

Types de textes et changement syntaxique

Pierre Larrivée¹ et Daniéla Capin²

¹ CRISCO (EA4255), Université de Caen

² *Scolia* / LILPA (EA 1339), Université de Strasbourg
pierre.larrivee@unicaen.fr

Résumé. A la différence des textes littéraires médiévaux, les textes juridiques médiévaux connaissent un taux de sujets non-exprimés très bas. Cette étude cherche à vérifier l'incidence de trois facteurs sur l'expression du sujet dans des textes juridiques du français médiéval : la longueur du texte, la personne verbale et la catégorie – nominale ou pronominale – du sujet, majoritaire dans ce type de textes. Réalisée sur deux corpus annotés de manière identique – un texte long (*Les Etablissements des Capétiens*) et des textes courts (des chartes du *Cartulaire* de l'Ile-de-France) – elle montre que, dans le cas des textes étudiés, la longueur du texte n'est pas significative, que la personne et la catégorie verbale du sujet ne constituent pas des facteurs concluants.

Abstract. Types of text and syntactic change French medieval literary texts have a higher ratio of null subjects than their legal counterparts. This study examines three orders of factors that have been alleged to explain the greater rate of subject expression in legal material. The length of the text, the pronominal person as well as the nominal or pronominal category of the subject are explored in two corpora, annotated according to the same principle – one long text (*Les Etablissements des Capétiens*) and series of short texts (charters from the *Cartulaire* de l'Ile-de-France). The study shows that in these texts no relationship can be found between the length of the text and the expression of the subject ; and that likewise the person and the category of the subject can not be considered as relevant factors for subject expression.

1 Introduction

Le recours obligé à des textes écrits pour la recherche diachronique a plusieurs répercussions méthodologiques. On accepte de façon générale que la pratique vernaculaire spontanée non-surveillée présente la grammaire la plus stable [7] et constitue le lieu d'impulsion privilégié des changements linguistiques [6]. Or, cette pratique est la moins susceptible d'être l'objet d'une mise à l'écrit, et sa mise à l'écrit est la moins susceptible d'être transmise à travers les siècles, parce qu'elle n'a pas la même légitimité que certaines productions religieuses ou politiques. On se trouve donc obligé parmi les (parties de) textes existants à retenir ceux qui sont susceptibles de s'approcher du vernaculaire, dans la mesure où l'on cherche à rendre compte de l'organisation et du changement de la langue ancienne en tant que reflet des connaissances immédiates effectives que possèdent les locuteurs des communautés concernées. C'est ce que cherchent à faire, entre autres, les travaux sur l'oral

représenté [5] dans les textes littéraires. Une autre approche consiste à se référer à des textes qui ne sont pas soumis à des visées stylistiques pouvant éloigner encore un peu plus du vernaculaire.

Quel que soit le choix, toutes les sources textuelles sont marquées par des propriétés textuelles que l'on peut aisément, si l'on n'y prend garde, prendre pour des propriétés du système de la langue. La forme textuelle (vers ou prose), le type (descriptif, explicatif, narratif, etc.) ou le domaine (littéraire, historique, juridique, etc.) influencent fortement les données. Ainsi, un travail récent de Simonenko, Crabbé et Prévost [13] montre que les textes en vers et les textes en prose donnent des trajectoires significativement différentes pour des changements syntaxiques majeurs de la langue ancienne comme la réalisation du sujet grammatical.

Qu'en serait-il si un ensemble textuel attestait plus précocement d'un changement syntaxique majeur dont on sait par ailleurs qu'il a eu lieu ? Faudrait-il y voir un meilleur reflet du changement linguistique effectif, ou une caractéristique imposée par le genre de texte lui-même ? Des travaux récents suggèrent que la perte du sujet nul (sujet non-exprimé) dans l'ancienne langue est attestée dans les textes légaux avec trois siècles d'avance sur les textes littéraires [2, 3 et 1]. De manière intuitive, on peut supposer que les textes légaux seraient plus susceptibles que d'autres d'instancier le sujet à cause des précisions exigées par la pratique juridique. Aussi judicieux qu'il puisse paraître, ce raisonnement est malheureusement faux si l'on se fie aux (rares) études existantes : les résultats obtenus après application du logiciel *Elliphant* [11] sur le sujet nul dans des textes juridiques, éducatifs et techniques en espagnol contemporain ne démontrent pas de taux fortement différencié sous ce rapport.

L'objectif de ce travail est d'évaluer trois hypothèses récentes produites dans des études cherchant à expliquer la perte du sujet nul dans les textes médiévaux et les raisons de l'émergence précoce du sujet obligatoire dans les textes juridiques. En effet, il a été proposé que l'émergence précoce du sujet y dépendrait des caractéristiques textuelles des sources. Ces caractéristiques incluraient la longueur des textes, la personne grammaticale et le nombre respectif des pronoms personnels et des groupes nominaux, ainsi que la proportion relative des pronoms personnels et des groupes nominaux en fonction de sujet. C'est donc l'impact de ces trois facteurs sur l'expression du sujet dans les textes légaux que nous nous proposons de mesurer. Le protocole utilisé pour ce faire est décrit dans les deux sections qui suivent.

2 Hypothèses

Notre étude s'intéresse à trois hypothèses récentes selon lesquelles ce seraient des caractéristiques des textes légaux qui y expliqueraient l'émergence apparemment plus précoce du sujet obligatoire. Chaque hypothèse examine une caractéristique particulière.

La première est proposée par Christiane Marchello-Nizia [9]. Elle suggère que la représentation plus précoce des changements syntaxiques majeurs dans les textes légaux serait fonction de leur brièveté : les textes plus courts seraient moins susceptibles d'offrir des chaînes anaphoriques que l'on retrouve dans les textes plus longs ; la reprise anaphorique étant souvent opérée par les sujets nuls, un taux supposé plus faible de reprises anaphoriques entraînerait des occurrences plus rares de sujet nul. La rareté du sujet nul donnerait l'impression d'une prépondérance du sujet obligatoire.

Une étude de Sophie Prévost [10] propose une autre hypothèse : dans son important corpus, essentiellement composé de textes littéraires, le sujet nul est plus fréquent avec les pronoms de troisième personne qu'avec ceux de première personne. Cependant, les mêmes taux de sujet nul se manifestent sans égard à la personne dans le seul texte juridique de ce corpus – les *Coutumes de Beauvaisis*. On suppose que cela serait dû à la rareté des pronoms autres que ceux de troisième personne dans les textes légaux, en revenant sur l'idée que ces

textes doivent de par leur nature être plus explicites quant à leur sujet grammatical de par leur nature.

Enfin, l'hypothèse de Simonenko, Crabbé et Prévost [12] aborde, entre autres facteurs, le rapport relatif entre pronoms personnels et groupes nominaux en fonction de sujet. Les sujets pronominaux étant plus susceptibles de ne pas être exprimés, leur plus faible représentation dans les textes légaux expliquerait les taux plus bas de sujets nuls, s'il s'avérait que les textes légaux contiennent plus de sujets nominaux que les textes littéraires.

Le présent article examine l'impact supposé de ces facteurs sur l'expression du sujet dans les textes juridiques. Il cherche à vérifier si la précocité de la disparition du sujet nul dans les textes légaux est fonction de leurs caractéristiques textuelles, comme on pourrait l'alléguer, ou s'il ne faut pas envisager l'hypothèse que, à cette époque, bien avant leur professionnalisation, les textes juridiques ne seraient pas plus proches de l'expression vernaculaire que les textes à visée stylistique du domaine littéraire. La méthode de cette vérification est décrite dans la prochaine section.

3 Données et méthode

Afin de vérifier les hypothèses évoquées sur l'émergence précoce du sujet obligatoire dans les textes légaux, nous avons choisi des textes du domaine juridique qui appartiennent à la période 1270-1280. Ils proviennent de deux régions limitrophes du Centre de la France. Ce choix est déterminé par le besoin d'écarter des variantes diatopiques importantes ; par ailleurs, ces régions ont joué un rôle important dans le développement de la norme linguistique unitaire. Deux ensembles de textes sont ainsi retenus. Les *Établissements des Capétiens* sont un code de lois coutumières, connu aussi sous le nom de *Lois de Saint-Louis*, de 213 chapitres sur 282 pages dans l'édition ancienne de Jourdan, Decrusy et Isembert (1821-1833), dont on retient les 300 et quelques premières propositions. L'ensemble de textes courts est représenté par 23 chartes du *Cartulaire* d'Île-de-France dont les transcriptions diplomatiques sont produites et mises en ligne par l'École Nationale des Chartes. Permettant de comparer deux types de textes légaux, le choix de textes courts et d'un texte long est motivé par l'objectif de vérifier l'hypothèse de l'impact de la longueur sur l'expression du sujet. La sélection des textes courts dans le grand ensemble du *Cartulaire* d'Île-de-France est guidée non seulement par la période, mais aussi par deux autres raisons : d'une part, l'ensemble contient des chartes longues (de plus de 450 mots) et courtes (moins de 200 mots) ; d'autre part, ces textes contiennent des marques de conjugaison verbale à une personne autre que la seule troisième personne du singulier. Ce choix permet de répondre à l'objectif de vérifier l'impact de la personne grammaticale que propose Sophie Prévost.

Les textes choisis sont téléchargés et mis en format autorisant le traitement de texte. Cela permet l'annotation exhaustive des propriétés pertinentes des propositions et le comptage automatique des propriétés codées (à l'aide de la fonction « recherche » dans MSWord). Les propriétés qui donnent lieu à l'annotation sont les suivantes : le statut de la proposition sous le rapport de sa dépendance hiérarchique (subordonnée, Sub) ou non (principale, Prc, dans laquelle on inclut les indépendantes) à une autre proposition ; l'expression ou non du sujet (Sujet Exprimé ou Sujet Non Exprimé), son statut morphosyntaxique (pronom personnel, SPP ; pronom relatif, SPrel ; autre pronom, Spro% ; groupe nominal, SLx) sa position (avant, Av, ou après, Ap, le verbe) et son statut sémantique (caractère référentiel, S, ou impersonnel, SI). Les différents types de coordinations (Coord 1 et Coord 2) seront explorés dans de futures études. Les passages suivants illustrent ces annotations :

- (1) Se aucuns vient devant aus, <Sub><SPro><Av> et muet question de marchié <Sub><Coord><SPro><Av>qu'il ait fait <Sub><SPP><Av>, rencontre un autre, ou

demande heritage <Sub><Coord><SPro><Av>, li prevost semondra celuy <Pr><SLx><Av> dont len se plaindra <Sub><SPP><Av>. Et quand les parties vendront à ce jour <Sub><SLx><Av>, li demandierres fera sa demande <Pr><SLx><Av>, et celuy à qui len demande <Sub><SPro><Av>, respondra, <Pr><SPro><Av> à cel jour meismes, se ce est de son faict<Sub><Spro %><Av>. (*Etablissements des Capétiens*, Chap.1)

(2) A touz cels qui verront <Sub><SPro><Av> cestes presentes lettres Gautier de Nemos, seigneur d'Achieres, salut en Nostre Seigneur. Sachent tuit <Pr><SPro><Ap> que nos quarente solz, que de cens que d'autre rente, assis a Semoi, a Meun, a Orlens, a Dornisi et en toz les lyeus devant diz, et totes les apartenances de celi cens, queles que eles soient <Sub><SPP><Av>, le quel cens ou les apartenances honorauble homme Remon, chevecier d'Orliens, achata <Sub><SLx><Av> de hanorable homme Gile de Boisson Caillart, deen de Seint Pere Empont, et a donné et annexé à la chevecerie d'Orliens <Sub><Coord2>, et le quel cens et toutes les apartenances li diz cheveciers tenoit <Sub><SLx><Av> en fié de mon seigneur Jahan Morin, chevalier, et icis Johans Morins lou tenoit <Sub><SLx><Av> ansinc de noz en fié, avons amorti <Sub><SPP><1PP><Av> et amortissons <Coord1> a la dite chevecerie d'Orliens come seignouz sire dou fié de qua la valor de diz quarente soz desus nommez et des apartenances, et volons <Sub><SPP><1PP><Coord2> que li diz quarente soz de cens et toutes les apartenances de celui cens soient tenues <Sub><SLx><Av> a toujours mes dou dit chevecier et de cels qui seront chevecier <Sub><SPro><Av> après lui an mein morte. (*Charte Orléans Sainte Croix*, 18 avril 1277)

Nous opérons ensuite le décompte de chaque propriété dans chaque ensemble. Nous comparons le ratio des manifestations de ces propriétés. Cela permet de tester l'impact de la longueur des textes, de la personne du pronom personnel et du rapport entre pronoms et groupes nominaux sur le taux d'expression du sujet. Les ratios sont soumis à des tests statistiques standards de khi2 pour établir le caractère significatif des variations quantitatives constatées. Sur la base des résultats des recherches antérieures, notre attente est que, dans des textes juridiques de la période 1270-1280, dans la zone géographique étudiée, la perte du sujet nul plus précoce n'est pas corrélée à la longueur des textes, ni à la personne verbale ou au type de sujet majoritairement utilisé, mais à la plus grande proximité de la langue des ces textes à la langue vernaculaire, du moins dans la période étudiée, soit, avant la professionnalisation des textes juridiques.

4 Longueur des textes et expression du sujet

Une hypothèse pour expliquer la disparition tardive des sujets non-exprimés dans les textes littéraires a été récemment avancée par Christiane Marchello-Nizia [9 et 10]. Elle considère la longueur du texte comme un des cinq facteurs externes expliquant le taux de sujets exprimés, aux côtés de la forme vers / prose, du genre, du registre et du dialecte. Elle suggère que les textes brefs contiennent une sur-représentation de sujets exprimés, en opposant des textes brefs anciens non littéraires (les *Serments de Strasbourg*, la *Cantilène de Sainte-Eulalie*, les chartes de Chièvres, Tournai et Arras) aux textes littéraires plus longs (*Passion*, *Saint-Alexis*, *Roland*, *Eneas*, *Yvain*, *Dole*, *Queste*). En agrégeant le taux de sujet nul des deux groupes de textes donnés, nous calculons une moyenne de 17,8 % pour le premier groupe de textes et 43 % pour le second, soit un rapport de sujets nuls du simple dans les textes courts à plus du double dans les textes longs.

Il semblait donc opportun de tester l'effet de la longueur du texte sur le taux de réalisation du sujet. Pour ce faire, nous avons choisi deux groupes de textes du même domaine dans la même région et de la même époque afin de minimiser les variations liées à ces facteurs. Nous avons étudié le début du texte coutumier *Établissement des Capétiens* à travers une séquence de 2505 mots dans laquelle nous avons annoté la réalisation des sujets. Face à lui, nous avons retenu 23 chartes du *Cartulaire d'Île-de-France* (1270-1280) qui totalisaient 6499 mots et où la réalisation du sujet a été systématiquement annotée. Les résultats sont les suivants.

L'extrait annoté des *Établissement des Capétiens* comptait 91 propositions principales et 225 propositions subordonnées, soit un total de 317 propositions conjuguées. De ces 317 propositions, 25 comptaient un sujet non exprimé, soit 17 sujets personnels non exprimés renvoyant à un antécédent et 8 sujets impersonnels non exprimés. Le rapport de sujets non exprimés au nombre de propositions représente un pourcentage de 7,89 %. Des 25 sujets non-exprimés, 9 se retrouvaient dans les 91 propositions principales (dont 2 sujets impersonnels non-exprimés) et 16 dans une des 225 propositions subordonnées (dont 6 sujets impersonnels non-exprimés). Le rapport de sujets non-exprimés à chaque type de proposition correspond donc à une proportion de 9,89 % dans les principales et 7,11 % dans les subordonnées. Retrouve-t-on les mêmes rapports dans les chartes choisies ?

Les 23 chartes regroupent 123 propositions principales et 238 propositions subordonnées, pour un total de 361 propositions verbales. De ces propositions, 24 sujets non-exprimés ont été dénombrés, soit 17 sujets personnels non-exprimés et 7 sujets impersonnels non-exprimés. Le rapport au nombre de propositions s'élève à 6,65 %, ce qui représente 1,24 % de moins que le rapport dans les textes longs. Parmi les 24 sujets non-exprimés, 9 se trouvent dans une des 123 propositions principales (dont 4 sujets impersonnels non-exprimés) et 12 dans une subordonnée (dont 3 sujets impersonnels non-exprimés). Cette distribution dans chaque type de proposition donne des proportions de 7,32 % pour les principales et 4,16 % pour les subordonnées. Un test khi2 montre qu'aucune des différences dans les chiffrages entre textes longs et textes courts n'est statistiquement significative.

La comparaison de la non-expression du sujet dans deux ensembles de textes du même domaine, de la même période et de la même région et qui ne se distinguent que par leur longueur montre donc que cette longueur n'a pas d'impact sur l'expression du sujet. Il serait possible que la longueur ait une influence à une échelle particulière. Pour aller plus loin, nous avons voulu comparer parmi les chartes les textes les plus courts et les plus longs. Nous avons analysé séparément quatre (4) chartes de plus de 450 mots et dix (10) chartes de moins de 200 mots. Leur identité est donnée dans les tableaux suivants (1 et 2) :

Tableau 1. Liste des chartes longues du *Cartulaire*

Pontoise Hôtel Dieu	septembre 1275	555 mots
Vaux de Cernay	28 août 1274	484 mots
Vaux de Cernay	1275	549 mots
Vaux de Cernay	mai 1279	488 mots
		2076 mots

Tableau 2. Liste des chartes courtes du *Cartulaire*

Pontoise Hôtel Dieu	avril 1276	190 mots
Pontoise Hôtel Dieu	juin 1276	173 mots
Pontoise Hôtel Dieu	novembre 1278	167 mots
Pontoise Hôtel Dieu	octobre 1279	147 mots
Saint-Christophe-en-Halatte	Août 1275 et Juin 1276	137 mots
Saint-Christophe-en-Halatte	Novembre 1276	140 mots
Saint-Christophe-en-Halatte	Avril 1278	87 mots

Saint-Christophe-en-Halatte	1277	85 mots
Saint-Christophe-en-Halatte	1279	41 mots
Vaux de Cernay	Juin 1278	165 mots
		1332 mots

Les textes de plus de 450 mots comportent 120 propositions conjuguées. Parmi elles, 115 comptent un sujet exprimé, et 5 un sujet non exprimé (dont 3 sujets impersonnels). Cela représente une proportion de 4,35 % des propositions à sujet nul. Les textes de moins de 200 mots livrent 90 propositions. 84 des sujets sont exprimés, et 6 ne le sont pas (dont 2 sujets impersonnels). La proportion que cela représente est de 7,14 % des propositions avec un sujet nul. La différence de proportion de sujets nuls dans les textes plus longs et les textes courts s'élève à 2,79 % et n'est pas statistiquement significative, à la différence des attentes exprimées voulant que les textes plus courts comportent plus de sujets exprimés.

L'examen comparatif d'un texte long et d'une série de textes courts pour la même époque, la même région et le même domaine montre qu'il n'y a pas de corrélation entre longueur du texte et taux de sujets réalisés. S'il y a une différence minime dans le taux de réalisation de sujets entre le texte coutumier long et les chartes plus courtes en faveur du premier, la différence entre les chartes longues et les chartes courtes se manifeste en faveur des secondes. Autrement dit, à cette étape de notre recherche, on ne constate pas d'influence de la longueur du texte sur la réalisation du sujet. Une autre influence potentielle sur l'expression du sujet est évaluée dans la section suivante.

4 Personne et nombre du sujet

Une des hypothèses récentes sur la disparition du sujet nul se base sur l'idée selon laquelle le pronom de première et de deuxième personne serait exprimé plus précocement que celui de troisième personne dans une perspective historique / diachronique. Selon Detges [4], la raison en serait que ces pronoms seraient plus susceptibles de faire l'objet d'une mise en relief pour des motifs énonciatifs. Cette valeur de mise en relief serait perdue avec un usage répété des pronoms concernés s'intégrant ainsi plus tôt dans le système grammatical. Dans une étude de 2015, Sophie Prévost relève les pronoms de première personne et de troisième personne dans un corpus de 14 textes étiquetés morpho-syntaxiquement allant de 1100 et 1400. Dans les textes littéraires, le taux agrégé de sujets nuls est de 40,84% pour les sujets nuls de première personne et de 72,25% de sujets nuls pour la troisième personne. « Le seul texte qui déroge à cette tendance massive », écrit l'auteure [10 : 14] « est un texte juridique (*Coutumes de Beauvaisis* de P. de Beaumanoir, 1283). Cela est dû à une expression élevée de P3 (elle est identique à celle de P1), ce qui s'explique probablement par la nécessité, dans ce type de texte, d'explicitier plus systématiquement les référents. »

Nous avons donc examiné le taux de réalisation du sujet avec les pronoms personnels selon la personne, nos chartes ayant été sélectionnées en fonction de la représentation de personnes verbales autres que la P3. En effet, intuitivement, les personnes autres que la troisième sont moins fréquentes dans les textes légaux que dans les textes littéraires. Si la personne joue le rôle attendu, dans l'absence de P1 et de P4 on devrait avoir un taux de sujet nul beaucoup plus élevé dans les textes légaux que dans les textes littéraires.

Qu'en est-il dans le corpus réuni pour l'étude ? Le texte coutumier des *Établissements des Capétiens* compte 317 propositions conjuguées. Sur les propositions conjuguées, 278 sont à la troisième personne du singulier ou du pluriel ; 39 comportent une première ou deuxième personne du singulier ou du pluriel. Parmi ces 39 cas de première ou deuxième personne, quel est le taux d'expression du sujet ? L'examen des données est catégorique : aucune forme verbale de première ou deuxième personne n'est liée à un pronom sujet non-

exprimé. Cela confirme l'idée que les pronoms de première et deuxième personne sont plus réalisés (avec un taux de 100%) que les pronoms de troisième personne (à un taux de 71,34%, avec 253 des 278 troisièmes personnes réalisées). Cette observation va donc dans le sens de l'hypothèse formulée par U.Detges [4].

Concernant l'ensemble des chartes choisies, on retrouve 361 propositions conjuguées, dont 263 à la troisième personne du singulier ou du pluriel et 98 – à la première ou deuxième personne du singulier ou du pluriel. Pour ces dernières, le taux d'expression du sujet s'élevait à 93,45 %, dans 89 des 98 occurrences, soit un taux plus faible que celui, catégorique, des *Établissements*. Le taux d'expression du sujet pour les expressions de troisième personne s'élève à 94,29 %, avec 248 des 263 cas. Ces chiffres montrent que dans les *Établissements*, le taux de sujets non-exprimés est plus faible pour les premières et deuxièmes personnes que pour les troisièmes. Cela contredit la tendance observée dans le *Cartulaire* ainsi que l'hypothèse évaluée dans cette section.

Les taux d'expression du sujet qui ne sont pas statistiquement significatifs dans ces textes vont dans des sens imprévus et contradictoires. Les chartes voient moins d'expression du sujet à la P1 (première personne du sg.) et à la P 4 (première personne du pl.). Le texte coutumier donne à voir plus de sujets exprimés de première et deuxième personne que de sujets de troisième personne.

Tableau 3. Taux de sujets non-exprimés pour les premières et deuxièmes personnes

Textes	1PS	1PP	2PS	2PP	Total
<i>Capétiens</i>	0/0	0/17	0/11	0/11	0/39
<i>Cartulaire</i>	3/39 (7,69%)	6/59 (10,17%)	0/0	0/0	9/98 (9,18%)

Ces chiffres permettent également d'évaluer l'idée selon laquelle le nombre du sujet nominal a un impact sur le taux de réalisation du sujet (John Charles Smith, communication personnelle). Le raisonnement est que les personnes plurielles tendent à être morphologiquement marquées, que cette marque contribue à rendre le sujet identifiable, son identifiabilité rendant moins pressante sa réalisation. Cette idée ne semble pas soutenue par les données recueillies. Même si le nombre d'occurrences est faible, la comparaison du pluriel et du singulier des première et deuxième personnes suggère une absence d'impact. D'une part, l'instanciation du sujet à ces personnes dans les *Établissements* fait que le taux de réalisation du sujet est identique au singulier et au pluriel. Dans le *Cartulaire*, il y a une différence réelle mais non significative, entre 92,41 % de réalisation du sujet à la première personne du singulier et 89,83% à la première personne de pluriel. Pour avoir une base comparative plus large, il serait intéressant de comparer la troisième personne du singulier et la troisième personne du pluriel, de même que les groupes nominaux singuliers et pluriels. Nous espérons aborder ces questions dans des travaux ultérieurs.

Cette section évalue l'impact de la personne sur la réalisation du sujet, selon l'hypothèse qu'elle pourrait expliquer la réalisation plus importante du sujet dans les textes légaux, même si les expressions de troisième personne sont intuitivement plus nombreuses dans ces textes que dans les textes littéraires. La vérification du taux de réalisation des sujets montre que dans les *Établissements* ce taux pour les première et deuxième personnes est plus élevé (100%) que le taux général pour la troisième personne (qui est de 92,11%). En revanche, dans les textes choisis du *Cartulaire*, le taux de réalisation des première et deuxième personnes est légèrement plus faible, s'élevant à 93,45 %, alors que, pour la troisième personne, il se trouve à 94,29 %. En résumé, la personne grammaticale n'est pas corrélée au

taux de réalisation du sujet. Nous verrons si une telle corrélation existe avec le statut pronominal ou nominal du sujet.

5 Sujets nominaux

Nous avons cherché à montrer, dans les deux dernières sections, quels étaient les facteurs qui pouvaient expliquer la réalisation du sujet pronominal dans les textes juridiques, Nous avons prouvé que la longueur des textes, la personne et le nombre des pronoms personnels sujets ne pouvaient pas expliquer l'expression plus précoce du sujet pronominal dans les textes juridiques par rapport aux textes littéraires de cette époque. Dans la présente section, nous examinons un facteur supplémentaire qui pourrait signaler des caractéristiques textuelles propres au domaine légal. Il s'agit du rapport des réalisations du sujet dans le cas d'un sujet pronominal et dans le cas d'un sujet nominal.

Des travaux en cours signalés par Alexandra Simonenko [12] semblent suggérer que le taux de réalisation du sujet pourrait dépendre du poids relatif des groupes nominaux et des pronoms dans différents types de texte. L'idée est que, comme le sujet non-exprimé est comparable à un pronom, un ensemble de textes qui comprendrait moins de pronoms qu'un autre aurait donc mécaniquement plus de groupes nominaux et donc un taux de réalisation du sujet plus élevé.

Qu'en est-il des les textes que nous avons réunis ? Ces textes ayant été soumis à une annotation précise (cf. *supra* 3), les résultats se laissent résumer par les chiffres suivants (cf. Tableau 4) : dans les *Établissements*, les sujets lexicaux (S_{Lex}) réalisés en tant que groupe nominal, avec un noyau nominal et un déterminant, représentent 59 des 317 propositions, soit 14,3 % de l'ensemble des sujets. Les sujets pronominaux (S_{Pro}), incluant les 10 impersonnels réalisés (S_I), s'élèvent à 233 occurrences, soit 73,5 % de tous les sujets, exprimés ou non-exprimés. Le compte des sujets non-exprimés (S_{NE}) est de 25 cas, c'est-à-dire comparable à 7,89% aux ratios fournis par Zimmerman [13]. Cette situation est-elle comparable à la situation du *Cartulaire* ? Le statut des sujets y est semblable, sauf pour un point : la proportion de groupes nominaux (S_{Lex}) est deux fois plus élevée, atteignant 29,36 %, soit 59 des 361 propositions. Les pronoms (S_{Pro}) y représentent 61,22 % pour un total de 231 occurrences, et les sujets non-exprimés (S_{NE}) – 24 occurrences – représentent 6,65 % de l'ensemble des sujets.

Tableau 4. Relevés par type de sujet

Textes	S _{Lex}	S _{Pro}	S _I	S _{NE}	Total
<i>Capétiens</i>	59 (14,3%)	223	10	25 (17 S _{NE} et 8 S _{INE})	317
<i>Cartulaire</i>	106 (29,4%)	222	9	24 (17 S _{NE} et 7 S _{INE})	361

Il n'y a donc pas de corrélation entre l'expression pronominale ou nominale du sujet et le taux de réalisation de celui-ci, puisque dans deux ensembles textuels ayant le même taux de réalisation du sujet, les sujets pronominaux exprimés varient du simple au double. Cela rappelle que les caractéristiques textuelles peuvent s'attacher non seulement à un domaine, mais aussi à un sous-ensemble de ce domaine. Pour donner un exemple tiré du travail en cours d'un des auteurs du présent article, le nom propre est abondant dans les comptes-rendus de délibération d'assises (dans le célèbre *Livre rouge d'Eu* par exemple), alors qu'il est à peu près absent des lois coutumières.

Les conclusions de cette étude sont présentées dans la section conclusive qui suit.

6 Discussion conclusive

Les linguistes qui ont travaillé sur la non-expression du sujet dans les textes médiévaux suggèrent que certaines propriétés des textes légaux – la longueur, la personne et le nombre du pronom personnel sujet, ainsi que la catégorie du sujet – pourraient expliquer le taux plus faible de sujets non-exprimés dans ces textes. Quel est l'impact de ces propriétés dans le cadre des textes examinés dans la présente étude?

- La longueur du texte n'est pas significative : *Les Etablissements des Capétiens*, un texte long, possède un taux de sujet nul (7,89 %) proche de celui des textes courts provenant du *Cartulaire* de l'Ile-de-France (6,65 %). Parmi les textes courts, ceux de plus de 200 mots ont un taux proche du texte long (7,14 %), alors que ceux de moins de 200 mots ont un taux plus faible (4,35%), sans que cela soit significatif.

- La personne grammaticale du sujet n'est pas probante : le taux de sujets non exprimés est plus faible pour les premières et deuxièmes personnes dans les *Etablissements des Capétiens* que pour l'ensemble des personnes, alors qu'il est plus élevé dans le *Cartulaire*.

- La catégorie grammaticale du sujet n'est pas significative : le taux de sujets lexicaux est deux fois plus important dans les textes courts que dans le texte long, mais il est sans impact sur le taux global de sujet non-exprimé.

A ceci, on doit ajouter le fait que la différence de sujets non-exprimés entre la principale et la subordonnée n'est pas significative au moment de la perte du sujet non-exprimé.

Aucune des propriétés signalées ne semble avoir joué un rôle important pour l'expression du sujet pronominal. Affranchis des contraintes stylistiques qui affectent les textes littéraires, enregistrant les résultats d'une évolution syntaxique bien avant les autres, les textes juridiques médiévaux seraient alors un meilleur reflet du changement linguistique effectif.

Références

1. L. Balon et P. Larrivée, «L’Ancien français n’est déjà plus une langue à sujet nul. Nouveau témoignage des textes légaux», *JFLS* **26**, 2, 221-237, (2016).
2. D. Capin et P. Larrivée, «La continuité référentielle du sujet nul dans les textes législatifs en français ancien », in E. Dupuis, K. Guilbert, *et al.* (éds), *Actes du Colloque le Choix des Mots*, PUR, (2016, à paraître).
3. D. Capin et P. Larrivée, «La disparition du sujet nul en ancien français et la continuité référentielle», in S. Prévost et B. Fagard (éds), *Le français en diachronie. Dépendances syntaxiques, morphosyntaxe verbale, grammaticalisation*, Berne, Peter Lang, 82-107, (2017).
4. U. Detges, «L’obligatorisation des pronoms sujets en ancien français dans une perspective pragmatique et comparative », *Verbum* **25**, 307-335, (2003).
5. C. Guillot, A. Lavrentiev, B. Pincemin, S. Heiden, «Le discours direct au Moyen Âge : vers une définition et une méthodologie d’analyse», in D. Lagorgette et P. Larrivée (dir.), *Actes du colloque Représentation du Sens Linguistique V*, Chambéry, Éditions de l’Université de Savoie, 17-41, (2014).
6. M. Kytö, «The emergence of American English. Evidence from seventeenth-century records in New England», in R. Hickey (dir.), *Legacies of Colonial English*, Cambridge, Cambridge University Press, 121-152, (2004).
7. W. Labov, *The social stratification of English in New York City*, Washington, Center for Applied Linguistics, (1966).
8. C. Marchello-Nizia, « De So à SV : Vers le sujet obligatoire et antéposé en français, les dernières phases d’un changement », *JFLS*, (2017).
9. C. Marchello-Nizia, «Les plus anciens textes français (IXe –mi XIIe s.) et l’évolution de la syntaxe propositionnelle», Présentation au Colloque Diachro VII, Paris, (2015).
10. S. Prévost, «Diachronie du français et linguistique de corpus : une approche quantitative renouvelée», *Langages* **197**, 23-45, (2015).
11. L. Rello, *Elliphant : A Machine Learning Method for Identifying Subject Ellipsis and Impersonal Constructions in Spanish*, Erasmus Mundus International Master Thesis, Barcelona, (2010).
12. A. Simonenko, B. Crabbe et S. Prévost, «Modelling historical changes as information of theoretic tradeoffs», Présentation au Colloque International SLE, Naples, (2016).
13. M. Zimmermann, *Expletive and Referential Subject Pronouns in Medieval French*, de Gruyter, (2014).