.1	33.1	53.8	15.9
<i>Coutumes</i> (1283)	traits picards	p	juridique	15.8	18.2	14.8	15.4
<i>Joinville</i> (1307)	non marqué	p	historique	12.3	15	52.6	10.2
<i>Froissart</i> (1369-1400)	franco-picard	p	historique	8.8	5.4	13.9	9.9
<i>Griseldis</i> (1395)	traits picards	v	littéraire	13.7	11.4	71.4	27.3
<i>Manieres</i> (1396-1399)	non marqué	m	didactique	4.2	3	50	7.3
<i>Quinze Joyes</i> (1400)	non marqué	p	littéraire	5.7	3.5	25.6	7.3

Tableau 2: Fréquence d'inversion de P1, P3, et P1+P3 : classement chronologique.

Il ressort par ailleurs de ces chiffres que, *Clari* mis à part, l'intervalle de variation des pourcentages d'inversion est relativement étroit, entre 3% et 33%, pour P1, entre 5.6% et 27.3% pour P3. Globalement, l'inversion pronominale est donc peu fréquente, hormis celle de P3 dans *Clari*, texte atypique qui associe absence d'inversion pour P1 et fréquence notablement élevée pour P3 (51%). Par ailleurs, si certains textes présentent des pourcentages d'inversion de P1 et P3 assez proches l'un de l'autre, d'autres au contraire accusent un écart notable : non seulement *Clari*, mais aussi *Aucassin*, *Queste*, *Griseldis*, *Manieres* et *Quinze Joyes*, avec un rapport supérieur à 2 entre les deux pourcentages, au profit de P1 ou de P3 selon les textes. Ainsi, sous les 22% d'inversion de la *Queste* se cachent des fréquences d'inversion de 33% pour P1 et de 16% pour P3.

Le calcul du khi2 permet de mieux apprécier le caractère éventuellement surprenant de la distribution de la position au regard de P1 ou P3.

#### 4.1.2. Valeurs de khi2 significatives

La distribution entre les variables 'position' et 'personne' est assez homogène dans la majorité des textes, ce dont témoignent des valeurs de khi2 qui oscillent entre 0.1 à 3.3. Ces chiffres correspondent à des probabilités qui se situent entre 0.07 et 0.74. Cela signifie que la distribution

5 Les fréquences relatives (en %) d'inversion de P1+P3, de P1, et de P3 sont calculées sur les ensembles respectifs P1+P3 exprimés, P1 exprimés, P3 exprimés.

6 J'indique la proportion (en %) de P1 inversés sur l'ensemble P1+P3 inversés.

observée a entre 7% et 74% de chances d'être le fruit du hasard. En revanche, les 5 textes dans lesquels les pourcentages d'inversion de P1 et P3 sont assez divergents présentent des valeurs de khi2 beaucoup plus élevées<sup>7</sup>. Voici ci-dessous les calculs de khi2 pour *Aucassin*, *Queste*, *Griseldis*, *Manieres* et *Quinze Joyes*.

texte + date	khi2	probabilité	texte + date	khi2	probabilité
Aucassin (déb. 13e)	5.81	0.016	Manières (1396-1399)	4.49	0.034
Queste (1230)	21.3	3,9*10-06	Quinze Joyes (1400)	4.97	0.026
<i>Griseldis</i> (1395)	3.98	0.046			

Tableau 3. Distribution de la position préverbale ou postverbale de P1 et P3 : valeur du khi2 et probabilité associée.

Faute de place, je ne présenterai pas ici le détail des tableaux des khi2. J'en retiendrai les points suivants : les cinq textes présentent des configurations partiellement différentes. Dans deux d'entre eux (*Aucassin* et *Queste*), c'est l'attraction entre position postverbale et P1 qui contribue le plus à la liaison générale entre les variables (environ 50%). Rappelons que *Queste* est le texte qui présente la fréquence d'inversion de P1 la plus élevée du corpus. En revanche, dans trois autres (*Griseldis*, *Manieres* et *Quinze Joyes*) c'est au contraire l'attraction entre P3 et cette même position qui est importante (rappelons que *Manieres* et *Quinze Joyes* présentent des pourcentages d'inversion très voisins). Au-delà de ces divergences, il ressort que c'est la valeur 'postverbale' de la variable 'position' qui joue un rôle déterminant, exerçant une force attractive ou au contraire répulsive vis-à-vis de P1 ou P3.

Que dire des relations qui s'établissent entre ces cinq textes à travers cette liaison forte entre P1 ou P3 et la position postverbale ? On notera que quatre d'entre eux appartiennent au domaine littéraire, et que *Griseldis*, *Manieres* et *Quinze Joyes* sont tous trois de la fin du 14<sup>e</sup> siècle. C'est là sans doute le fait le plus marquant. Hormis cette convergence temporelle, aucune affinité ne semble se nouer entre textes présentant des khi2 voisins.

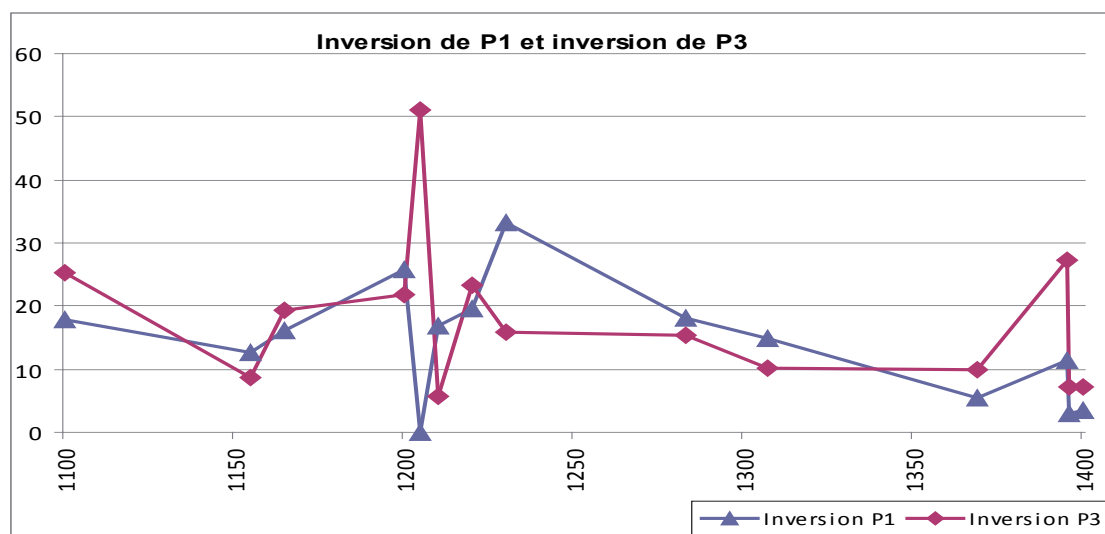
#### 4.1.3. Représentations graphiques

Le graphique 1, ci-dessous, met au jour une amplitude de variation globalement faible sur l'ensemble de la période (*Clari* mis à part), et une progression de ces fréquences à travers les différents textes qui n'a rien d'une ligne régulièrement descendante. Il n'apparaît donc pas d'évolution diachronique spectaculaire. On peut néanmoins tenter de dégager quelques convergences au regard de différents critères : date, domaine et genre, forme et dialecte.

La datation des textes du corpus permet de distinguer nettement deux groupes : d'une part les 3 textes de la fin du 14<sup>e</sup>, très resserrés dans le temps, auxquels on peut adjoindre *Froissart* (1369-1400) ; d'autre part les textes du premier tiers du 13<sup>e</sup> siècle (*Ami*, *Clari*, *Aucassin*, *Miracles* et *Queste*) ; *Roland* est quant à lui nettement détaché au début du 12<sup>e</sup> siècle, tandis qu'*Eneas* est

7 *Clari* a été écarté en raison de la nullité de l'un des effectifs (absence de P1 postverbal).

à mi-chemin entre *Roland* et les textes du début du 13<sup>e</sup> siècle. *Coustumes* et *Joinville* forment un autre groupe, au tournant du 14<sup>e</sup> siècle. Mais il est malaisé de dégager des affinités au sein des regroupements, si ce n'est pour *Manieres* et *Quinze Joye*, qui présentent des fréquences d'inversion voisines pour P1 et pour P3. On peut rapprocher *Froissart* de ces deux textes : les fréquences, bien qu'un peu plus élevées, ne dépassent en effet pas la barre des 10%, P3 s'imposant devant P1, comme dans les deux autres textes. Ces trois textes mis à part, les affinités entre textes quant à leurs fréquences d'inversion ne sont guère liées à leur proximité temporelle. Une seule certitude se dégage : à partir de *Coustumes*, et mis à part le pourcentage de P1 dans *Griseldis*, toutes les fréquences d'inversion de ce corpus sont inférieures à 20%.



Graphique 1 : Inversion P1 et inversion P3

Si l'on considère le critère dialectal, on a du mal à opérer des regroupements entre les trois textes anglo-normands ou normands du 12<sup>e</sup> siècle. Il en va de même pour les textes picardisants. Pour ce qui est du domaine, les trois textes historiques – *Clari*, *Joinville* et *Froissart* – présentent des configurations assez différentes. Parmi les textes littéraires, on peut considérer qu'il y a une certaine proximité entre *Roland* et *Amile*, bien qu'ils divergent sur l'inversion de P1. Rien de tel en revanche entre les 3 romans (*Beroul*, *Eneas*, *Queste*) et le recueil de nouvelles *Quinze Joyes*.

Un désordre assez grand règne donc parmi les textes. Même si les choses semblent commencer à s'ordonner – à la baisse – à la fin du 13<sup>e</sup> siècle, la variation idiolectale apparaît encore comme la grande maîtresse des fréquences de l'inversion.

#### 4.2. La non-expression du sujet

Voici quelques exemples :

(5) [...] de lui **firent** segnor et mestre./Puis **ont gardé** devers senestre (*Eneas*, v. 76-77)

(6) Je ne seroie pas tant ose / Que je i osasse venir / Trop **demor** ci, n'en **quier** mentir. / (*Beroul*, v. 60-64)



(7) Et depuis qu'il l'a congneü / En **a** deux beaux enfans **eü** / Qu'on ne scet qu'ils sont devenuz./ Trop en **est blamez** et **tenuz** / A rigoureux de ses subgez. (*Griseldis*, p. 70)

(8) et les hales sont faites a la guise des cloistres de ces moignes blans, mes je croi que de trop il n'en soit nul si grant. Et vous **dirai** pour quoy il le me semble,... (*Joinville*, p. 48)

#### 4.2.1. Les fréquences

Texte + date	Dialecte	Forme	Domaine	non- expres. P1+P3 <sup>4</sup>	P1		non- expres. P3
					non- expres. P1	P1/P1+P3 non expr. <sup>5</sup>	
<i>Roland</i> (1100)	anglo-normand	v	littéraire	89.3	62.2	11	89.3
<i>Eneas</i> (1155)	normand	v	littéraire	82.8	71.7	14.3	85
<i>Béroul</i> (1165-1200)	traits normands	v	littéraire	82.7	61.7	18.8	89.7
<i>Ami Amile</i> (1200)	non marqué	v	littéraire	81.5	64.1	19.7	87.3
<i>Clari</i> (1205)	picard	p	historique	80.8	25	1.3	83.3
<i>Aucassin</i> (début 13 <sup>e</sup> )	traits picards	m	littéraire	69.7	37.5	9.8	77
<i>Miracles</i> (1220)	non marqué	v	religieux	84	70.5	16.3	87.3
<i>Queste</i> (1230)	non marqué	p	littéraire	68.6	21.5	4.5	76.5
<i>Coutumes</i> (1283)	traits picards	p	juridique	26.9	26.7	12.7	27
<i>Joinville</i> (1307)	non marqué	p	historique	44.1	21.1	4.5	54.1
<i>Froissart</i> (1369-1400)	franco-picard	p	historique	68.6	24.6	3.4	73.2
<i>Griseldis</i> (1395)	traits picards	v	littéraire	69.6	65.1	69.7	82.8
<i>Manieres</i> (1396-1399)	non marqué	m	didactique	16.7	4.5	16.8	36.6
<i>Quinze Joyes</i> (1400)	non marqué	p	littéraire	46.1	15.5	8.9	57.3

Tableau 4. Fréquences de non-expression de P1, P3, et P1+P3.

La proportion de P1 non exprimés sur l'ensemble P1 + P3 non exprimés connaît une variation bien moindre que pour l'inversion : P1 est largement minoritaire, à l'exception de *Griseldis*. De plus, certaines régularités se dégagent: dans les textes dans lesquels la proportion de P1 est inférieure à 10%, la fréquence de sa non-expression ne dépasse pas 30% (sauf *Aucassin*). Ce point mis à part, on observe que la non-expression est globalement (P1 et P3 confondus) très fréquente dans tous les textes. A l'exception de *Coutumes*, *Joinville*, *Manieres* et *Quinze Joyes*, sa fréquence dépasse 68%, et même 80% dans 6 d'entre eux. L'amplitude de variation de la non-expression est donc très large, de 16.7% (*Manieres*) à 89.3% (*Roland*). Toutefois, si l'on considère les fréquences respectives de P1 et de P3, l'analyse se révèle plus complexe. Tout d'abord la fréquence de non-expression de P3 est bien plus élevée que celle de P1, à

8 Les fréquences relatives (en %) de non-expression de P1+P3, de P1, et de P3 sont calculées sur les ensembles respectifs P1+P3 (exprimés et non exprimés), P1 (exprimés et non exprimés) et P3 (exprimés et non exprimés)

9 J'indique la proportion (en %) de P1 non exprimés sur l'ensemble P1+P3 non exprimés.

l'exception de *Coustumes* (les pourcentages de P1 et P3 plafonnent à 27%). Dans la moitié des autres textes, le rapport entre la fréquence de non-expression de P3 et celle de P1 oscille entre 3 et 4. La non-expression de P3 est même 8 fois plus fréquente dans *Manieres*. L'amplitude de variation des fréquences reste élevée dans tous les cas : de 5% à 71.7% pour P1, et de 27% à 94% pour P3. A nouveau s'impose la nécessité de ne pas écraser les spécificités de P1 et P3 sous les fréquences globales. Ainsi les pourcentages de non-expression de P1+P3 de *Froissart* et de *Griseldis* avoisinent 69%, chiffre qui recouvre une fréquence de non-expression de P1 de 24.6% pour *Froissart*, de 65% pour *Griseldis*.

#### 4.2.2. Valeurs de khi2 significatives

La conjonction de la moindre représentation de P1 parmi les Sp non exprimés et de sa fréquence de non-expression presque toujours inférieure à celles de P3 le laissait attendre : les khi2 calculés pour chacun des textes atteignent des valeurs très élevées et sont donc associés à des probabilités excessivement faibles que la distribution observée soit le fruit du hasard. Un texte cependant se singularise : *Coustumes*, pour lequel le calcul du khi2 met au jour une probabilité de 97% que la distribution observée soit le fruit du hasard.

Texte + date	dialecte	forme	domaine	khi2	Probabilité
<i>Roland</i> (1100)	anglo-normand	v	littéraire	245.4	2.54*10 <sup>-55</sup>
<i>Eneas</i> (1155)	normand	v	littéraire	35.1	3.11*10 <sup>-9</sup>
<i>Béroul</i> (1165-1200)	traits normands	v	littéraire	159.3	1.64*10 <sup>-36</sup>
<i>Ami Amile</i> (1200)	non marqué	v	littéraire	79.8	4.17*10 <sup>-19</sup>
<i>Clari</i> (1205)	picard	p	historique	104.5	1.54*10 <sup>-24</sup>
<i>Aucassin</i> (début 13 <sup>ème</sup> )	traits picards	m	littéraire	62.6	2.54*10 <sup>-15</sup>
<i>Miracles</i> (1220)	non marqué	v	religieux	26.3	2.88*10 <sup>-7</sup>
<i>Queste</i> (1230)	non marqué	p	littéraire	296	2.4*10 <sup>-66</sup>
<i>Coustumes</i> (1283)	traits picards	p	juridique	0.001	0.97
<i>Joinville</i> (1307)	non marqué	p	historique	104.1	1.91*10 <sup>-24</sup>
<i>Froissart</i> (1369-1400)	franco-picard	p	historique	121.1	3.69*10 <sup>-28</sup>
<i>Griseldis</i> (1395)	traits picards	v	littéraire	14.2	0.00016
<i>Manieres</i> (1396-1399)	non marqué	m	didactique	98.6	3.1*10 <sup>-23</sup>
<i>Quinze Joyes</i> (1400)	non marqué	p	littéraire	193.8	4.6*10 <sup>-44</sup>

Tableau 5 : Distribution de l'expression et de la non-expression de P1 et P3 : khi2 et probabilité

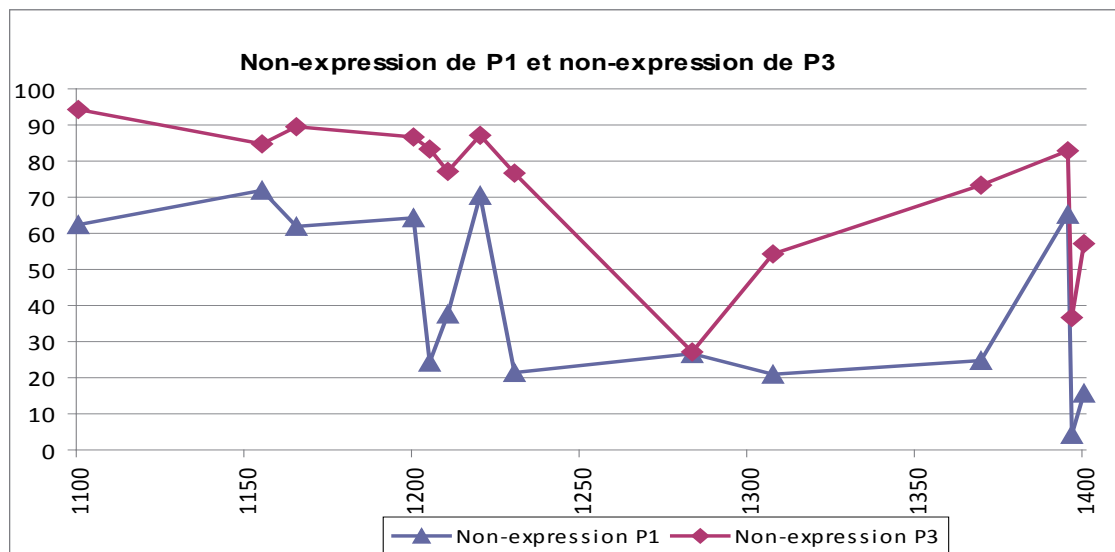
Hormis *Coustumes*, tous les textes présentent une configuration similaire : une attraction entre la non-expression et P3, et une répulsion entre cette même non-expression et P1, même si la régularité de ces liaisons recouvre des attractions et des répulsions plus ou moins fortes, ainsi que des divergences variables entre effectifs réels et théoriques, la plus fréquente consistant en la conjugaison d'une attraction entre P1 et son expression, et d'une répulsion entre P1 et sa non-expression (*Graal*, *Quinze Joyes*, *Froissart*, *Clari*, *Joinville*, *Aucassin* et *Eneas*). Cette configuration se retrouve dans les trois textes historiques (*Clari*, *Joinville* et *Froissart*).

#### 4.2.3. Représentations graphiques :

Bien que le graphique 2 ci-dessous ne traduise pas d'évolution chronologique continue, on observe néanmoins des tendances plus nettes que pour l'inversion. Si l'on considère le paramètre chronologique, on constate que le début de la période présente une relative stabilité, au moins en ce qui concerne la courbe de P3. De *Roland* à *Queste*, la non-expression de P3 dépasse 70%. C'est en revanche moins net pour P1. Relativement stable (entre 60% et 70%) pour *Roland*, *Eneas* et *Beroul*, les 3 textes les plus anciens, mais aussi pour *Amile* (1200), la fréquence de la non-expression de P1 connaît une chute vertigineuse dans *Clari*, texte par ailleurs très atypique en ce qui concerne l'inversion. La non-expression de P1 reste faible dans *Aucassin*, remonte dans *Miracles*, mais connaît une nouvelle chute dans *Queste*. On notera que *Aucassin* et *Queste*, dans lesquels la non-expression de P1 est donc nettement en deçà de celle de P3, sont deux textes qui présentent à l'inverse une fréquence d'inversion de P1 supérieure à celle de P3. C'est globalement plus confus après *Queste*. *Coustumes* accuse une chute généralisée des fréquences, avec en outre une parfaite concordance (quantitative au moins) entre la non-expression de P1 et celle de P3. La non-expression repart à la hausse dans les 3 textes suivants, *Joinville*, *Froissart* et *Griseldis*, ce texte renouant avec des fréquences proches de celles des textes les plus anciens. Les fréquences connaissent une nouvelle chute dans *Manieres*, particulièrement en ce qui concerne P1 (4.5%), suivie d'une légère remontée dans *Quinze Joyes*. Alors que *Manieres* et *Quinze Joyes* présentent de fortes affinités en ce qui concerne l'inversion de Sp, les deux textes divergent assez nettement au regard de sa non-expression. Hormis pour plusieurs textes du début de la période qui ont des configurations de non-expression de Sp assez proches, le critère chronologique n'apparaît pas, dans ce corpus, comme nettement déterminant. Les deux derniers textes de la fin de la période (*Manieres* et *Quinze Joyes*) ont des fréquences moindres que bon nombre des textes précédents : peut-être cela préfigure-t-il le début d'une baisse généralisée, mais il faudrait, pour le confirmer, considérer d'autres textes, contemporains et postérieurs.

Le critère dialectal semble un peu plus discriminant pour la non-expression de Sp que pour son inversion. Il reste un critère à manier avec précaution, dans la mesure où ses différentes valeurs recouvrent en partie les valeurs des autres critères. Ainsi, il apparaît que les textes présentant des traits anglo-normands ou normands se trouvent au 12<sup>e</sup>, tandis que ceux à caractère picardisants se trouvent au 13<sup>e</sup> et encore au 14<sup>e</sup> siècle. De plus, deux des textes picardisants sont historiques, et les deux textes « normands » (*Eneas* et *Beroul*), non seulement sont assez proches chronologiquement, mais sont aussi tous deux des romans. Les mêmes précautions sont de mise en ce qui concerne le domaine, dans la mesure où les affinités observées jusqu'ici recourent des affinités sur le plan dialectal et/ou chronologique. Rappelons cependant que nous avons mis au jour, dans le cadre du calcul du khi2, des affinités entre les trois textes historiques. Reste enfin le critère de la forme. Aucune spécificité au regard de ce critère ne s'est dégagée pour l'inversion. On constate en revanche que les deux textes du début du 13<sup>e</sup> et de la fin du 14<sup>e</sup> qui présentent des affinités avec les textes plus anciens sont précisément ceux qui sont en vers : *Miracles* et *Griseldis*, que tout sépare par ailleurs. Deux textes ne permettent certes pas de poser une corrélation entre non-expression élevée et forme versifiée : il faudra explorer d'autres textes pour le vérifier.

En conclusion, on retiendra que, plus que pour l'inversion, les configurations de la non-expression de Sp dans les différents textes font apparaître quelques tendances, sur les plans chronologique, dialectal, du domaine (et du genre), et de la forme.



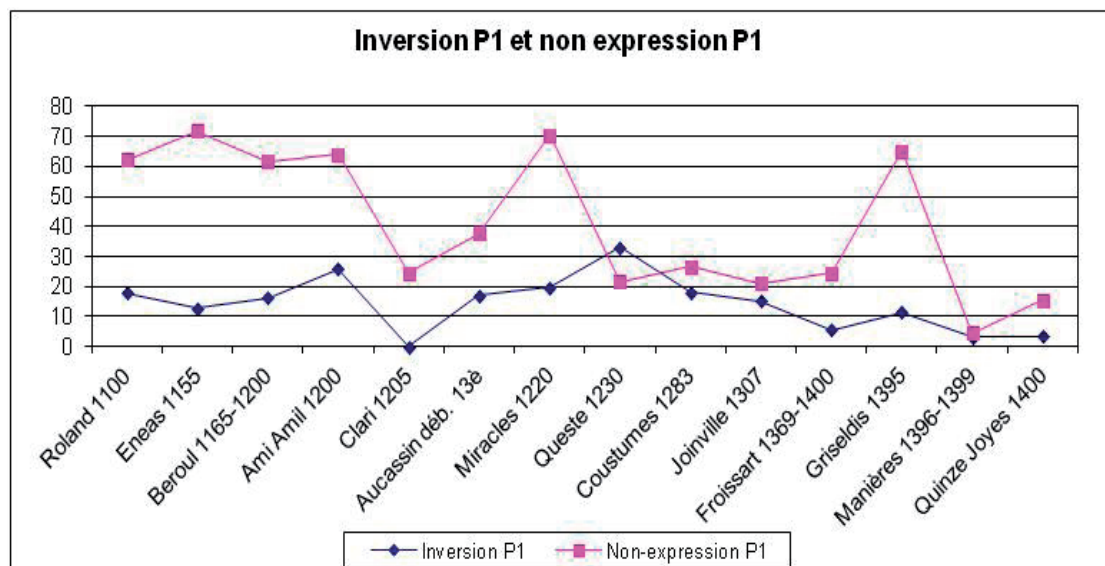
Graphique 2 : Non-expression P1 et non expression P3

#### 4.3. Inversion et non-expression de Sp : peut-on dégager des (cor)relations ?

Les fréquences liées à P3 présentent quelques affinités : *Amile* et *Miracles*, dont la fréquence de non-expression de P3 est de 87.3%, tendent à inverser P3 dans des proportions similaires (21.8% et 23.2%). De ces deux textes on peut rapprocher *Beroul*, dont les fréquences de non-expression et d'inversion sont de 89.7% et de 19.3%. Ces trois textes se situent au tournant des 12<sup>e</sup> et 13<sup>e</sup> siècles, mais ils n'ont sinon que peu de points communs. Les autres textes présentant une fréquence de non-expression de P3 élevée (entre 83% et 85%) ont en revanche des tendances à inverser P3 très variables : 8.7% pour *Eneas*, 51.1% pour *Clari* et 27.3% pour *Griseldis*. Si l'on considère à l'inverse les deux textes dans lesquels la non-expression de P3 est la plus faible (*Coustumes* : 27% et *Manieres* : 36.6%), on constate de mêmes divergences en matière d'inversion (respectivement 15.4% et 7.3%).

C'est cependant du côté de P1 que l'absence de convergences entre les deux fréquences est la plus frappante. Parmi les quatre textes qui offrent une fréquence de non-expression oscillant entre 61 % et 65%, deux ont des fréquences d'inversion assez proches (*Roland* : 17.8%, *Beroul* : 16.1%), mais dans les deux autres la proportion d'inversion de P1 est au contraire très différente : 25.9% dans *Amile*, et 11.4% dans *Griseldis*. Notons que le texte dans lequel la non-expression est la plus fréquente (*Eneas* : 71.7%) est aussi parmi ceux qui offrent une fréquence d'inversion peu élevée (12.6% dans *Eneas*). Si l'on considère les fréquences d'inversion, on constate que les quatre textes les plus tardifs (*Froissart*, *Griseldis*, *Manieres* et *Quinze Joyes*) offrent les chiffres les plus bas (de 3% à 11.4%), mais présentent en revanche des fréquences d'expression excessivement variables, de 4.5% à 65.1%. On pourrait poursuivre ainsi, et souligner la difficulté à déceler des convergences régulières. Signalons pour finir que le texte qui présente la fréquence d'inversion de P1 la plus forte, et largement en tête des autres (33.1%), est aussi celui dont la fréquence de non-expression de P1 est relativement basse : 21.5%. *Queste* est d'autant plus spécifique à cet égard que c'est le seul texte du corpus dans lequel la fréquence d'inversion de P1 est supérieure à celle de sa non-expression.

Les graphiques présentés ci-dessous mettent en perspective inversion et non-expression pour chacune des personnes.

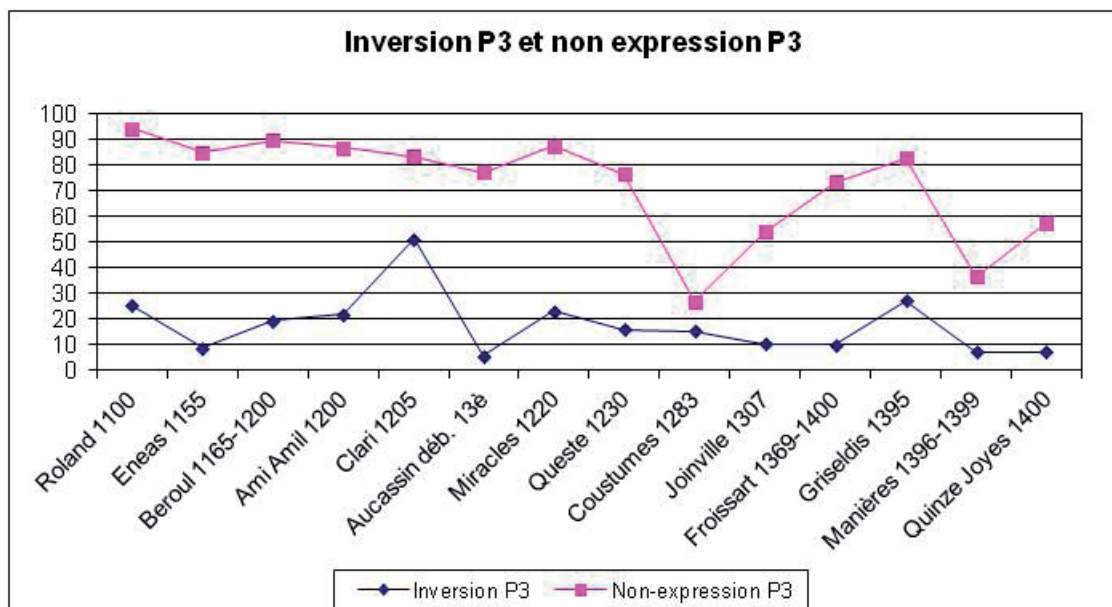


Graphique 3 : Inversion et non-expression de P1..

L'observation du graphique 3 montre que les deux lignes suivent un mouvement relativement parallèle jusque *Miracles* (1220), qui présente un point culminant remarquable au niveau de la non-expression. A partir du texte suivant (*Queste*, 1230), les lignes divergent nettement (hormis de *Coutumes* à *Joinville*). Signalons d'une part le croisement des lignes au niveau de *Queste*, qui, de manière très atypique, présente une fréquence d'inversion de P1 supérieure à sa non-expression. Signalons d'autre part, à la fin du 14<sup>e</sup> siècle, le caractère pareillement singulier de *Griseldis*, qui accuse une forte poussée de la non-expression de P1.

On observe pour les tracés de P3 (graphique 4 ci-dessous) des mouvements plus convergents entre les deux lignes de fréquence. En effet, hormis les deux points que constituent *Clari* pour l'inversion (la plus fréquente du corpus) et *Coutumes* pour la non-expression (la plus rare du corpus), les deux lignes dessinent une progression bien plus voisine que pour P1.

Cette plus grande convergence entre les fréquences d'inversion et de non-expression de P3 qu'entre celles de P1 est confirmée par le calcul du coefficient de corrélation. Celui-ci a été établi pour évaluer la liaison entre l'inversion et la non-expression de P1 d'une part, et entre celles de P3 d'autre part. Avec une légère dérogation à ce qui se fait le plus couramment, il a été calculé à partir des fréquences relatives et non pas des fréquences absolues. Ce choix se justifie par la différence de taille entre les textes. Le calcul de la moyenne, sur lequel se fonde celui des écarts réduits, aurait été biaisé par la forte variabilité de la taille des textes.



Graphique 4: Inversion et non-expression de P3

Sans surprise, le coefficient de corrélation est légèrement plus élevé pour P3 (0.43) que pour P1 (0.37). L'écart entre les deux n'est cependant pas spectaculaire. Surtout, les deux valeurs correspondent à une probabilité supérieure à 10% d'atteindre ou de dépasser ces coefficients. On ne peut donc rejeter l'hypothèse nulle : il n'y a pas de dépendance statistique entre les deux séries de variables (inversion et non-expression) et cela quelle que soit la personne. S'il est apparu des affinités plus grandes entre inversion et non-expression de P3, on ne peut malgré tout parler de corrélation au sens statistique du terme.

## 5. Conclusion

A l'issue de cette étude, on retiendra la difficulté, sur le plan quantitatif, à dégager des tendances sans écraser les spécificités. C'est donc bien la variabilité des fréquences qui domine, entre personnes au sein d'un même texte, et d'un texte à l'autre. Nous avons vu par ailleurs que quatre textes ont un comportement particulièrement singulier : *Clari*, dans lequel l'inversion en général est atypique, de même que la non-expression de P3 (rare) ; *Coutumes*, qui offre des fréquences de non-expression très basses et proches les unes des autres ; *Griseldis*, dont la non-expression en général et l'inversion de P3 sont au contraire très élevées ; et *Queste* enfin, qui présente une fréquence d'inversion de P1 supérieure à celle de sa non-expression. Que faire de ces textes singuliers ? Faut-il les traiter à part ?

Si la pertinence du critère chronologique est avérée sur le long cours (non-expression et inversion ont régressé), il ne s'impose pas comme caractère déterminant au sein de la période considérée : il n'y a pas de baisse régulière. Les autres critères envisagés (dialecte, domaine et forme) ne jouent de leur côté qu'un rôle ponctuel. J'ai par ailleurs insisté sur le problème de 'parasitage' entre les paramètres.

Assurément le corpus mérite d'être élargi, afin de décliner de manière plus systématique les différents paramètres, et afin aussi de mieux représenter le milieu des 13<sup>e</sup> et 14<sup>e</sup> siècles. J'ai

justifié ce choix d'un point de vue méthodologique. Mais il répondait aussi à des motivations linguistiques : j'ai fait l'hypothèse qu'il s'agissait des deux bornes entre lesquelles les changements s'opéraient. De fait, il apparaît une relative homogénéité parmi les textes de la fin du 12<sup>e</sup> siècle, tant en ce qui concerne l'expression que la position du sujet. Dans une moindre mesure, on observe aussi des affinités fréquentielles entre les textes de la fin du 14<sup>e</sup> siècle, au moins en ce qui concerne l'inversion de Sp. Entre ces deux bornes, la situation est bien plus désordonnée et complexe. Il semble donc – mais cela méritera d'être confirmé – que la phase de transition se caractérise, non pas par une baisse régulière, mais par des pratiques très diversifiées d'un texte à l'autre. On peut faire l'hypothèse que ce moment de 'passage', et donc de désorganisation partielle du système, ouvre la voie à une variété d'usages plus grande que pendant les phases de relative stabilité.

Il est par ailleurs apparu que l'évolution de la non-expression de P1 est plus chaotique que celle de P3. Or, globalement, la non-expression de P1 est bien moins fréquente que celle de P3. Il se peut qu'il y ait un lien entre la fréquence d'une construction et la régularité plus ou moins grande de son évolution. Cette hypothèse doit cependant être envisagée avec précautions : elle se fonde sur un phénomène qui implique la personne du locuteur. Or celle-ci est le lieu privilégié de l'expressivité et de la subjectivité : que ses modalités d'expression soient soumises à une forte variabilité d'un texte à l'autre n'est pas surprenant. On notera cependant qu'il n'y a pas une telle différence P1/P3 au niveau de l'inversion.

Les remarques formulées ci-dessus méritent d'être confirmées (ou non) par l'exploration de données supplémentaires, de même que doit être menée une analyse qualitative approfondie.

## Références

- Buridant, C. 2000. *Grammaire nouvelle de l'ancien français*. Paris : Sedes.
- Combettes, B. (1988). *Recherches sur l'ordre des éléments de la phrase en moyen français* (Thèse pour le Doctorat d'Etat, Université de Nancy ; exemplaire dactylographié).
- Detges, U. (2003). Du sujet parlant au sujet grammatical. L'obligatorisation des pronoms sujets en ancien français dans une perspective pragmatique et comparative. *Verbum*, XXV (3) : 307-333.
- Foulet, L. (1930, 1<sup>ère</sup> ed. 1919). *Petite syntaxe de l'ancien français*. Paris : Champion.
- Marchello-Nizia, C. (1995). *L'évolution du français : ordre des mots, démonstratifs, accent tonique*. Paris : Armand Colin.
- Prévost, S. (2011a). Expression et position du sujet pronominal en français, *Mémoires de la Société de Linguistique de Paris*, Tome XIX : 13-33.
- Prévost, S. (2011b) *Expression et position du sujet pronominal du 12<sup>e</sup> au 14<sup>e</sup> siècle : une approche quantitative*. Mémoire de recherche inédite pour l'HDR, soutenue le 8/07/11 à l'ENS de Lyon.
- Skårup, P. (1975). *Les premières zones de la proposition en ancien français. Essai de syntaxe de position*. Etudes romanes de l'Université de Copenhague, *Revue Romane*, numéro spécial 6, Akademisk Forlag.
- Vance, B. (1997). *Syntactic Change in Medieval French : Verb-Second and Null Subjects*, Dordrecht-Boston-Londres : Kluwer Academic Publishers.
- Vennemann, T. (1976). Topics, subjects and word-order : from SXV to SVX via TVX. in J.M Anderson et C. Jones eds. *Proceedings of the first international congress of Historical Linguistics*. Amsterdam: 339-376.