
HOW LEXICAL MERGER CAN DRIVE GRAMMATICALIZATION: THIRD PERSON PRONOUNS FROM LATIN TO OLD FRENCH*

LIEVEN DANCKAERT
CNRS - UNIVERSITÉ DE LILLE

LILIANE HAEGEMAN
GHENT UNIVERSITY

SOPHIE PRÉVOST
CNRS - ENS-UNIVERSITÉ PSL - SORBONNE
NOUVELLE PARIS 3

ABSTRACT The aim of this paper is to elucidate the genesis of Romance third person personal pronouns deriving from Latin *ILLE* and *IPSE*, with special reference to the history of French. Drawing a parallel with the taxonomy of phonological mergers from Labov (1994), we argue that the Late Latin competition between *ILLE* and *IPSE* was resolved through a series of lexical mergers (i.e. the opposite of better known lexical splits). Concretely, we propose that strong personal pronouns (such as French *lui*) arose through merger of *ILLE* and *IPSE*, to the effect that the union of the feature sets of the latter two elements was transferred to the newly formed category. In contrast, weak pronouns (like French *il*) only retain the intersection of the feature sets of *ILLE* and *IPSE*. By creating new functional categories, lexical merger thus acts as a driving force behind grammaticalization.

1 INTRODUCTION

1.1 *Pronouns and determiners from Latin to Romance*

As is well known, Latin did not have definite articles; in contrast, all of the Romance daughter languages do have this category, which in most cases is derived either from the Latin demonstrative pronoun *ILLE*, or, less often, from the intensifying reflexive *IPSE* (key references from the very bulky literature

* Many thanks to the audience at DiGS 21, two anonymous reviewers, and Johanna Wood for much-appreciated feedback.

on this subject include Meader 1901, Trager 1932, Abel 1971, Selig 1992, Giusti 2001, Carlier & De Mulder 2010, Ledgeway 2015, Lüdtke 2015, Ledgeway & Smith 2016; in particular on the remnants of *IPSE*, see Aebischer 1948 and Sornicola 2008, 2012). The grammaticalization of definite articles is arguably related to the afterlife of Latin *ILLE* and *IPSE* in contexts other than that of adnominal determiners, namely third person personal pronouns. Aspects of this development were studied in, among others, Wanner (1987), Sornicola (1998, 2001), Schøsler & Strudsholm (2013), Prévost (2018) and Danckaert & Prévost (submitted). Relatively few studies have been devoted to these two developments in conjunction (but see e.g. Renzi 1997; Vincent 1997; Borsdal Hertzzenberg 2015).

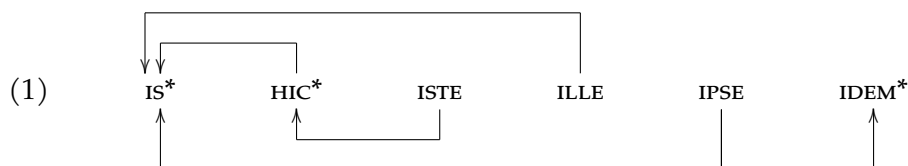
As noted in Cappellaro (2016: 722–723), one question which has received relatively little attention is how exactly true third person personal pronouns (i.e. a category distinct from, and perhaps in opposition to, demonstrative pronouns) came into being in the history of Latin/Romance. At first sight, the relevant development does not seem to be a ‘classical’ case of grammaticalization: for one thing, just as is the case with the development of definite articles, we are not dealing with a lexical category that develops into a functional one. In addition, some of the newly created items, in particular strong personal pronouns (on which, see Section 1.3) are also not phonologically, and perhaps not even semantically, reduced with respect to the pre-existing categories. With respect to this last point, Cappellaro (2016: 723) correctly points out that the change is not simply one from ‘deictic’ to ‘anaphoric’ (pace Renzi 1997): Latin demonstratives could express anaphoric reference, and strong third person pronouns such as present-day Italian *lui* and French *lui* can still be used deictically.

The present paper aims to shed some light on the origins of third person personal pronouns in the Romance language family (with special reference to French): by applying insights from the literature on phonological change to the realm of morphosyntax, we propose that the merger of two competing (in the sense of Kroch 1989, 1994) functional categories, in this case *ILLE* and *IPSE*, can give rise to the creation of new functional material. We suggest that recognizing diachronic processes of this type may deepen our understanding of certain cases of grammaticalization.

1.2 Third person reference in Latin

To overtly express third person reference, speakers of Latin could make use of a series of pronouns which are traditionally labelled ‘demonstratives’ (*IS*, *HIC*, *ISTE*, and *ILLE*); other options include the reflexive intensifier *IPSE* (‘self, the very’) and the identity term *IDEM* ‘the same’, to which we could add what

is known as the ‘connecting relative’ QUI (on which, see [Danckaert 2012](#): 181–185, and references cited there).¹ Importantly, the relative frequencies of these elements do not remain stable over time. [Väänänen \(1981: 121\)](#) summarizes the development from Classical to Late Latin as follows (forms marked with an asterisk do not survive in Romance; the arrows shows which item took the place of which):



In Section 3 of this paper, we will offer corpus evidence confirming that this basic picture is correct, supplementing it with some new observations about the syntactic distribution of the bare (i.e. non-adnominal) nominative forms of these pronouns.²

As pointed out in [Danckaert \(2012: 217–223\)](#), there is every reason to believe that the traditional view that the elements listed in (1) are not personal pronouns is indeed correct. [Cardinaletti & Starke \(1999a: 284\)](#) list a number of diagnostics to distinguish (overt) personal pronouns from demonstratives. One of them is that the former are subject to Principle B of the Binding Theory, and the latter to Principle C. In other words, only personal pronouns can appear in the c-command domain of a co-referential category.³ For example, in Latin we never find demonstrative pronouns acting as the subject of a complement clause, in a configuration whereby the relevant demonstrative is coreferent with the matrix subject. In Old French on the other hand, this construal can readily be found:

- (2) (a) *Oliver_i sent [qu' il_i est a mort nasfret]*
 Oliver feel.PRS.3SG that he be.PRS.3SG to death wounded
 ‘Olivier feels that he is wounded to death.’
 (*Roland*, ca. 1100, v. 1965)

¹ For general discussion of pronominal reference in (the history of) Latin, the interested reader is referred to [Pieroni \(2010\)](#) and [Bordal Hertenberg \(2015\)](#).

² Given its overall low token frequency, in what follows we will not further take into account *IDEM*.

³ This claim is in need of some qualification, as in certain circumstances it does in fact seem possible for demonstratives to be c-commanded by a co-indexed constituent (see e.g. [Hinterwimmer & Brocher 2018](#) (on German)). However, this observation does not affect our argument, which is that the absence of third person pronouns in ‘Principle B’ configurations can be taken to mean that the relevant items were not personal pronouns.

- (b) *et il_i dist [que si fera il_i volentiers por*
and he feel.PRS.3SG that si do.FUT.3SG he willingly to
fere a mon seignor Gauvain compaignie]
do.INF to my Lord Gauvain company
‘and he says that he will do it willingly to keep Lord Gauvain
company.’

(*Queste del Saint Graal*, ca. 1225, p. 161b)

To the best of our knowledge, in the entire Latinity the counterpart of (2) involving ILLE, IS, HIC or ISTE is never attested: though strictly speaking an *argumentum ex silentio*, this observation does suggest that these elements are demonstratives, not personal pronouns.

Note that the non-attestation of embedded demonstratives c-commanded by a co-indexed category is not due to the fact that Latin was a *pro*-drop language. In languages that allow for zero anaphora and which have third person personal pronouns, the latter are often dispreferred in the c-command domain of coreferential expressions, *pro* always being the default option in those environments. Only when *pro*-drop is for some (pragmatic) reason dispreferred is an overt (strong) personal pronoun felicitous (for relevant discussion, see e.g. Lobo & Silva 2015: 197–199). We can illustrate this point by means of the Spanish examples in (3)–(4) (from Larson & Luján 1989: 6, their (27)–(28)). The question-answer pair in (3) starts with a long-distance object question; therefore, the embedded object in (3b) can be said to be a narrow focus. In the same clause, the subject is part of the presupposition: for this reason *pro* is strongly preferred over an overt pronoun. In (3b), the hash-sign indicates pragmatic infelicity, rather than ungrammaticality.

- (3) (a) *Qué cree Juan_i [que Ø_i obtendrá en ese*
what think.PRS.3SG Juan that obtain.FUT.3SG in this
concurso]?
contest
‘What does John think he will get in the competition?’
 (b) *Juan_i cree [que Ø_i/#el_i ganará el premio]*
Juan think.PRS.3SG that he win.FUT.3SG the award
‘John thinks that he/#HE will win the award.’

This exchange is minimally different from that in (4): here the a-example is a long-distance-subject question, and as a result, the embedded subject in the corresponding b-example acts as a new information focus. When the intended coreference relations obtain, this interpretation requires the overt lexicalization of the embedded pronoun:

- (4) (a) *Qujén_j cree Juan_i [que t_j ganará el premio]?*
 who think.PRS.3SG Juan that win.FUT.3SG the award
 ‘Who does John think will win the award?’
 (b) *Juan_i cree [que #Ø_i/el_i ganará el premio]*
 Juan think.PRS.3SG that he win.FUT.3SG the award
 ‘John thinks that #he/HE will win the award.’

We can conclude that there is no principled ban in null subject languages on strong personal pronouns appearing in the c-command domain of a coreferential element. As a result, we can maintain that the absence of this pattern in third person environments in the Latin corpus suggests that the language only had demonstrative third person pronouns. This conclusion is confirmed by the fact that the first and second person singular pronouns *ego* ‘I’ and *tu* ‘you’ can in fact be c-commanded by a coreferential category:⁴

- (5) *“Vere”, inquit, “Ø_i dico uobis, [quod ego_i*
in.truth say.PRF.3SG say.PRS.1SG you.DAT.PL that I.NOM
audiui uoces angelorum in caelis
hear.PRF.1SG voices.ACC angels.GEN in heavens.ABL
“Sanctus” in laude Domini proclamantes]”.
Sanctus in praise.ABL Lord.GEN proclaiming.NOM.M.PL
 ‘He said: “I tell you in truth that I personally heard angels’
 voices in heaven proclaiming “Sanctus” to praise the Lord”.’
 (Greg. Tur., *vit. patr.* 16.2, late sixth c. CE)
- (6) *quid enim Ø_i non dicis, [quod tu_i benigne*
why PRT not say.PRS.2SG that you.NOM.SG kind.ADV
iam in ea operatus es]?
already in DEM.ABL.F.SG worked.NOM.M.SG be.PRS.2SG
 ‘Why don’t you say that you have already worked kindly in
 her?’

(Greg. M., *in cant.* 44, late sixth c. CE)

Finally, and very importantly, the coreference restriction just illustrated does not hold for *IPSE*, no doubt because the original function of this element is not

⁴ The examples in (5) to (7) were all drawn from later authors, who frequently use finite complement clauses, in environments where earlier writers would have used an *AcI* (i.e. an infinitival clause with an accusative subject). It is not easy to find similar examples from Early or Classical Latin, but we can safely assume that this is due to the scarcity of finite, and non-obviative, complement clauses in earlier texts, not to the properties of Early and Classical Latin personal pronouns. On the Early, Classical and Late Latin corpus samples that we will investigate in this study, see Sections 2.1–2.2.

that of a (deictic or anaphoric) pronoun; rather, in an example like (7) *ipse* presumably acts as a secondary predicate co-indexed with a phonologically null subject (on IPSE, see e.g. Bertocchi 2000 and Pieroni 2007; see Gast 2006 for cross-linguistic parallels). In any event, there is no reason to assume that IPSE was at any point subject to Principle C. As we will argue below, this point is quite important for the later development of Romance third person pronouns.

- (7) *unde probamus quod \emptyset_i iam nouerat*
 whence prove.PRS.1SG that already know.PLPFV.3SG
[quia \emptyset_i ipse_i baptizaret in spiritu
 that REFL.NOM.M.SG baptise.IPFV.SBJV.3SG in spirit.ABL
sancto]?
 holy.ABL
 ‘How can we prove that he already knew that he would himself
 baptise in the Holy Spirit?’
 (Aug., in euang. Ioh. 124.8, ca. 407 CE)

1.3 Aims and structure of the paper

The main goal of the paper is to elucidate how two types of Romance third person personal pronouns came into being. We will focus in particular on (historical) French, but our analysis should apply to other Romance varieties as well. For reasons of space, we will only consider subject pronouns.⁵ More precisely, adopting the three-way typology of pronouns developed in Cardinaletti & Starke (1999b), we will be concerned with strong and weak subject pronouns, and thus not with subject clitics such as those found in many Northern Italian dialects (see e.g. Poletto 2000 and Poletto & Tortora 2016). Examples of the two types of pronouns that we will be concerned with include French *lui* and Italian *lui* (which are both strong), and French *il* and Italian *egli* (which are weak).⁶

The core idea to be worked out is that the competition (Kroch 1989, 1994) between Late Latin *ILLE* and *IPSE* was resolved through various types of lexi-

⁵ We will also not be concerned with the related development of the rise of definite articles in Romance, which as mentioned also derive from *ILLE* and *IPSE*.

⁶ Cardinaletti & Starke (1999b) list the following differences between strong and weak pronouns: (i) only the former can be coordinated (with another strong pronoun, or with a full lexical DP), (ii) weak but not strong pronouns may be phonologically reduced, (iii) only strong pronouns can appear as complements of prepositions, and (iv) for prosodic reasons, weak pronouns have to surface in dedicated functional projections in the clause, whereas strong pronouns are allowed to occur VP-internally as well as in right- and left-peripheral positions. The authors also propose that there are semantic differences between the two categories: we refer to the original paper, as well as to Cardinaletti & Starke (1999a) for full discussion.

cal merger of these two items, rather than through extinction or specialization of one element, and the concomitant generalization of the other. In particular, drawing upon the typology of phonological mergers proposed in Labov (1994), we propose that *ILLE* and *IPSE* underwent ‘merger by expansion’ to yield strong personal pronouns, and that a second operation of ‘merger by approximation’ resulted in the creation of weak pronouns; thirdly, the phonological shape of a newly formed pronoun is always the result of ‘merger by transfer’ (which can either go towards *ILLE* or to *IPSE*). Our analysis thus suggests that at least in some cases, the need to resolve competition between two (and perhaps even more) lexical items may drive morphosyntactic change, and thus grammaticalization. If on the right track, our analysis also entails that the parallels between phonological and lexical change (at least as far as changes affecting the feature composition of functional vocabulary items is concerned) may be deeper than previously thought.

Before we proceed, we would like to emphasize that the genesis of third personal personal pronouns is – though not unrelated to – independent of another major development, namely the loss of *pro*-drop in certain parts of the *Romània*. The independence of the two developments clearly comes to the fore when we consider that all Romance languages have third person personal pronouns, but only some (most notably Gallo-Romance) varieties lost the possibility of having referential null subjects. However, in Section 5 we briefly speculate on the syntax of *pro*-drop in Old French.

The paper is structured as follows. In Section 2, we present the results of a series of corpus studies documenting the frequency and distribution of pronominal subjects in Latin and Old French. The descriptive generalizations emerging from these corpus studies constitute the empirical foundation of the following discussion. We pay special attention to the situation in Old French, where the differential behaviour of main and embedded clauses makes it somewhat difficult to properly interpret the data. In Section 3, we turn to the pragmatic status of *ILLE* and *IPSE* in Late Latin, pointing out certain distributional and interpretive differences between these two third person reference devices. In Section 4, we first elaborate on various types of phonological mergers, as discussed in Labov (1994) and related work. We go on to apply this apparatus to the development of third person reference markers in the history of Latin/Romance. In Section 5, we briefly touch upon an issue that is arguably related to – but as mentioned above, clearly distinct from – the rise of weak personal pronouns in the history of French, namely the demise of *pro*-drop. Section 6 concludes the paper.

2 SUBJECT PRONOUNS FROM LATIN TO OLD FRENCH: THE LAY OF THE LAND

The aim of this section is to formulate a number of empirical generalizations concerning the frequency and syntactic distribution of overt personal and demonstrative pronouns acting as the syntactic subject of finite clauses. When quantifying the rate of these items, we will use the incidence of clauses with a referential null subject as a baseline, but we repeat that our focus is never on (the syntax of) *pro*-drop per se: rather, what we are interested in is a diachronic comparison between pronominal third person subjects on the one hand, and first and second person subjects on the other.

Although the broad outlines of the data that we are concerned with are generally known, there is at present no reliable corpus study to document the finer details of the relevant developments. As we will show, the corpus data do in fact reveal some rather surprising contrasts between (Late) Latin and Old French. In particular, our results show that the two languages behave very differently with respect to the parameters ‘Clause type’ ([\pm root clause]), and to a lesser extent ‘Person’ (operationalized as [\pm third person subject]). The basic generalizations are the following. First, in Late Latin overt subject pronouns (personal pronouns in the case of first and second person, demonstratives and *ipse* for the third person) are generally more common in main than in embedded clauses. In Old French, we find exactly the opposite. Second, in Latin, Person effects only play a relatively minor role when it comes to predicting the incidence of overt pronominal subjects. This is most notably the case in Late Latin, where overt demonstrative pronouns and overt personal pronouns occur at similar rates (at least when acting as a grammatical subject). In Old French on the other hand, overt first and second person pronouns are generally more common than their third person counterparts. Crucially, closer inspection reveals complex interactions between the factors ‘Clause type’ and ‘Person’ in Old French, to the effect that this last generalization does not hold in embedded clauses.

2.1 *Early and Classical Latin*

To gauge the incidence of overt subject pronouns in Early and Classical Latin we analysed a corpus consisting of the seven text samples listed in Table 1.^{7,8} One important factor we took into account when putting together this corpus

⁷ With the term ‘Early Latin’, we refer to the language of all Latin texts dating from before 100 BCE. The label ‘Classical Latin’ is reserved to the period of *ca.* 100 BCE until *ca.* 120 CE (roughly speaking from Cicero until Suetonius). In other words, the term ‘Classical’ is not used with any stylistic connotation.

⁸ All texts for the corpus study reported on in Section 2.1 were taken from the Brepols database, available online at www.brepolis.net.

Author	Text	Date (approx.)	N of pronominal subjects
Plautus	<i>Amphitruo</i>	185 BCE	1068
Terence	<i>Andria</i>	165 BCE	1027
Cicero	<i>In Verrem</i> II.1	70 BCE	1032
Livy	<i>Ab Urbe condita</i> 21	5 CE	658
Seneca	<i>Ad Lucilium</i> 1–25	50 CE	605
Petronius	<i>Satyricon</i>	60 CE	1001

Table 1 Corpus of Early and Classical Latin texts.

is whether or not a given text contains enough tokens of all Person-Number combinations. For this reason we left aside among other things technical treatises, which typically contain little or no second persons (singular and plural), nor many first person verbs (in particular plurals, and to a lesser extent first person singulars).

Only finite clauses with a pronominal (null or overt; personal or demonstrative) subject were taken into account. For third person contexts, we included the pronouns *IS*, *HIC*, *ISTE*, *ILLE* and *IPSE*, but we took care to only count the latter as a genuine subject pronoun if the ‘reinforced reflexive’ reading illustrated in (7) is not available.⁹ We excluded all clauses which appear as a second, third etc. conjunct and which contain what appears to be a null subject coreferential with that of the first conjunct. The reasons for this is that the relevant tokens are structurally ambiguous, in that they are amenable to a phrase structure analysis with ‘low’ coordination, i.e. of two (extended) VPs out of which a subject has been extracted in an across-the-board manner. Clauses with an impersonal finite verb such as *oportet* ‘it befits’ and *licet* ‘it is allowed’ were also excluded from our dataset, as in this environment there is no alternation between null and overt subjects. The same methodological choices were made for the two other data sets that we will report on shortly (with the proviso that the set of overt subject pronouns is of course different in Old French: see Section 2.3 for further clarifications).

Our findings are summarized in Figure 1, where we have plotted the coefficients of a (fixed effects) logistic regression, modeling the likelihood for

⁹ For the time being, we will not further differentiate between these five items, but we refer to Section 3 for additional discussion of their relative rate and distribution.

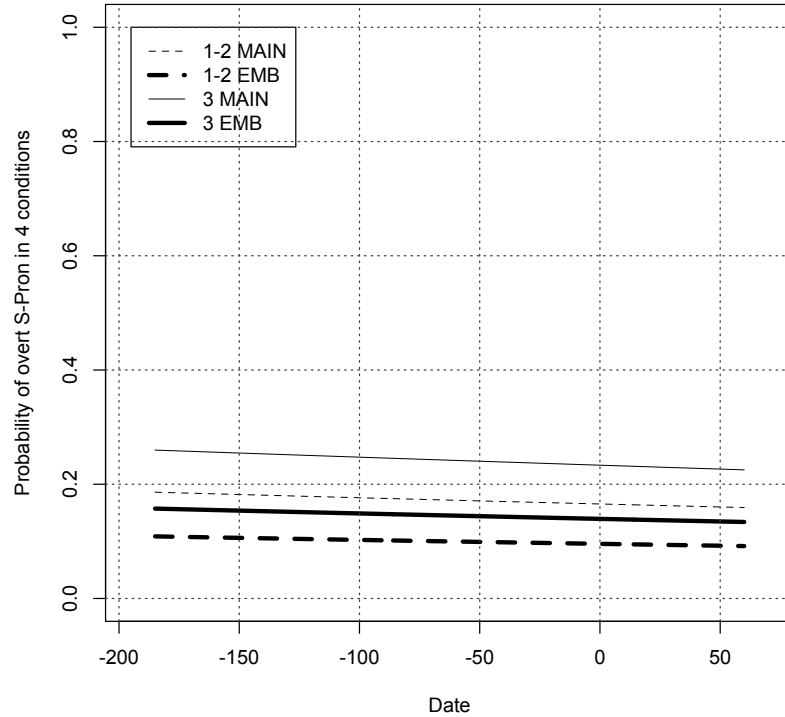


Figure 1 Rate of overt pronominal subjects in Early and Classical Latin (*ca.* BCE - 60 CE): Effects of 'Clause type' and 'Person'.

a pronominal subject to be expressed overtly (we converted the estimated log odds to a probability scale from 0 to 1). The horizontal axis documents the time dimension, which in this cases ranges from 185 BCE until 60 CE. As elaborated on above, we took into account distinctions between main and embedded clauses, and between first/second and third persons (i.e. overt personal pronouns vs. demonstratives).

The basic generalizations that emerge from the statistical analysis can be summarized as follows. First, overt subjects are overall more likely to be found in third person contexts (cf. the solid lines in Figure 1; estimated log odds ($= \beta$) = 0.429, $p < 0.001$), and less likely in embedded clauses (cf. the thicker lines in Figure 1; $\beta = -0.628$, $p < 0.001$). In addition, there is a significant interaction between the factors 'Date' and 'Clause type', which says that the observed discrepancy between main and embedded clauses becomes stronger over time ($\beta = 0.003$, $p < 0.001$). A main effect of 'Date' turned out not to be significant, nor did any of the other interaction terms between the

afore-mentioned predictors. In any event, as can be deduced from Figure 1, though significant, the size of the three effects just mentioned is overall relatively small.

Note, in passing, that it is not immediately obvious how the relative preponderance of overt third person subjects is to be understood.¹⁰ It is tempting to interpret this effect in terms of the Accessibility Marking Scale from (Ariel 1990: 73), according to which demonstratives tend to refer to less accessible referents than (personal) pronouns. Assuming that less accessible referents need more explicit coding, the relatively high frequencies of bare demonstrative subjects would indeed be expected. On the other hand, as we will see later, pronominal subjects in Old French main clauses are sensitive to Person effects which are quite different from the ones discussed here. This suggests that Person-asymmetries related to the frequency of occurrence of various types of pronouns are not always to be explained in pragmatic terms only (or at least not in terms of Accessibility). Clearly, what is needed at this point is a detailed study of the pragmatic properties of various (overt and covert) pronouns and their antecedents. From a methodological point of view, the multifactorial, corpus-based approach pursued in Bordal Hertenberg (2015) seems like a good starting point to design such a large-scale study, preferably with some diachronic depth. Needless to say, such an enterprise goes well beyond the scope of this study (but see Section 3 of this paper, as well as Danckaert & Prévost submitted for some initial remarks).

Turning to the observed Clause type effect, nothing in what we have said thus far makes us predict that there should be any difference between main and embedded clauses, but as we will see, this factor is particularly important when it comes to modeling the transition from Latin to French.

2.2 Merovingian Latin

Let us now turn to pronominal subjects in Late Latin. We analysed a sample of seven Merovingian and Early Carolingian Latin texts dating from between 520 and 835 CE, which were all composed in the geographical area roughly corresponding to present-day France. We used the text editions from the *Monumenta Germaniae Historica* (viz. the *Scriptores rerum Merovingicarum*, cf. Krusch & Levison 1884–1920). A full description of this sub-corpus is given in Table 2.

¹⁰ We are indebted to a reviewer for very insightful discussion of the issues discussed in this paragraph.

Text	Date (approx.)	N of pronominal subjects
<i>Vita Genovefae</i>	6 th c.	259
Gregory of Tours, <i>Historiae</i> 6	580–585	856
Fredegar, book 4	658–660	568
<i>Vita Eligii</i> , book 1	ca. 670–680	453
<i>Liber Historiae Francorum</i>	727	674
<i>Vita Trudonis</i>	784–791	408
<i>Vita Galli Walahfrido</i>	833	631

Table 2 Corpus of Late Latin texts.

As above, we ran a logistic regression on this dataset with ‘Date’, ‘Clause type’ and ‘Person’ as fixed effects. The results we obtained are summarized in Figure 2.

In the model from which we extracted the coefficients plotted in Figure 2, only an interaction term between Person and Date ($\beta = 0.002$, $p = 0.04$) and one between Clause type and Date ($\beta = -0.002$, $p = 0.03$) came out as (borderline) significant at the 0.05 level. As to the factor Person, the obtained result would mean that originally first and second person subjects were more likely to be overt, but that towards the early ninth century this is no longer true (compare the trajectories of the two dotted and the two solid lines in the graph: they both stay parallel to each other, but the two pairs cross each other around 650 CE). Similarly, the two thicker lines and the two thinner ones also cross each other, which is indicative of an interaction effect between Clause type and Date to the effect that, in our later texts, it is no longer generally the case that overt pronominal subjects are preferred in main clauses. However, in addition to the p-values reported above not being far below the 0.05 threshold, the estimated log odds (i.e. the β -coefficients) are particularly low, which means that the relevant significant effects are only very small. In sum, there are some reasons to doubt that the effects plotted in Figure 2 are reliable. To have a better idea of what is going on in this sub-corpus, we ran a more powerful mixed-effects analysis, without a time variable, but with the individual source texts as random effects. The strongest such model we could build is one with random intercepts for ‘Text’ and random slopes for the two levels of the factor ‘Clause type’, as well as a fixed effect for Clause type. A fixed

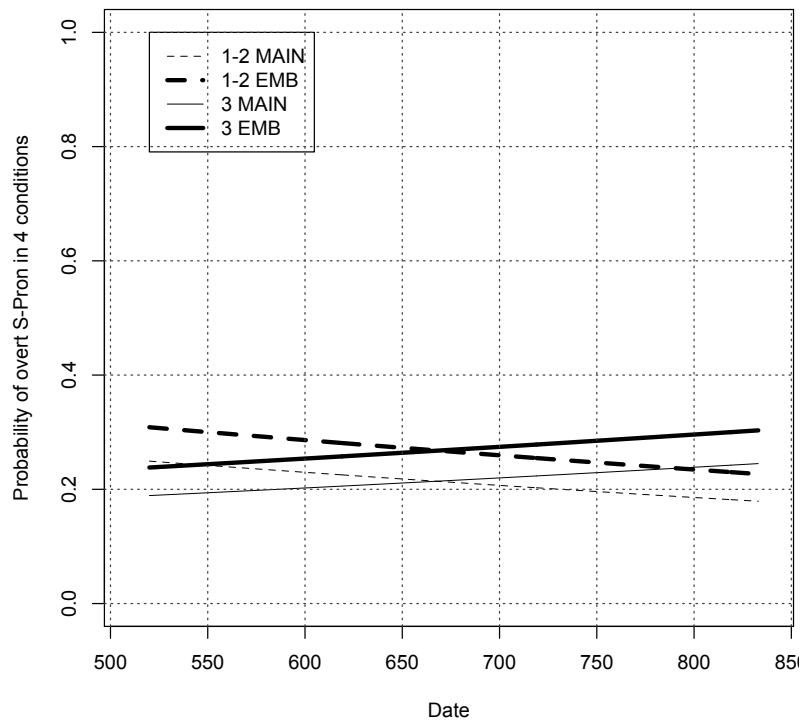


Figure 2 Rate of overt pronominal subjects in Late Latin (*ca.* 550–833 CE): Effects of ‘Clause type’ and ‘Person’.

main effect for Person did not turn out to be significant, neither when considered in isolation, nor when added to the model with significant Clause type-effects. The estimated log odds for the factor ‘Clause type’ indicates that overt pronominal subjects are dispreferred in embedded clauses ($\beta = 1.247$, $p < 0.001$).

Let us then compare the results of the two Latin sub-corpora. The role of the factor ‘Clause type’ seems to remain pretty constant, in that overt pronominal subjects are always more frequent in main clauses. In contrast, the role of ‘Person’ seems to change in the transition from Classical to Late Latin: the once robust contrast between first and second person on the one hand, and third person on the other, seems to have evaporated, or at least to have become very weak, in the Merovingian period. We can interpret this observation to mean that at least from a quantitative perspective (i.e. as far as rates of usage are concerned), the various third person pronouns start to behave more and more like true personal pronouns, a result which is, of course, entirely

expected (given what we know about later developments). In Section 3, we will further elaborate on the discourse interpretation of in particular *ILLE* and *IPSE* in Late Latin (but as mentioned we will not attempt a systematic comparison between the conditions of usage associated with third and first/second person pronominal subjects).

2.3 (Early) Old French

Finally, to study subject pronouns in Old French, we analysed 11 texts dating from *ca.* 1000 to 1250, which are listed in Table 3.¹¹ As to the overt subject pronouns, we took into account only personal pronouns (for a complete list of possible forms that Old French subject pronouns could take, see [Burdant 2000](#): 408), leaving aside the various newly formed demonstratives (on which, see [Guillot 2015](#), [Guillot-Barbance 2017](#)). To maximally ensure that the Old French data can be compared with the results obtained from our Latin corpus samples, here too we did not take into account clauses with an impersonal verb. As shown in Figure 3, in the first centuries of the second millennium the rate of overt subjects changes very quickly.

Text	Date (approx.)	N of pronominal subjects
<i>Passion</i>	<i>ca.</i> 1000	304
<i>Vie de Saint Alexis</i>	<i>ca.</i> 1050	471
<i>Chanson de Roland</i>	<i>ca.</i> 1100	600
<i>Eneas</i>	<i>ca.</i> 1155	607
Beroul, <i>Tristan</i>	late 12 th c.	540
Chretien de Troies, <i>Yvain</i>	1180	580
R. de Clari, <i>La Conqueste de Constantinople</i>	after 1205	434
<i>Aucassin et Nicolette</i>	late 12 th /early 13 th c.	601
<i>Queste del Saint Graal</i>	1225	580
J. Renart, <i>Roman de la rose</i> <i>ou de G. de Dole</i>	<i>ca.</i> 1228	591
J. Sarrasin, <i>Lettre à Nicolas Arrode</i>	1249	92

Table 3 Corpus of (Early) Old French texts.

¹¹ For this part of the corpus study we used texts from the *Base de Français Médiéval* (<http://bfm.ens-lyon.fr/>).

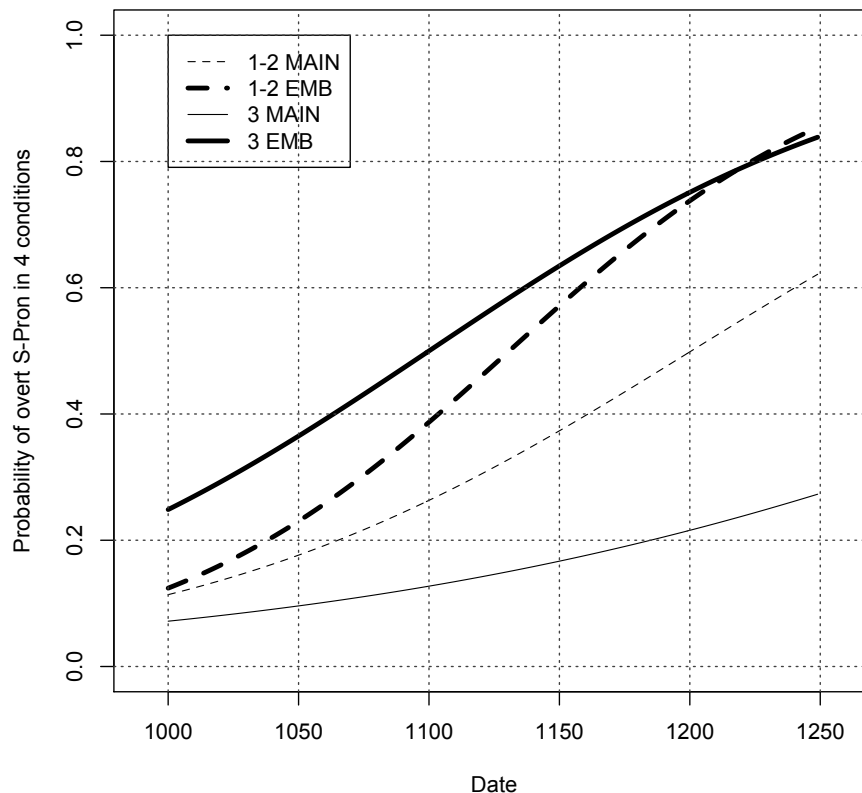


Figure 3 Rate of overt pronominal subjects in (Early) Old French (*ca.* 1000–1250): Effects of ‘Clause type’ and ‘Person’.

Most importantly, and in strong contrast with Latin, our data show that in Old French overt pronominal subjects are generally preferred in embedded clauses (cf. the thicker lines in Figure 3), and dispreferred in main clauses. This confirms a claim that goes back to at least Foulet (1930) and Franzén (1939). In addition, embedded clauses initially favour overt third person pronouns. However, after *ca.* 1200, Person distinctions no longer play a role in embedded clauses. In main clauses, overt third person pronouns are always less frequent than their first and second person counterparts, and this tendency becomes stronger over time. To give a more precise idea of the data summarized in Figure 3, in Table 4 we give the full details of the logistic model on the basis of which this plot was produced.

We would like to interpret the discrepancy between Old French main and embedded clauses to mean that the language developed a new mechanism to license null subjects in main clauses, which explains why the demise of *pro*-drop is delayed in this syntactic environment. We will offer some speculation

	Estimate	Std. error	p-value
(Intercept)	-12.282893	1.237960	<2e-16 ***
Person_3	3.377843	1.407830	0.016425 *
Clause type_EMB	-4.610373	1.324668	0.000501 ***
Date	0.010231	0.001057	<2e-16 ***
Person_3:Clause type_EMB	1.355303	0.147920	<2e-16 ***
Person_3:Date	-0.003885	0.001200	0.001200 **
Clause type_EMB:Date	0.004709	0.001128	2.98e-05 ***

Table 4 Overt and covert subject pronouns in Old French over time: Effects of ‘Person’ and ‘Clause type’.

about the nature of this new licensing mechanism in Section 5. In contrast, we take it that embedded null subjects are the (disappearing) remnants of the type of null subject that was available in Latin, but crucially, we still have to explain why in Old French embedded clauses, overt third person personal pronouns are initially more frequent than first and second person pronouns, despite the fact that the former were not inherited as such from the Latin mother language.

2.4 Taking stock: the explananda

In order to answer the main research question of this paper (how did Romance third person personal pronouns come into being?), we need to take into account a number of considerations. First, as documented in great detail in Cappellaro (2016), in many modern Romance varieties the paradigm of strong (or tonic, to use the author’s terminology) third person personal pronouns contains descendants of both *ILLE* and *IPSE*. We would like to follow Cappellaro (2016: 725) (see also Cappellaro 2018: 125–134) in interpreting this state of affairs to mean that in Late Latin, *ILLE* and *IPSE* competed to express third person reference. In addition, as mentioned in the introduction the emergence of third person personal pronouns cannot easily be treated as a run-of-the-mill case of grammaticalization, in particular not when strong pronouns are considered: not only is the change one from one functional (rather than lexical) category to another, it is also not clear in which sense the newly formed strong personal pronouns are semantically, structurally and/or phonologically reduced with respect to their demonstrative/reflexive ances-

tors. In contrast, the genesis of weak (atonic) pronouns in Old French constitutes a more likely case of ‘classical’ grammaticalization, as the newly created category clearly is reduced (with respect to demonstrative as well as strong personal pronouns), both structurally and phonologically (see [Cardinaletti & Starke 1999a,b](#)). A final question that remains to be answered is whether Old French weak pronouns were derived directly from Latin *ILLE*, or rather from (Old French) strong personal pronouns: we will come back to this issue in Section 4.

3 DISCOURSE PROPERTIES OF LATE LATIN *ILLE* AND *IPSE*

Let us now have a closer look at the interpretive properties of *ILLE* and *IPSE* in Late Latin. As has often been pointed out, despite there being a clear sense in which the two items ‘compete’ for the same function in the relevant period, *ILLE* and *IPSE* are certainly not fully equivalent (see among others [Selig 1992](#); [Vincent 1997](#)). Some new corpus findings suggest that this is indeed correct: concretely, *ILLE* gradually develops a particularly strong preference for appearing in main clauses, whereas in Late Latin the most frequent third person pronoun in embedded clauses is *IPSE*. The relevant quantitative data we observed in the two subcorpora introduced in Sections 2.1 and 2.2 are summarized in Table 5. Among other things, the data in this table confirm that, in Late Latin, *ILLE* is by far the most frequently used third person pronominal reference device. We also indicate whether or not the distribution of a given pronoun is significantly skewed towards main or embedded clauses; the expected frequencies used for the χ^2 -tests were calculated on the basis of the total number of (finite) main and embedded clauses with an (overt or otherwise) pronominal subject in each period. The significant results appear in the shaded cells.

	Early and Classical Latin			Late Latin		
	N main	N embedded	χ^2 , p =	N main	N embedded	χ^2 , p =
HIC	107	49	0.009	42	28	0.132
ILLE	70	39	0.228	236	22	<0.001
IPSE	27	23	0.655	52	39	0.229
IS	50	37	0.991	1	0	NA
ISTE	44	45	0.119	4	3	1

Table 5 Distribution of Latin third person subject pronouns.

Late Latin *ILLE*'s preference for main clauses suggests that in the relevant period, the pronoun typically functions as a syntactic topic, a type of constituent whose distribution is known to be restricted in embedded clauses (Emonds 1970; Hooper & Thompson 1973; Ross 1973; Haegeman 2012). This observation can be taken to mean that *ILLE* took over some of the functions of its erstwhile competitors *HIC* and *IS*, which in Danckaert (2012: 207–257) were shown to readily function as left-peripheral topics in Classical Latin main clauses, but not in embedded ones.¹² Another factor conspiring against the usage of *ILLE* in embedded clauses may have been the fact that the pronoun was subject to Principle C (cf. Section 1.2).

In our Late Latin corpus, bare *ILLE* in subject function typically indicates a topic shift, very often indicating a change of speaker (compare Frascarelli & Hinterhölzl (2007)'s notion of 'Shifting Topic'). Consider, for example, the little stretch of narrative discourse in (8), which features three discourse referents, namely *Wiomadus*, *Egidius*, and a group of Franks whose identity is not further specified. In this example, none of the three occurrences of subject *ILLE* shifts the topic away from one that was newly introduced by means of a full NP, but rather, an *ILLE*-topic either succeeds another *ILLE*-topic, or an instance of zero anaphora:

- (8) \emptyset_i hortabatur Egidio_j [aliquos Francos]_k
 encourage.IPFV.3SG Egidius.ACC some.ACC Franks.ACC
dolose oppremere. ille_j audiens
 cunning.ADV oppress.INF DEM.NOM.M.SG hearing.NOM.M.SG
consilium eius_i acrius coepit
 advice.ACC DEM.GEN.M.SG sharp.COMP.ADV begin.PRF.3SG
oppremere eos_k illi_i uero in timore
 oppress.INF DEM.ACC.M.PL DEM.NOM.M.PL PRT in fear.ABL
ac seditione uersi, iterum consilium a
 and uproar.ABL turned.NOM.M.PL again advice.ACC from
Wiomado_j expetierunt, qualiter \emptyset_k facere
 Wiomadus.ABL ask.PRF.3PL how do.INF
deberent. at ille_j dixit
 have_to.IPFV.SBJV.3PL but DEM.NOM.M.SG say.PRF.3SG
eis_k: [...].
 DEM.DAT.M.PL

¹² The data in Table 5 also suggest that that *HIC* was originally especially fit to mark discourse topics (i.e. a well-known main clause phenomenon), given the early preference for bare nominative forms of *HIC* to occur significantly more frequently in main clauses than in embedded domains. Early and Classical Latin *is*, on the other hand, seems to have been less narrowly specialized.

‘He (viz. Wiomadus) encouraged Egidius to cunningly oppress some of the Franks. He followed his advice, and started to oppress them relentlessly. Out of fear, the Franks then revolted and asked Wiomadus for advice as to what they should do. He answered to them: [...].’

(*Liber Historiae Francorum* 7, 727 CE)

Alternatively, *ILLE* is a more neutral anaphoric marker indicating topic continuity (as in (9)), a usage which may correspond to the notion of ‘Familiarity Topic’ from Frascarelli & Hinterhölzl (2007).¹³

- (9)

[_{DP} <i>Stilla,</i>	[_{CP} <i>quem</i>	<i>comitem</i>	<i>superius</i>	
star.NOM	which.ACC.F.SG	comet.ACC	higher.ADV	
<i>nominavi]]_i,</i>	<i>apparuit,</i>	<i>ita ut</i>	<i>in circuitu</i>	
call.PRF.1SG	appear.PRF.3SG	so that	in surrounding.ABL	
<i>eius_i</i>	<i>magna</i>	<i>nigrido</i>	<i>esset;</i>	<i>et</i>
DEM.GEN.F.SG	great.NOM	blackness.NOM	be.IPFV.SBJV.3SG	and
<i>illa_i,</i>	<i>tamquam</i>	<i>se</i>	<i>in foramen</i>	<i>aliquod</i>
DEM.NOM.F.SG	as.if	REFL.ACC	in opening.ACC	some.ACC
<i>posita,</i>	<i>inter</i>	<i>tenebras</i>	<i>relucebat.</i>	
placed.NOM	between	darkness.ACC	shine.IPFV.3SG	

‘A star, which I earlier called a comet, appeared, in such a way that everything around it was cloaked in thick blackness. And this star shone a bright light through the darkness, as if it were placed in some opening in the sky.’

(Greg. Tur., *Hist.* 6.14, ca. 595 CE)

On the other hand, subject *IPSE* is either used with its original Classical Latin force (roughly meaning ‘he [...] himself, she [...] herself etc.’), or it indicates that some familiar discourse participant is explicitly being contrasted with the current discourse topic (as in (10), where the topic changes from the churches destroyed by Clovis to Clovis himself), yielding a configuration which is more strongly contrastive than that of the topic shifting usage of *ILLE* illustrated in (8).

¹³ Another typical function of Late Latin *ILLE* often discussed in the literature is the one coined *definite Erstnennung* (‘definite first mention’) in Selig (1992), which often involves (an NP modified by) *ILLE* acting as the head of a relative clause. However, at least as far as bare (i.e. non-adnominal) *ILLE* in subject function is concerned, this usage is not particularly frequent in the Late Latin corpus we have investigated for this study.

- (10) *Eo tempore [multae ecclesiae]_i a*
DEM.ABL.N.SG time.ABL many.NOM churches.NOM by
Chlodouecho_j exercitu depredatae sunt.
Clovis.GEN army.ABL looted.NOM.F.PL be.PRS.3PL
Erat=que ipse_j tunc fanaticus et
be.IPFV.3SG=and he.NOM.M.SG then fanatic.NOM and
paganus.
pagan.NOM
‘At that time many churches were looted by Clovis’ army. He
was a fanatic, and a pagan.’
(Liber Historiae Francorum 7, 727 CE)

We conclude that although there may have been some functional overlap between *ILLE* and *IPSE*, the two pronouns were clearly not freely interchangeable. With this established, we return to the later developments of Romance pronominal paradigms.

4 LEXICAL MERGERS IN PRONOMINAL PARADIGMS

As documented in great detail in Labov’s work on sound changes in American English (see Labov 1994 for a comprehensive synthesis), many instances of phonological change involve the merger of two previously distinct phonemes into one ‘atomic’ category. Apparently less frequently, phonological change goes in the other direction, yielding a phonological split. In the realm of lexical change, lexical splits are of course very well documented. However, much less has been said about the merger of two distinct lexical items, be it content (open class) or function (closed class) words. The guiding idea behind our analysis is that lexical merger is exactly what happened to *ILLE* and *IPSE*.¹⁴

4.1 A typology of phonological mergers

Labov (1994: 321–323) discusses three “radically different” mechanisms of merger in phonology (for seminal discussion, see also Trudgill & Foxcroft 1978). Focusing, in particular, on vocalic mergers, Labov first distinguishes ‘merger by expansion’, which he describes as a process through which “[t]he

¹⁴ In addition to that, we certainly also have to assume that one or more lexical splits took place: whereas both *ILLE* and *IPSE* could act either as adnominal determiners or as free-standing pronouns, Romance definite articles and third person personal pronouns are clearly distinct. It is in all likelihood necessary to assume another split inside the category of personal pronouns, which would account for the fact that Romance object clitics invariably derive from *ILLE* not *IPSE* (Vincent 1997).

phonetic range of the new phoneme is roughly equivalent to the union of the range of the two phonemes that merged". As an example, Labov cites the development of tense /oh/ and lax /o/ phonemes in Eastern Pennsylvania, as studied in [Herold \(1990\)](#): whereas older speakers use the two variants in clearly distinct phonetic contexts, younger speakers randomly interchange them. A second type is 'merger by approximation', defined as "the gradual approximation of the phonetic targets of two phonemes until they are nondistinct", whereby the resulting phoneme "may show a mean value intermediate between those of the original two." In the latter case, the output of this process is distinct from both input phonemes.¹⁵ According to the description of [Lennig \(1978\)](#), this is what was happening to the vowels /a/ and /ɑ/ in the variety of French spoken in Paris in the mid-1970s. A third and final type of change is called 'merger by transfer', which is "a unidirectional process in which words are transferred gradually from one phonemic category to another. As a rule, it is not consistent with a result that shows an intermediate phonetic form." Labov refers to [Milroy \(1980\)](#) for discussion of this type of merger in the history of Belfast English (on the differences between merger by approximation and merger by transfer, see also [Trudgill & Foxcroft 1978](#)).

For reasons of space, we cannot here further elaborate on the various case studies which have led to the establishment of this typology: for full discussion, we refer to Labov's original work, and also to [Babel, McAuliffe & Haber \(2013\)](#), who provide illustrations which nicely visualize the relevant processes (see in particular Figure 1 on p. 2 of their article).

4.2 *Lexical mergers of pronouns*

In a nutshell, we propose that once language learners started to analyse *ILLE* and *IPSE* as competing for the same function, namely that of a relatively neutral third person reference device, the two items were merged together. In order to resolve the interpretive clash between the discourse properties not shared by both elements, merger happened in such a way that the resulting lexeme (let us call it *LUI*) retained features of both *ILLE* and *IPSE*. More precisely, we can say that *LUI* ends up being endowed with – roughly speaking – the union

¹⁵ According to [Labov \(1994: 321\)](#), merger by approximation may also result in a phoneme whose realizations do, in fact, have the same mean value as one of the input phonemes: an example of this would come from Greek, where no less than eight originally distinct phonemes were eventually merged into one front vowel /i/ (cf. [Labov 1994: 229](#)). As pointed out in [Johnson \(2007: 4\)](#), this subtype of merger by approximation may formally be very similar to merger by transfer: the differences between the two processes reside in, among other things, the role of sociolinguistic factors (which are stronger in the case of merger by transfer, and above the level of conscious awareness) and the rate of change (merger by transfer being slower, cf. [Labov 1994: 323](#)).

of the feature sets of *ILLE* and *IPSE*: the relevant process is therefore an instance of merger by expansion. This development affected the whole of Romance, witness the fact that all Romance varieties have third person personal pronouns distinct from demonstratives (Cappellaro 2016). In addition, in Old French (and other varieties which developed weak pronouns), a second lexical item (*IL*) was created which ended up retaining only features that were originally shared by *ILLE* and *IPSE* (say roughly the intersection of the feature sets of those two items). This last process can be characterized as an instance of merger by approximation. Finally, in all cases, the phonological shape of a newly created pronoun was determined through merger by transfer; witness the fact that despite the considerable amount of variation in Romance pronominal paradigms, no mixed forms seem to exist containing one or more phonological segments inherited from *ILLE*, and one or more that derive from *IPSE*. Let us then discuss the genesis of strong *LUI* and weak *IL* in more detail, starting with the former.

For the sake of concreteness, and abstracting away from features which are not essential to the analysis (such as ‘Number’), let us assume the following sets of properties to characterize Late Latin *ILLE* and *IPSE*:

- | | |
|---|-------------|
| (11) (a) [+ 3 rd Person, + Topic continuity, + Deixis, DP] | <i>ILLE</i> |
| (b) [+ 3 rd Person, + Contrast, – Deixis, ϕ P] | <i>IPSE</i> |

Let us briefly comment upon these feature sets. First, we take it that the ‘+ 3rd Person’ property is obvious; suffice it to add that we could also assume that ‘3rd Person’ is in fact the absence of Person (following a tradition going back to Benveniste 1966: 251–257). In a similar vein, we also don’t want to exclude that the notion of ‘Person’ can (or has to) be decomposed in a number of independent features, such as [\pm Author] and [\pm Participant] (Harbour 2016). Second, for the characterizations ‘+ Topic continuity’ (*ILLE*) and ‘+ Contrast’ (*IPSE*), we refer to the discussion in Section 3. Next, the ‘+ Deixis’ property expresses the fact that *ILLE* can (but need not be) used deictically: we take it that the structural correlate of this property is a functional layer which in the case of *ILLE* is always projected, and can optionally be ‘activated’, e.g. by means of an interpretable Deixis-feature.¹⁶ For *IPSE*, the ‘– Deixis’ setting captures the exclusively anaphoric nature of the element, i.e. the fact that it is referential, but not compatible with a deictic use. Finally, we would like to relate the presence or absence of a Deixis-projection to a difference *qua* categorial status of the two pronouns under discussion: concretely, adopting

16 For a recent analysis of the structural representation of (spatial) deixis, see Lander & Haegeman (2018).

the terminology from Déchaine & Wiltschko (2002), we characterize *ILLE* as a DP, and *IPSE* as a ϕ P.¹⁷ Analysing *ILLE* as a DP correctly captures the fact that this item (in its bare use) has the same syntactic distribution as DPs projected by a lexical noun: it can only be used in argument and adjunct positions, never as a predicate. In this respect it differs from *IPSE*, which we argued above (cf. (7)) can act as a (secondary) predicate. We therefore consider *IPSE* to be a ϕ P, i.e. a nominal category with ϕ -features but no outer D-shell.

When *ILLE* and *IPSE* undergo merger by expansion, a new element *LUI* is created with the feature set [+3rd Person, + Emphatic, + Deixis, DP], which comes about in the following way. First, the ‘+3rd Person’ property is simply inherited from both merging elements. Second, lexical merger of an element endowed with a feature ‘+ Topic continuity’ and one with a ‘+ Contrast’ property clearly cannot give rise to an element which (always) expresses both of these notions simultaneously, as this would yield an interpretive clash. Instead, we take it that what arises is an item that is merely emphatic, not narrowly specialized to express topicality, focality, contrast etc. This gives the correct result, as *LUI* can be used to express all of those nuances, whenever *pro* (or *IL*) is for some reason pragmatically infelicitous. Third, recall that we take the setting ‘+ Deixis’ to be correlated with the presence of a functional projection. It suffices that this category is present in one of the two merging items (in this case *ILLE*) in order for it to be inherited by *LUI*. This correctly accounts for the fact that strong personal pronouns can be used anaphorically as well as deictically (cf. Section 1.1). Finally, by the same logic as that applied to the presence of a Deixis-projection, merger-by-expansion of a full DP (which itself contains a ϕ P) and a structurally smaller bare ϕ P is expected to yield a DP. This accounts for the fact that Romance (strong) personal pronouns can no longer occur in the predicative construal exemplified in (7) (which we assume to be different from cases of subject doubling by means of a strong pronoun in e.g. Italian; see Belletti 2005 for a possible analysis of the latter phenomenon). Finally, the binding properties of *LUI* can be derived as follows. Recall from Section 1.2 that Romance strong personal pronouns are subject to Principle B and not Principle C of the Binding Theory. Cru-

17 Déchaine & Wiltschko (2002) propose a three-way topology of pronouns which is highly reminiscent of that of Cardinaletti & Starke (1999b) introduced in Section 1.2. Briefly put, they distinguish DPs, ϕ Ps and NPs, whereby ϕ Ps contain an NP-layer, and DPs a ϕ P (and an NP). For reasons of space, we cannot here attempt a detailed comparison between the two proposals: suffice it to say that despite some similarities, the two theories differ in a number of crucial respects: for one thing, Déchaine & Wiltschko (2002: 409, fn. 1) point out that their categorization cuts across distinctions between phrasal pronouns, clitics and agreement morphemes. See also Déchaine & Wiltschko (2002: 439) for a brief comparison between the two proposals, and for a possibility to analyse Cardinaletti & Starke’s strong, weak and clitic pronouns as three distinct types of their own ϕ P.

cially, this property was already present in (Late) Latin *IPSE* (cf. (7)): we propose that it was retained by the newly formed pronouns. Specifically, we conceive of Principles C and B as standing in a subset-superset relation, rather than as being opposites.¹⁸ Concretely, the distribution of elements subject to Principle C ('X should never be c-commanded by a coreferential category') is more restricted than that of elements obeying Principle B ('X should not be *locally* c-commanded by a coreferential category'). Put differently, the range of syntactic domains that an R-expression (a lexical DP, or a demonstrative) can appear in is a proper subset of that in which a personal pronoun can appear. In the case at hand, the newly created strong pronoun *LUI* can appear in the set of binding configurations that is the union of the sets of environments that *ILLE* and *IPSE* can appear in, which happens to coincide with the range of environments that *IPSE* can appear in.

Turning then to Old French *IL*, we can say that this element emerged as a *bona fide*, unmarked personal pronoun, which is devoid of *ILLE*'s topic feature and of the contrastive semantics typical of *IPSE*. Instead, it only expresses the notion 'third person', which is arguably the one semantic feature shared by both *ILLE* and (anaphoric) *IPSE*. This development would be an instance of merger by approximation, with the proviso that there was already some functional overlap between the two merging elements. The binding properties of weak pronouns (which just as their strong counterparts obey Principle B) can be derived from a proposal made in [Cardinaletti & Starke \(1999b: 192\)](#), to the effect that weak pronouns per definition lack the structural layer where indexical features are hosted: this effectively disqualifies weak pronouns as R-expressions.¹⁹

Importantly, we assume that Old French strong *LUI* and weak *IL* were created around the same time, and by this token that weak pronouns are not derived from strong personal pronouns. The main advantage of this line of analysis is that it can account for one of the more puzzling corpus findings reported in Section 2.3, namely that despite being an innovation, overt third person subject pronouns are initially more frequent than overt first and second person pronouns (in embedded clauses, that is). Concretely, we interpret this last observation to mean that the creation of a paradigm of weak personal pronouns was set in motion in third person contexts. A plausible alter-

¹⁸ We thank an anonymous reviewer for urging us to clarify this issue.

¹⁹ Note that the same result cannot be derived through set theoretic computations (i.e. by following a logic similar to that applied to the binding properties of *LUI*): the intersection between a set and its proper subset being equal to that proper subset, we would incorrectly predict that weak pronouns should have the same binding properties as *ILLE*, and thus be subject to Principle C, contrary to fact. It seems reasonable to assume that the structure proposed by [Cardinaletti & Starke \(1999b\)](#), along with its semantic correlates, takes primacy.

native would be to say that the creation of a full set of weak subject pronouns is a secondary development, which took place after the category strong *LUI* had been innovated. However, this analysis has the drawback that it leaves the ‘Person’ effect present in the Old French corpus unaccounted for. Note also, that it is certainly not the case that weak pronouns are simply phonologically weaker versions of strong pronouns; witness the fact that French weak *il* is derived from the Latin nominative form *ille*, and strong *lui* (indirectly) from the dative *illi*. If, on the other hand, we assume that the genesis of strong and weak pronouns derived from *ILLE* corresponds to two independent but simultaneous processes, the observed time-lag between the spread of overt weak pronouns in third person and non-third person environments is, in fact, predicted. Weak first and second person pronouns come into being later, perhaps through analogy with weak *IL*.

4.3 *Lexical merger and grammaticalization*

Given the analysis detailed in the previous section, we can say that by resolving morphosyntactic competition, lexical merger effectively acts as a ‘catalyst’ for grammaticalization. When the process involved is merger by expansion, the outcome is a novel functional vocabulary item which is not reduced with respect to its historical sources (in fact it may actually be bigger). On the other hand, merger by approximation gives rise to a more prototypical case of grammaticalization, with the loss of functional features and/or phonological segments. In addition to a number of generally accepted structural processes underlying, or even driving, grammaticalization, such as upward reanalysis (Roberts & Roussou 2003) and economy-based constraints like the Head Preference Principle and Late Merge Principle from van Gelderen (2011: 13–14), lexical merger may thus be an additional mechanism which can give rise to the creation of new functional categories. Importantly, in contrast with the other processes mentioned, the primary locus of change would be the lexicon, rather than phrase structure syntax.

5 SOME NOTES ON *PRO-DROP* IN OLD FRENCH

Before we conclude, we would like to briefly touch upon two issues related to the loss of *pro-drop* in Old French (a development which as we pointed out earlier is not to be conflated with the genesis of weak personal pronouns in the language). For recent discussion of this much-debated topic, and in particular the relation between the erosion of agreement morphology and the demise of zero anaphora, we refer to Zimmermann (2014) and especially Simonenko, Crabbé & Prévost (2019), and to the references cited there.

First, there does not seem to be any reason to conclude that Old French was a partial null subject language (in the sense of [Holmberg 2005, 2010](#); [Holmberg, Nayudu & Sheehan 2009](#); see also [Barbosa 2019](#); [Roberts 2019: 207–216](#)): the hallmark of such a system, viz. the possibility for a (main clause) third person singular null subject to receive a generic interpretation (corresponding to that of English *one* and French *on*), does not seem to be present in Old French ([Buridant 2000: 409, §326](#)). In addition, as pointed out in [Danckaert & Prévost \(submitted\)](#), there is no evidence that the presence of a c-commanding coreferential category in the immediately superordinate clause favours the occurrence of embedded null subjects, which again suggests that despite the observed Person effects, Old French was not a partial null subject language (*contra* [Ingham 2018](#)).

Secondly, we would like to offer some speculation about null subjects in Old French main clauses, whose productivity we hypothesized earlier is to be ascribed to an innovation which must have taken place after the Late Latin period. According to the influential Government and Binding analysis of [Adams \(1987\)](#), main clause null subjects in Old French are licensed by V-to-C movement: in particular, the finite verb was argued to license *pro* under Government. We will here briefly explore an alternative to this idea, which does not refer to the now abandoned notion of Government. With [Poletto \(2000\)](#) and [Poletto & Tortora \(2016\)](#), we can assume that in the high functional field four projections are available which are related to subject agreement (but which are independent of subject agreement morphology on finite verbs). These projections can optionally be lexicalized overtly: this is famously the case in many Northern Italian dialects. The two lowest of the relevant functional heads, which are located just below the lower edge of the CP-domain, encode second person singular and third person (singular and plural) (for an overview, see [Poletto 2000: 38](#)). Interestingly, there is evidence that in Old French main clauses, it is subjects of exactly these kinds that are most likely to be left unexpressed. The results of a logistic regression modeling the influence of the factors Person and Number on the expression of pronominal subjects are summarized in Table 6, where the level ‘first person singular’ is set as a baseline. Note that we lumped together singular and plural third persons. In addition, given that for certain Person-Number combinations (in particular first and second person plural) our token counts are unfortunately rather low, estimates are averaged over all texts in our corpus: in other words, we did not take into account the diachronic dimension, nor any random effects associated with individual texts (as a result, the figures in Table 6 may have to be interpreted with some caution).

	Estimate	Std. error	p-value
(Intercept)	-0.31415	0.08618	0.000267 ***
2SG	-0.57005	0.28600	0.046243 *
3SG+PL	-1.18675	0.09961	<2e-16 ***
1PL	0.06526	0.21474	0.761218
2PL	-0.15875	0.16218	0.327629

Table 6 Person and *pro*-drop in Old French declarative main clauses.

Let us assume that verb movement in Old French main clauses targets a position in the low left periphery, say Fin, which sits above the two lowest subject agreement projections from [Poletto \(2000\)](#), but below the two highest ones (on Fin as a target for verb movement, see [Haegeman 1996](#); [Poletto 2002](#); [Wolfe 2019](#)). We can then assume that on its way to Fin, the verb passes through, and thereby ‘activates’ the lower regions of the subject agreement field: this would correctly account for the prevalence of third person and second person singular null subjects in Old French main clauses. Null subjects with different person-number specifications could either be generated by the inherited agreement-based *pro*-drop grammar (which there is at this point no reason to believe declined faster (or slower) in main than in embedded clauses), or perhaps by a grammar involving V-movement to Force, i.e. to a position much higher than the entire subject agreement field. As pointed out in [Poletto \(2018\)](#), an analysis that links V-to-C movement to the licensing of null subjects predicts that rates of *pro*-drop should be similar in declarative and interrogative main clauses (which does not in fact seem to be the case in Old Italian; for relevant discussion, see also [Cognola & Walkden 2019](#)). In future research it will be interesting to test whether effects similar to those summarized in Table 6 can also be observed in interrogative main clauses in Old French.

6 CONCLUSION

In this paper we have considered how Romance third person personal pronouns came into being, a question which in the literature has received relatively little attention. On the assumption that this development does not bear the signature of a prototypical case of grammaticalization, we suggested that the two most prominent Late Latin lexical items expressing third person pronominal reference, to wit *ILLE* and *IPSE*, were merged together to yield true (tonic) personal pronouns, endowed with the union of the feature sets of both

input items. We extended this analysis to account for the rise of weak subject pronouns in Gallo-Romance: we suggested that this development also involved lexical merger of the same two items, but differently from the previous process, only the intersection of the feature sets of *ILLE* and *IPSE* was preserved. Our main theoretical claim is that in addition to other processes, such as upward reanalysis and phrase-to-head reanalysis, lexical mergers may also play a role in driving the creation of new functional material. In future research, it will be interesting to explore whether mergers of lexical items are perhaps more widespread than previously assumed.

REFERENCES

- Abel, Fritz. 1971. *L'adjectif démonstratif dans la langue de la Bible: étude sur la formation des systèmes déictiques et de l'article défini des langues romanes*. Tübingen: Niemeyer.
- Adams, Marianne. 1987. From Old French to the theory of pro-drop. *Natural Language & Linguistic Theory* 5. 1–32.
- Aebischer, Paul. 1948. Contribution à la protohistoire des articles *ille* et *ipse* dans les langues romanes. *Cultura Neolatina* 8. 181–203.
- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Babel, Molly, Michael McAuliffe & Graham Haber. 2013. Can mergers-in-progress be unmerged in speech accommodation? *Frontiers in Psychology* 4. Article 653, 1–14.
- Barbosa, Pilar. 2019. *pro* as a minimal nP: Toward a unified approach to *pro*-drop. *Linguistic Inquiry* 50. 487–526.
- Belletti, Adriana. 2005. Extended doubling and the VP periphery. *Probus* 17. 1–35.
- Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.
- Bertocchi, Alessandra. 2000. *Ipse* as an intensifier. In Gualtiero Calboli (ed.), *Papers on Grammar V*, 15–30. Bologna: CLUEB.
- Bordal Hertenberg, Mari Johanne. 2015. *Third person reference in Late Latin: Demonstratives, definite articles and personal pronouns in the Itinerarium Egeriae*. Berlin: Mouton De Gruyter.
- Buridant, Claude. 2000. *Grammaire nouvelle de l'ancien français*. Paris: SEDES.
- Cappellaro, Chiara. 2016. Tonic pronominal systems: Morphophonology. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 722–741. Oxford: Oxford University Press.
- Cappellaro, Chiara. 2018. Genesis and diachronic persistence of overabundance: Data from Romance languages. In Antje Dammel, Matthias Eitelmann & Mirjam Schmuck (eds.), *Reorganising grammatical variation: Di-*

- achronic studies in the retention, redistribution and refunctionalisation of linguistic variants*, 119–148. Amsterdam: Benjamins.
- Cardinaletti, Anna & Michal Starke. 1999a. Responses and demonstratives. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, 273–290. Berlin: Mouton De Gruyter.
- Cardinaletti, Anna & Michal Starke. 1999b. The typology of structural deficiency: A case study of the three grammatical classes. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, 145–233. Berlin: Mouton De Gruyter.
- Carlier, Anne & Walter De Mulder. 2010. The emergence of the definite article: *Ille* in competition with *ipse* in Late Latin. In Kristin Davidse, Lieven Vandelanotte & Hubert Cuyckens (eds.), *Subjectification, intersubjectification and grammaticalization*, 241–275. Berlin: Mouton De Gruyter.
- Cognola, Federica & George Walkden. 2019. *pro*-drop in interrogatives and declaratives: A parallel study of Old High German and Old Italian. *Linguistik Online* 100. 95–140.
- Danckaert, Lieven. 2012. *Latin embedded clauses: The left periphery*. Amsterdam: Benjamins.
- Danckaert, Lieven & Sophie Prévost. submitted. The rise of obligatory subject pronouns. In Anne Carlier, Lieven Danckaert, Céline Guillot-Barbance, Maria Selig & Jasper Vangaeve (eds.), *Bridging the gap between Late Latin and Early Old French*, Berlin: Walter de Gruyter.
- Déchaine, Marie-Rose & Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33. 409–442.
- Emonds, Joseph. 1970. *Root and structure-preserving transformations*: MIT dissertation.
- Foulet, Lucien. 1930. *Petite syntaxe de l'ancien français*. Paris: Champion.
- Franzén, Torsten. 1939. *Etude sur la syntaxe des pronoms personnels sujets en ancien français*. Uppsala: Almqvist & Wiksells.
- Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form: Generalizations across languages*, 87–116. Amsterdam: Benjamins.
- Gast, Volker. 2006. *The grammar of identity: Intensifiers and reflexives in Germanic languages*. London: Routledge.
- van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the Language Faculty*. Oxford: Oxford University Press.
- Giusti, Giuliana. 2001. The birth of a functional category. In Guglielmo Cinque & Giampaolo Salvi (eds.), *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, 157–171. Amsterdam: Elsevier.

- Guillot, Céline. 2015. From Old French and Middle French to Contemporary French. In Konstanze Jungbluth & Federica Da Milano (eds.), *Manual of deixis in Romance languages*, 558–580. Berlin: Mouton De Gruyter.
- Guillot-Barbance, Céline. 2017. *Le démonstratif en français: Etude de sémantique grammaticale diachronique (9ème-15ème siècles)*. Leuven: Peeters.
- Haegeman, Liliane. 1996. Verb second, the split CP and null subjects in early Dutch finite clauses. *Geneva Generative Papers* 4. 133–175.
- Haegeman, Liliane. 2012. *Adverbial clauses, Main Clause Phenomena and the composition of the left periphery*. Oxford: Oxford University Press.
- Harbour, Daniel. 2016. *Impossible persons*. Cambridge, MA: The MIT Press.
- Herold, Ruth. 1990. *Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania*: University of Pennsylvania dissertation.
- Hinterwimmer, Stefan & Andreas Brocher. 2018. An experimental investigation of the binding options of demonstrative pronouns in German. *Glossa: a journal of general linguistics* 3(1). 77, 1–25.
- Holmberg, Anders. 2005. Is there a little *pro*?: Evidence from Finnish. *Linguistic Inquiry* 533–564.
- Holmberg, Anders. 2010. Null subject parameters. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in Minimalist theory*, 88–124. Cambridge: Cambridge University Press.
- Holmberg, Anders, Aarti Nayudu & Michelle Sheehan. 2009. Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63. 59–97.
- Hooper, Joan & Sandra Thompson. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4. 465–497.
- Ingham, Richard. 2018. Topic, Focus and null subjects in Old French. *Canadian Journal of Linguistics* 63. 242–263.
- Johnson, Daniel Ezra. 2007. *Stability and change along a dialect boundary: The low vowels of Southeastern New England*: University of Pennsylvania dissertation.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1. 199–244.
- Kroch, Anthony. 1994. Morphosyntactic variation. *Chicago Linguistics Society* 30(2). 180–201.
- Krusch, Bruno & Wilhelm Levison (eds.). 1884–1920. *Monumenta Germaniae Historica: Scriptores rerum Merovingicarum (8 vols.)*. Hannover/Leipzig: Hahn.

- Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.
- Lander, Eric & Liliane Haegeman. 2018. The nanosyntax of spatial deixis. *Studia Linguistica* 72. 362–427.
- Larson, Richard & Marta Lujàn. 1989. Emphatic pronouns. Unpublished manuscript, Stony Brook University New York and University of Austin, Texas, available at <https://linguistics.stonybrook.edu/faculty/richard.larson/files/publications/larson-lujan89.pdf> (last accessed 19 March 2021).
- Ledgeway, Adam. 2015. Varieties in Italy. In Konstanze Jungbluth & Federica Da Milano (eds.), *Manual of deixis in Romance languages*, 75–139. Berlin: Mouton De Gruyter.
- Ledgeway, Adam & John Charles Smith. 2016. Deixis. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 879–896. Oxford: Oxford University Press.
- Lennig, Matthew. 1978. *Acoustic measurement of linguistic change: The modern Paris vowel system*. University of Pennsylvania dissertation.
- Lobo, Maria & Carolina Silva. 2015. Resolving pronoun ambiguity in European Portuguese. In Jiyoung Choi, Hamida Demirdache & Oana Lungu (eds.), *Language acquisition at the interfaces: Proceedings of GALA 2015*, 196–215. Newcastle upon Tyne: Cambridge Scholars Press.
- Lüdtke, Jens. 2015. From Latin and Vulgar Latin to Romance languages. In Konstanze Jungbluth & Federica Da Milano (eds.), *Manual of deixis in Romance languages*, 537–557. Berlin: Mouton De Gruyter.
- Meader, Clarence. 1901. *The Latin pronouns is, hic, iste, ipse: A semasiological study*. New York: Macmillan.
- Milroy, James. 1980. Lexical alternation and the history of English: Evidence from an urban vernacular. In Elizabeth Closs Traugott, Rebecca Labrum & Susan Shepherd (eds.), *Papers from the fourth International Conference on Historical Linguistics*, 355–362. Amsterdam: Benjamins.
- Pieroni, Silvia. 2007. *Ipse*: Relationships with grammatical functions and person. In Gérald Purnelle & Joseph Denooz (eds.), *Ordre et cohérence en latin*, 97–110. Genève: Droz.
- Pieroni, Silvia. 2010. Deixis and anaphora. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New perspectives on historical Latin syntax, vol. 3, constituent syntax: Quantification, numerals, possession, anaphora*, 389–502. Berlin: Mouton De Gruyter.
- Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford: Oxford University Press.

- Poletto, Cecilia. 2002. On V2 and V3 sequences in Rhaetoromance. In Sjeff Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic microvariation*, Available at <http://www.meertens.nl/books/synmic/pdf/poletto.pdf> (last accessed 19 March 2021).
- Poletto, Cecilia. 2018. The twists and shakes of pro drop: On the licensing of null topics in Old Italian varieties. Paper presented at DiGS 20, University of York; slides available at https://www.york.ac.uk/media/languageandlinguistics/documents/conferences/digs20/Poletto_slides_DiGS20.pdf (last accessed 19 March 2021).
- Poletto, Cecilia & Christina Tortora. 2016. Subject clitics. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 772–785. Oxford: Oxford University Press.
- Prévost, Sophie. 2018. Increase of pronominal subjects in Old French: Evidence for a starting-point in Late Latin. In Anne Carlier & Céline Guillot-Barbance (eds.), *Latin tardif, français ancien: Continuités et ruptures*, chap. 171–200. Berlin: Mouton De Gruyter.
- Renzi, Lorenzo. 1997. Fissione di lat. ILLE nelle lingue romanze. In Günter Holtus, Johannes Kramer & Wolfgang Sieward (eds.), *Italica et Romanica: Festschrift für Max Pfister zum 65. Geburtstag, Vol. 2: Morphologie, Syntax, Wortbildung; Grammatikographie*, 7–18. Tübingen: Niemeyer.
- Roberts, Ian. 2019. *Parameter hierarchies and Universal Grammar*. Oxford: Oxford University Press.
- Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A Minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
- Ross, John. 1973. The Penthouse Principle and the order of constituents. In Claudia Corum, Thomas Smith-Stark & Ann Weiser (eds.), *You take the high node and i'll take the low node: The differences between main and subordinate clauses*, 397–422. Chicago, IL: Chicago Linguistic Society.
- Schösler, Lene & Erling Strudsholm. 2013. Preservation, modification, and innovation. paradigmatic reorganisation of the system of personal pronouns – from Latin into Modern Italian. In Kristen Jeppesen Kragh & Jan Lindschouw (eds.), *Deixis and pronouns in Romance languages*, 49–68. Amsterdam: Benjamins.
- Selig, Maria. 1992. *Die Entwicklung der Nominaldeterminanten im Spätlatein: Romanischer Sprachwandel und lateinische Schriftlichkeit*. Tübingen: Niemeyer.
- Simonenko, Alexandra, Benoît Crabbé & Sophie Prévost. 2019. Agreement syncretisation and the loss of null subjects: Quantificational models for Medieval French. *Language Variation and Change* 31. 275–301.

- Sornicola, Rosanna. 1998. Tra tipologia e storia: I pronomi soggetto e le colonie gallo-italiche. In Giovanni Ruffino (ed.), *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza: Centro di studi filologici e linguistici siciliani, Università di Palermo, 18-24 settembre 1995*, 639–658. Tübingen: Niemeyer.
- Sornicola, Rosanna. 2001. Diacronia e pancronia nella sintassi e semantica dei pronomi soggetto tra latino e lingue romanze. In Valeria Viparelli (ed.), *Ricerche linguistiche tra antico e moderno*, 29–66. Napoli: Liguori.
- Sornicola, Rosanna. 2008. Sul problema dei resti di *ipse* nella Romània. In Romano Lazzeroni, Emanuele Banfi, Giuliano Bernini, Marina Chini & Giovanna Marotta (eds.), *Diachronica et synchronica: Studi in onore di Anna Giacalone Ramat*, 537–562. Pisa: Edizioni ETS.
- Sornicola, Rosanna. 2012. Multifunzionalità di *ipse* in latino e polimorfismo degli esiti romanzi. In Frédérique Biville, Marie-Karine Lhommé & Daniel Vallat (eds.), *Latin vulgaire, latin tardif IX: Actes du IXe colloque international sur le latin vulgaire et tardif, Lyon, 2-6 septembre 2009*, 347–362. Lyon: Maison de l'Orient et de la Méditerranée.
- Trager, George. 1932. *The use of the Latin demonstratives (especially ILLE and IPSE) up to 600 A.D., as the source of the Romance article*. New York: Institute of French Studies.
- Trudgill, Peter & Tina Foxcroft. 1978. On the sociolinguistics of vocalic mergers: Transfer and approximation in East Anglia. In *Sociolinguistic patterns in British English*, 69–79. London: Edward Arnold.
- Väänänen, Veikko. 1981. *Introduction au latin vulgaire (troisième édition revue et augmentée)*. Paris: Klincksieck.
- Vincent, Nigel. 1997. The emergence of the D-system in Romance. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters of morphosyntactic change*, 149–169. Cambridge: Cambridge University Press.
- Wanner, Dieter. 1987. *The development of Romance clitic pronouns: From Latin to Old Romance*. Berlin: Mouton De Gruyter.
- Wolfe, Sam. 2019. *Verb Second in Old Romance*. Oxford: Oxford University Press.
- Zimmermann, Michael. 2014. *Expletive and referential subject pronouns in Medieval French*. Berlin: Walter de Gruyter.

Lieven Danckaert
STL - Savoirs, Textes, Langage, UMR 8163
Rue du Barreau BP 60149
59653 Villeneuve-d'Ascq Cedex
France
lieven.danckaert@univ-lille.fr
<https://pro.univ-lille.fr/lieven-danckaert>

Liliane Haegeman
Vakgroep Taalkunde (UGent)
Blandijnberg 2
9000 Gent
Belgium
Liliane.Haegeman@UGent.be
[https://research.flw.ugent.be/
nl/liliane.haegeman](https://research.flw.ugent.be/nl/liliane.haegeman)

Sophie Prévost
LATTICE - Langues, Textes, Traitements
informatiques et Cognition (UMR 8094)
1 rue Maurice Arnoux
92120 Montrouge
France
sophie.prevast@ens.fr
[https://www.lattice.cnrs.fr/amp/membres/
direction/sophie-prevost/](https://www.lattice.cnrs.fr/amp/membres/direction/sophie-prevost/)