

The evolution of word order frequencies in medieval English and French

Anthony Kroch
Beatrice Santorini
Department of Linguistics
University of Pennsylvania

www.ling.upenn.edu/~kroch/handouts/hessenl2.pdf

Sources

English Parsed Corpora I

- Anthony Kroch and Ann Taylor. *Penn-Helsinki Parsed Corpus of Middle English, second edition*. University of Pennsylvania, 2000. (<http://www.ling.upenn.edu/hist-corpora>)

1.3 million words

- Anthony Kroch, Beatrice Santorini, and Ariel Diertani. *Penn-Helsinki Parsed Corpus of Early Modern English*. University of Pennsylvania, 2004.

1.8 million words

- Anthony Kroch, Beatrice Santorini, and Ariel Diertani. *Penn Parsed Corpus of Modern British English*. University of Pennsylvania, 2010.

1.0 million words

English Parsed Corpora II

- Ann Taylor, Anthony Warner, Susan Pintzuk, and Frank Beths. *York-Toronto-Helsinki Parsed Corpus of Old English Prose, first edition*. Oxford Text Archive, 2003.
(<http://www-users.york.ac.uk/~lang22/YCOE/YcoeHome.htm>)

1.5 million words

- Ann Taylor, Arja Nurmi, Anthony Warner, Susan Pintzuk, and Terttu Nevalainen. *Parsed Corpus of Early English Correspondence, first edition*. Oxford Text Archive, 2006.

2.2 million words

French Parsed Corpora

- France Martineau et al. *MCVF Corpus of Historical French*. University of Ottawa, 2010. (<http://www.arts.uottawa.ca/voies/>)
 \approx 1 million words
- Anthony Kroch and Beatrice Santorini. *Penn Supplement to the MCVF Corpus of Historical French*. University of Pennsylvania, 2012. (<http://www.ling.upenn.edu/hist-corpora/>)
 \approx 150,000 words

The loss of verb-second word order
and the decline of topicalization

Topicalized object sentences with V2 word order in English and French

- (1) These thinges shuld a mayde bisely kepe with gret drede....
(CMAELR4-M4,10.259)
- (2) Alle sal man comunlike recaieue
(CMBENRUL-M3,35.1131)
- (3) messe e matines ad li reis escultet
mass and matins has the king heard
(Roland 11.139)
- (4) l'estreu li tint sun uncle Guinemer
the stirrup him held his uncle Guinemer
(Roland 27.329)

Topicalized object sentences with V3 word order in English and French

(1) Thes same wordys God schall say to you at +te day of dome
(CMMIRK-M34, 71.1908)

(2) þat Octa scal ifinden
(LAYAMON-MI, 508.466)

(3) ceste franchise il tient de Deiu
this privilege he has from God
(Somme, 80.268)

(4) aultre remyde je n'y voy
other remedy I not there see
(XV Joies, III v.1209)

Adverb initial sentences with V2 word order in English and French

(1) and þanne schulde þy talkyngge be sauery and swete
(CMAELR3-M23, 49.734)

(2) and elles miȝt neuer man in þis liif abide it ne bere it.
(CMCLOUD-M3, 84.453)

(3) or est ele bien venue
now is she welcome
(Yvain 43.1440)

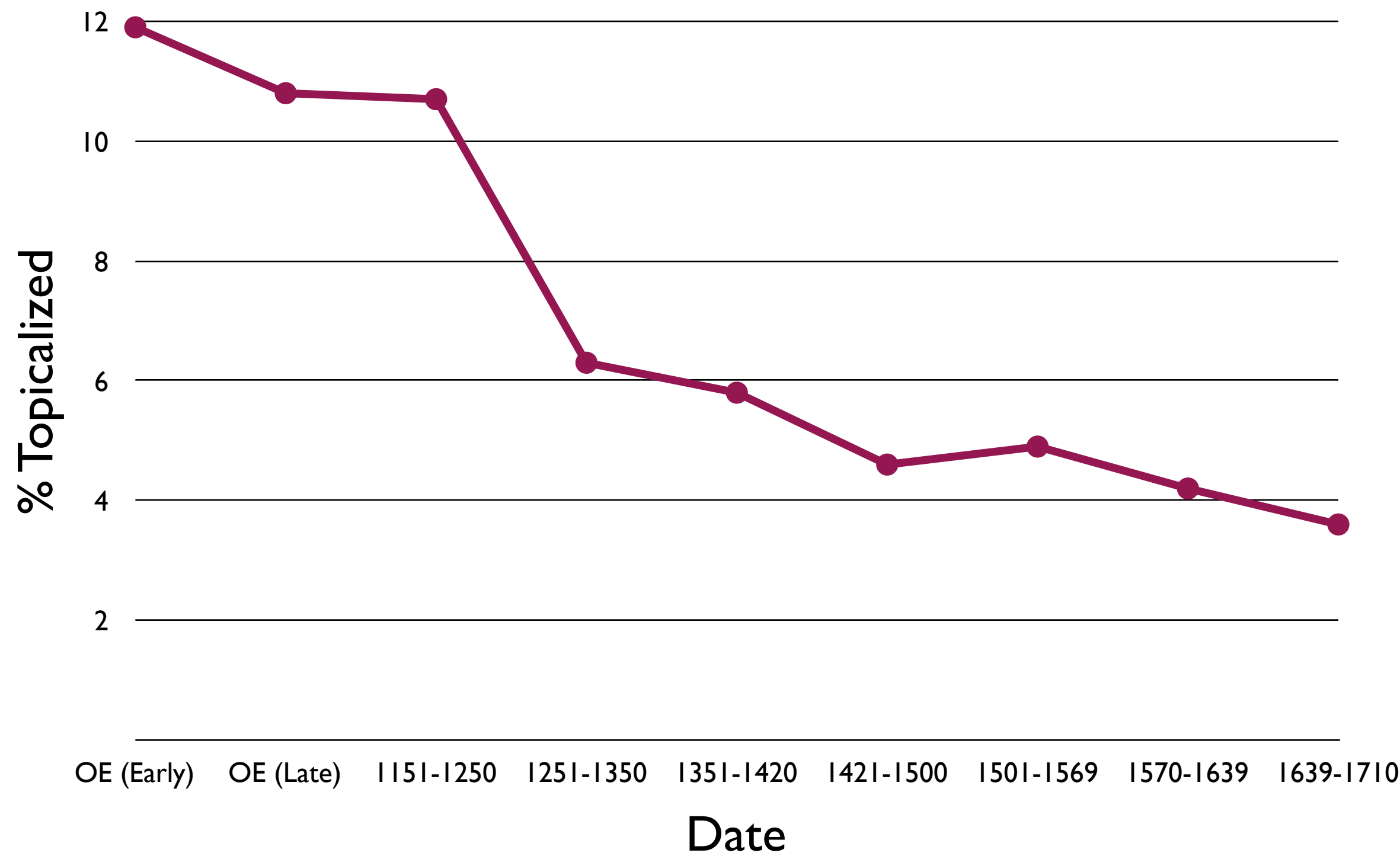
(4) ben l'entendit il arcevesques Turpin
well it understands the archbishop Turpin
(Roland 96.1237)

Adverb initial sentences with V3 word order in English and French

- (1) eftir þe dede sal þe not mis-like
(CMBENRUL-M3, 8.238))
- (2) And þan þe swan shal bene slayne wiþ sorwe
(CMBRUT3-M3, 74.2230)
- (3) unkes nuls hom ne vit tel ajustee.
never no man not saw such joust
(Roland 238.3299)
- (4) premierement il devient taverniers
first he becomes innkeeper
(Somme 47.306)

English

Decline of direct object topicalization in English

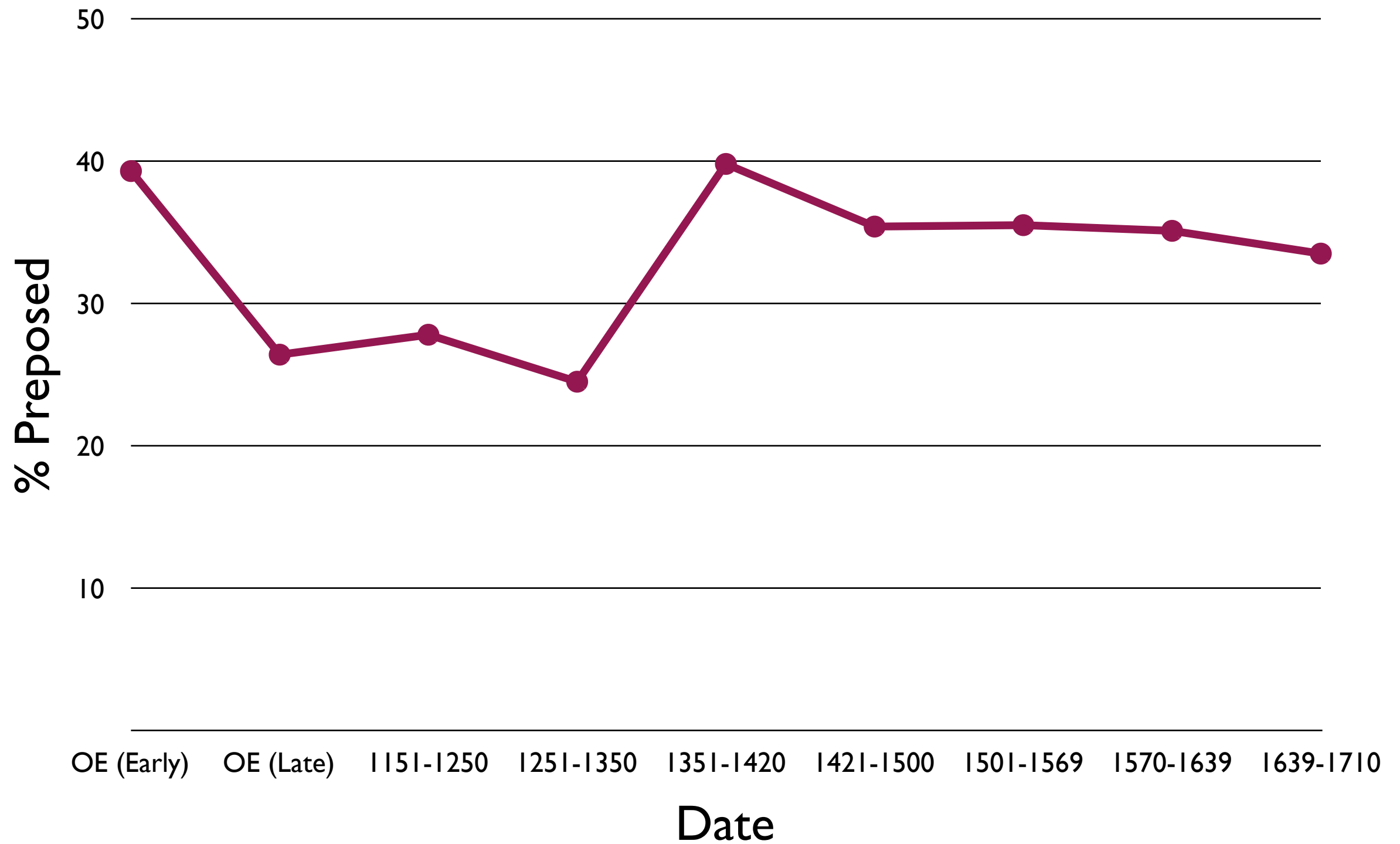


Frequency of direct object topicalization in modern spoken Dutch (Bouma 2008)

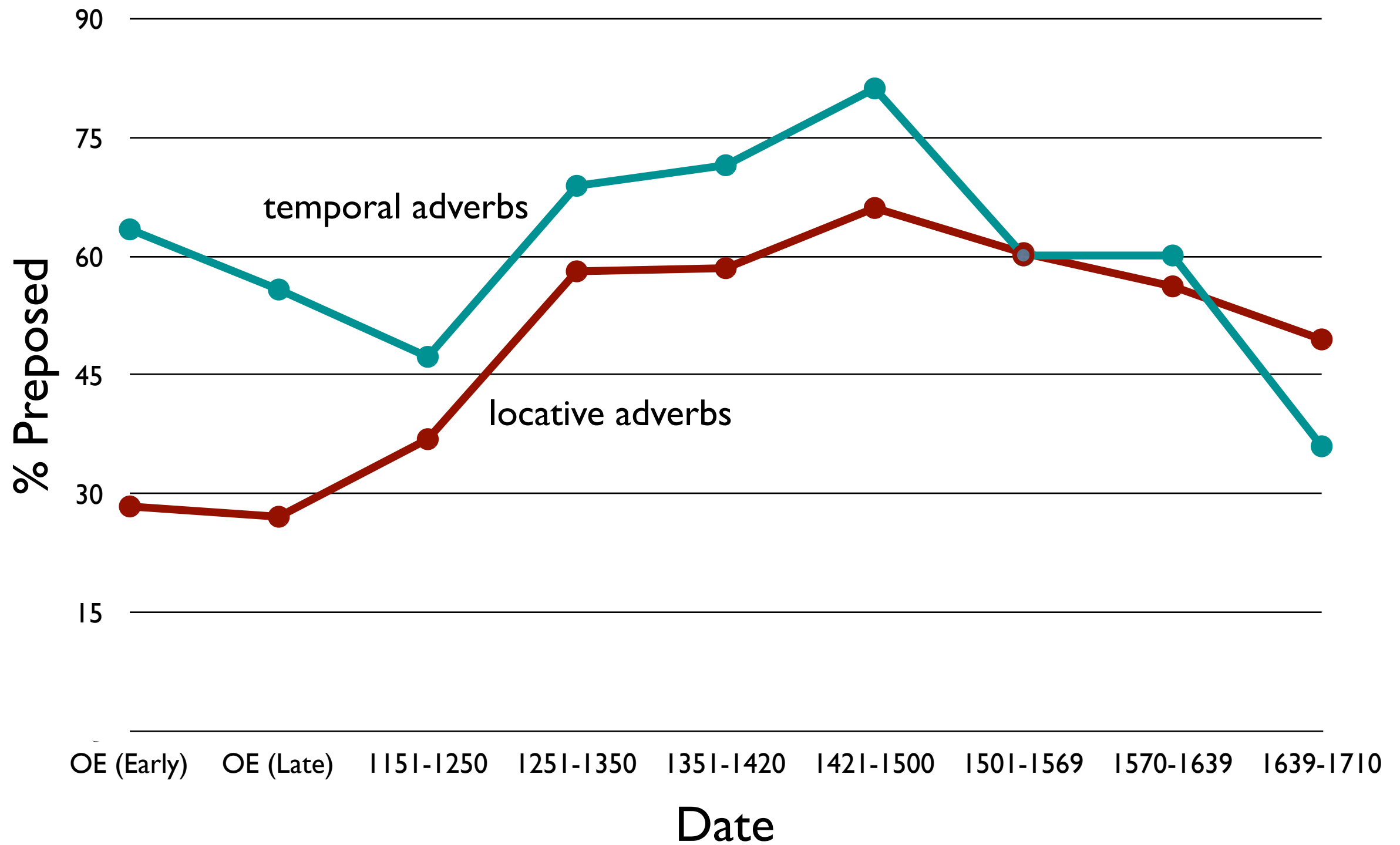
Table 4.2: Summary of Vorfeld occupation of arguments.

Argument	Vorfeld		Prop est (%)
	yes	no	pt
subject	43 523	18 597	70.1
direct object	3 418	20 432	14.3
indirect object	38	815	4.5

Evolution of PP preposing in English



Evolution of adverb fronting in English



The history of topicalization in English (Speyer 2008)

- Why does topicalization decline in Middle English but not disappear? If the change is parametric, it should go to completion. Otherwise, topicalization, a clear case of stylistic variation, might be expected to be stable in frequency over time.
- This question finds an answer in the specific interaction between parametric settings and stylistic variation in the history of English.

An illustrative case in the New Testament

Matthew 13.28

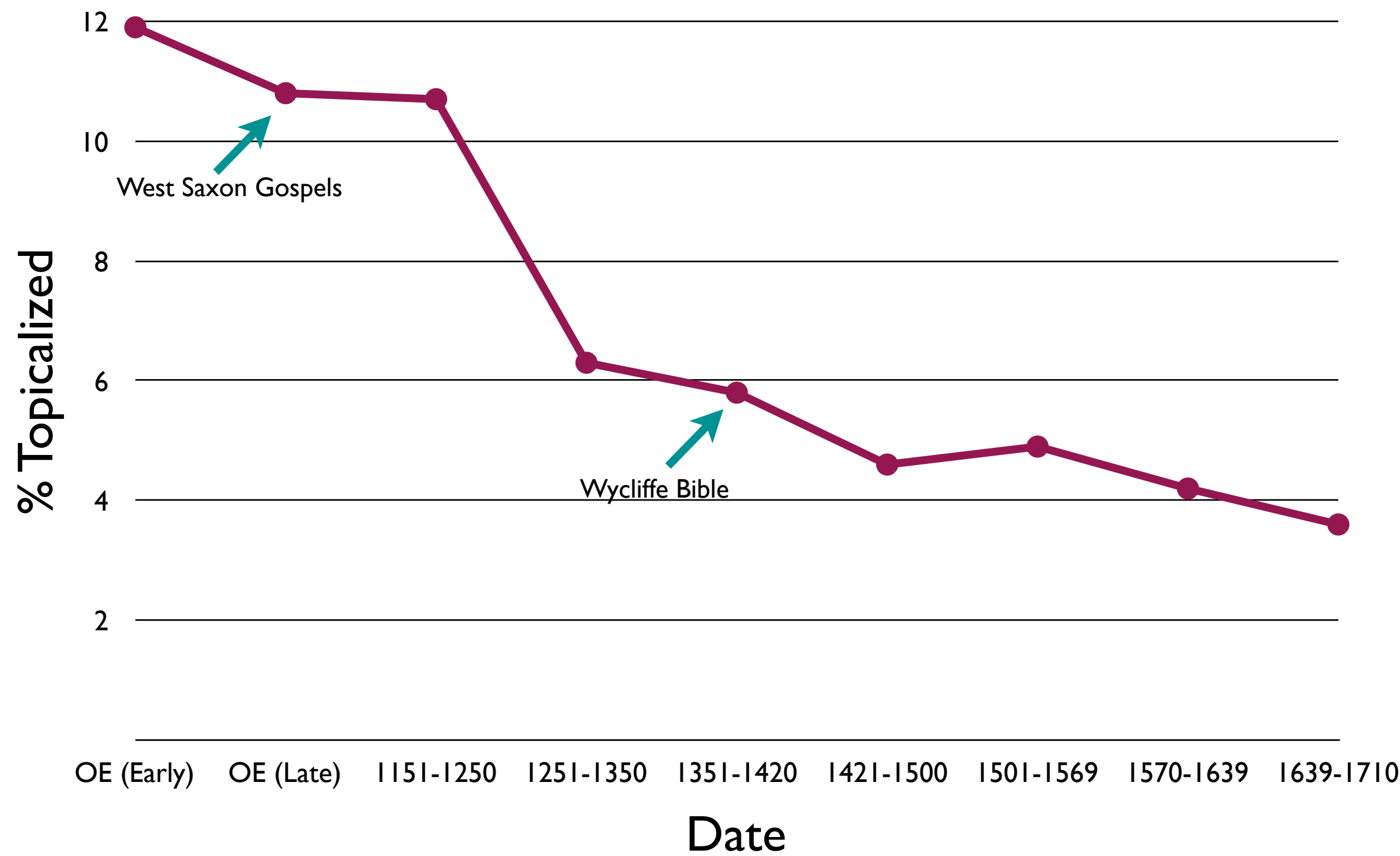
(1) Ðā cwæþ hē, **Þæt dyde** unhold mann.

West Saxon Gospels ca. 1000

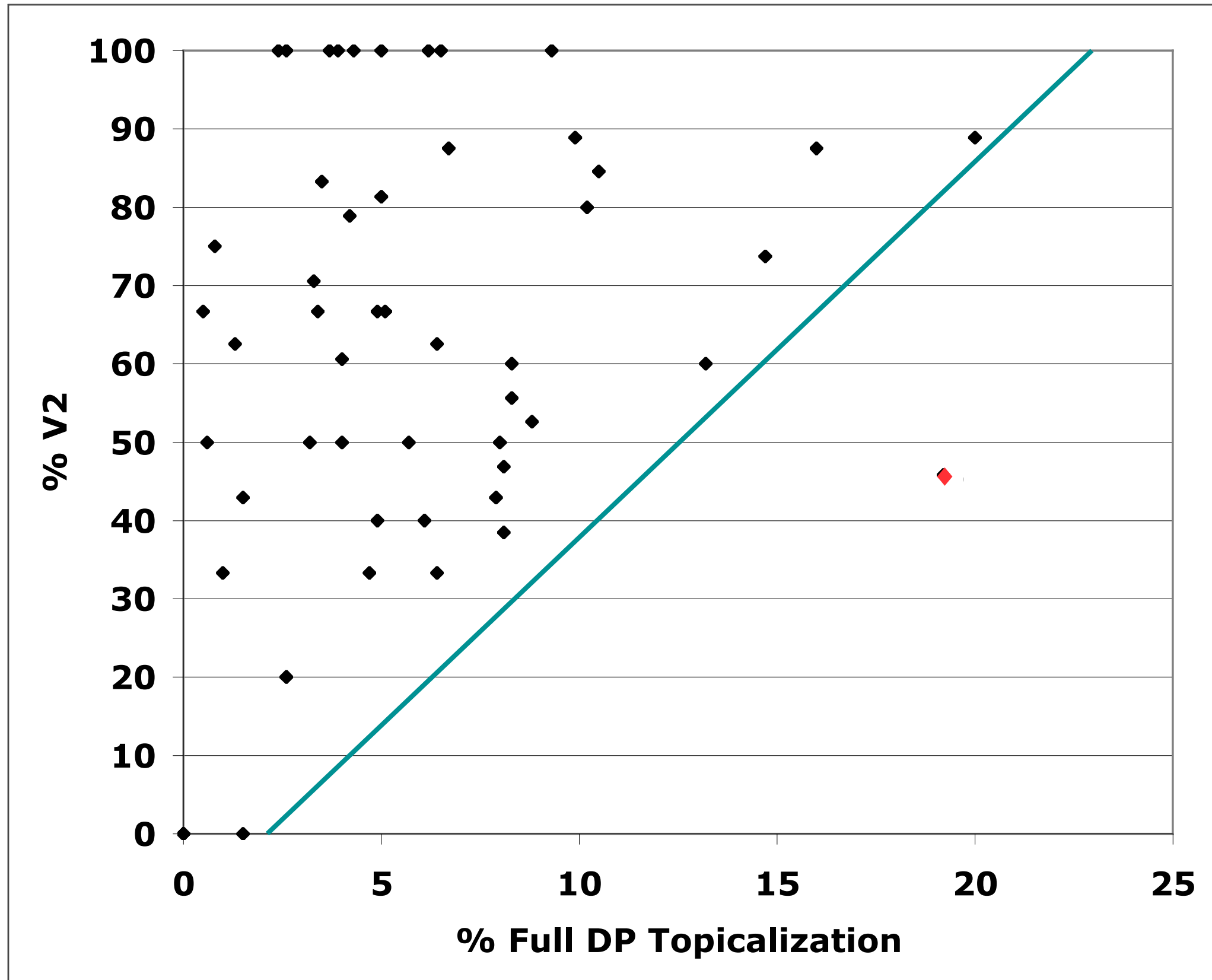
(2) And he seide to hem, An enemy **hath** do~ **this thing**.

Wycliffe Bible ca. 1380

Decline of direct object topicalization in English



Correlation between frequencies of object topicalization and of V2 in Middle English texts (Wallenberg 2007)



Distribution of subject types in a corpus of topicalized and non-topicalized sentences in natural speech

personal pronoun	demonstrative pronoun	full noun phrase
140	20	142
46.4	6.6	47.0

Subject type in sentences with *in situ* objects

personal pronoun	demonstrative pronoun	full noun phrase
181	2	17
90.5 %	1 %	8.5 %

Subject type in sentences with topicalized objects

Clash avoidance

- The type of topicalization that declines:

(1) The *nèwspaper* *Jóhn* read; the *nòvel* *Máry* did.

(Compare: The *nèwspaper* read *Jóhn*.)

- The type of topicalization that doesn't:

(2) The *nèwspaper* I *réad*; the *nòvel* I *díd*n't.

Translating German topicalized arguments into English in three modern German novels (Speyer 2008)

Topicalized to topicalized:

G: **Mahlkes Haupt** bedeckte dieser Hut **besonders peinlich**.

E: **On Mahlke's head** this hat made a **particularly painful impression**.

Topicalized to non-topicalized:

G: **Zu den sechs** kamen noch **drei weitere**.

E: **Three others** joined **these six** in the afternoon.

Accent placement and topicalization frequencies in translating German topicalized arguments into English

	focus accent on the German subject	accent elsewhere
topicalization in the English translation	0	31
no topicalization in the English	25	100

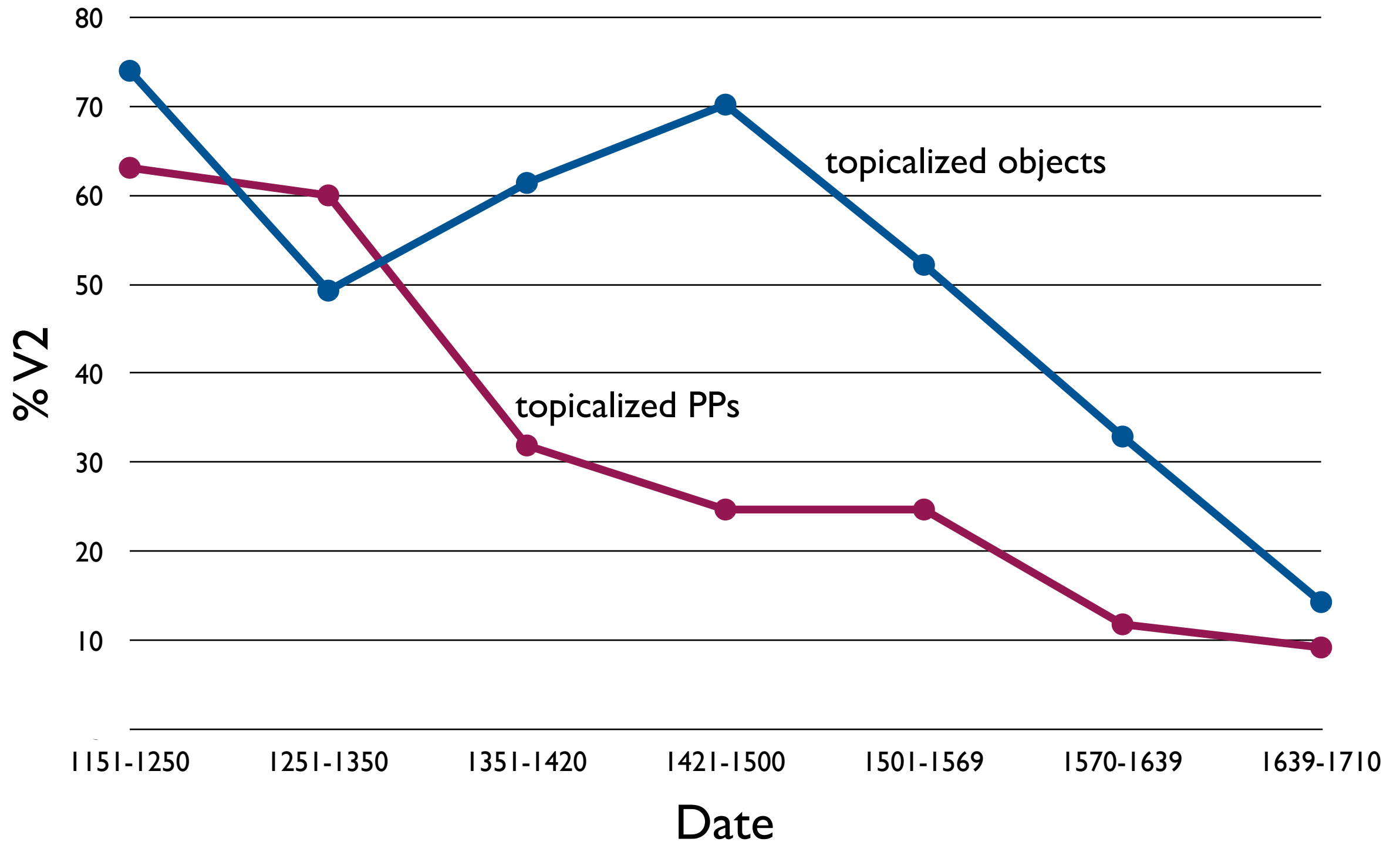
Distribution of contrastive topicalization by focus (second accent) placement in Middle English (Speyer 2008)

focus position distribution of cases	focus on subject	focus on tensed verb	focus elsewhere
N (total= 207)	113	29	65
% inversion	89	14	71
% of cases	55	14	31

Frequency of matrix clause topicalization and V2 in Middle and Early Modern English

time period sentence type	me1	me2	me3	me4	eme1	eme2	eme3
# sent. with DO	2855	1300	4615	2271	3229	3584	2544
# topicalized	219	69	145	66	67	82	28
% topicalized	7.7	5.3	3.1	2.9	2.1	2.3	1.1
# V2	162	34	89	46	35	27	4
% V2	74	49.3	61.4	70.2	52.2	32.9	14.3

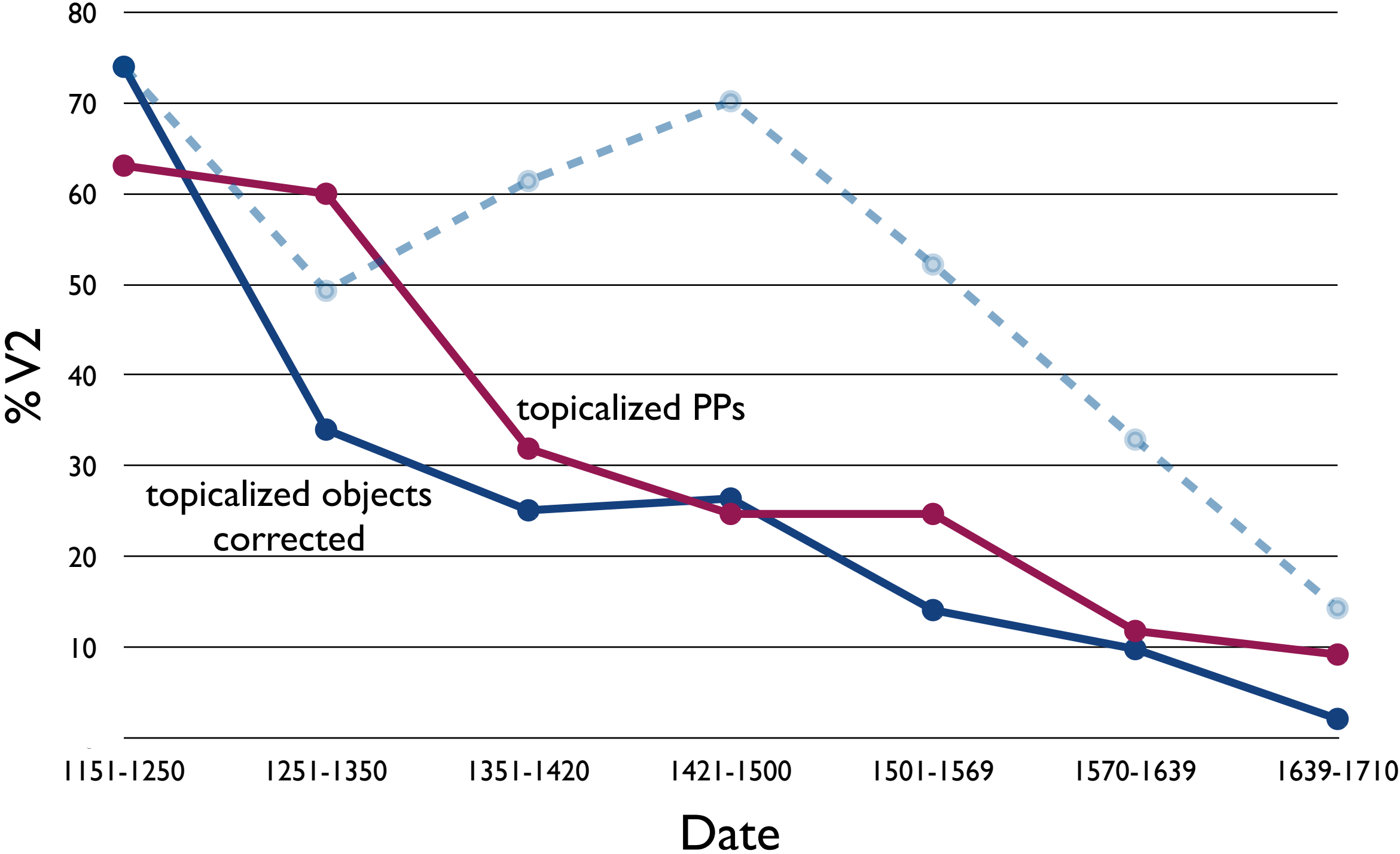
Rate of V2 loss in English with topicalized objects and PPs



Corrected frequency of matrix clause topicalization and V2 in Middle and Early Modern English

time period sentence type	me1	me2	me3	me4	eme1	eme2	eme3
# sent. with DO	2855	1300	4615	2271	3229	3584	2544
# topicalized	219	69	145	66	67	82	28
would have been topicalized	219	100	354	174	248	275	195
actual rate of V2	74	49.3	61.4	70.2	52.2	32.9	14.3
corrected rate V2	74.0	34.0	25.1	26.4	14.1	9.8	2.1

Rate of V2 loss in English corrected for clash avoidance

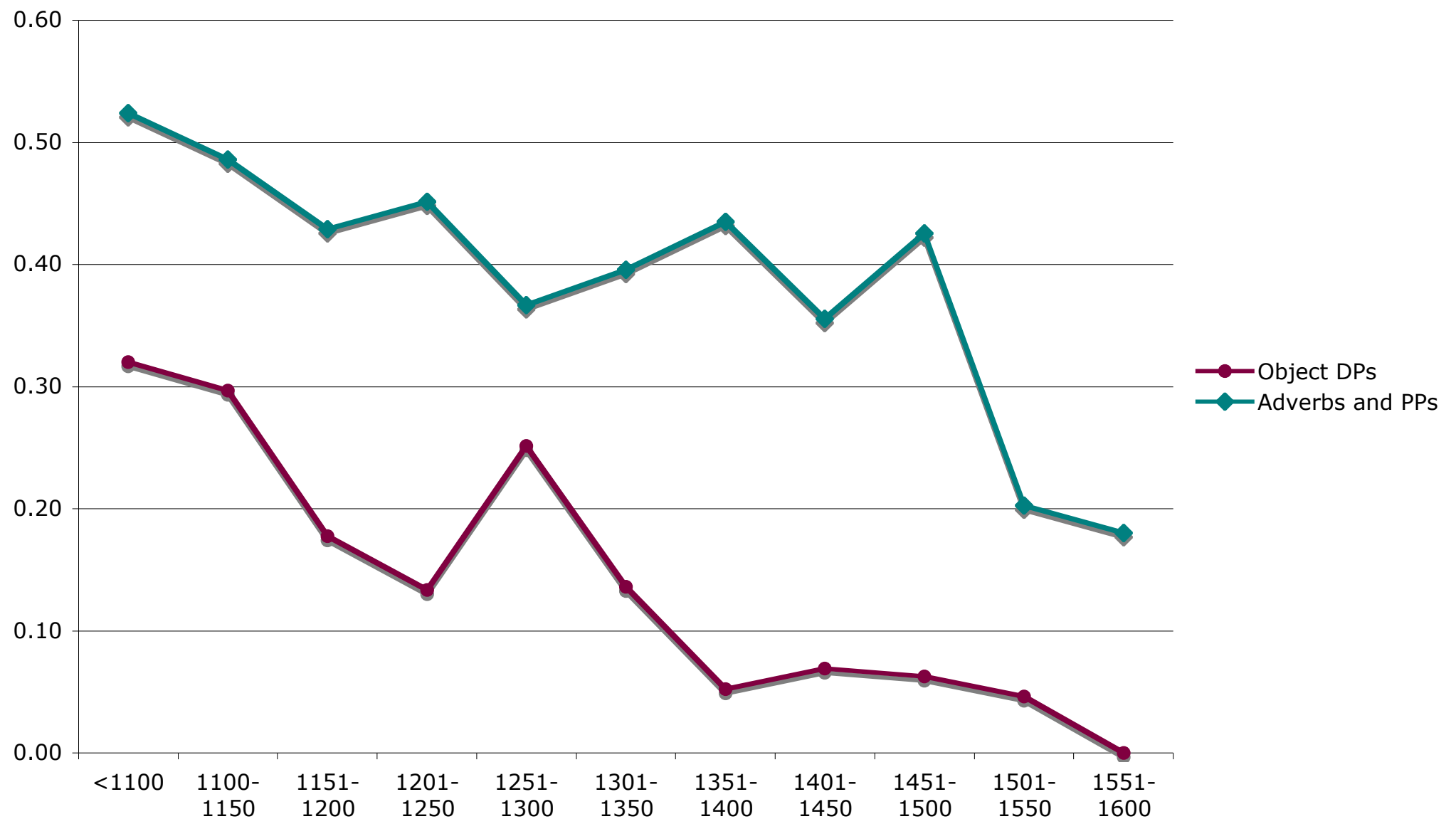


French

V2 in Old and Middle French

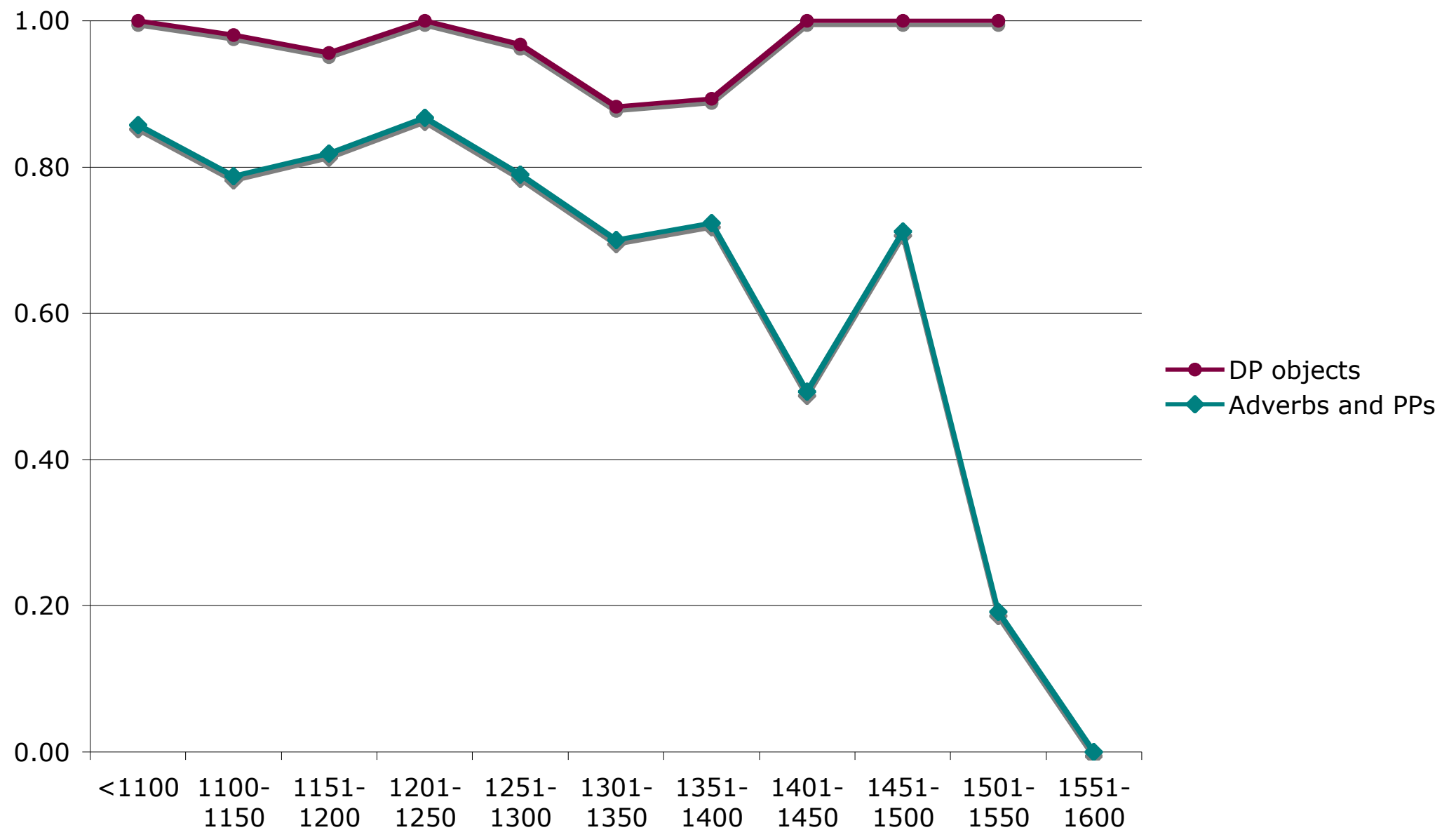
- (1) l'estreu li tint sun uncle Guinemer
the stirrup him held his uncle Guinemer
Roland 27.329
- (2) Espaigne vus durat il en fiet
Spain you will-give he in fief
Roland, 36.446
- (3) or est ele bien venue
now is she welcome
Yvain 43.1440

Decline of XP fronting in French in sentences with overt subjects



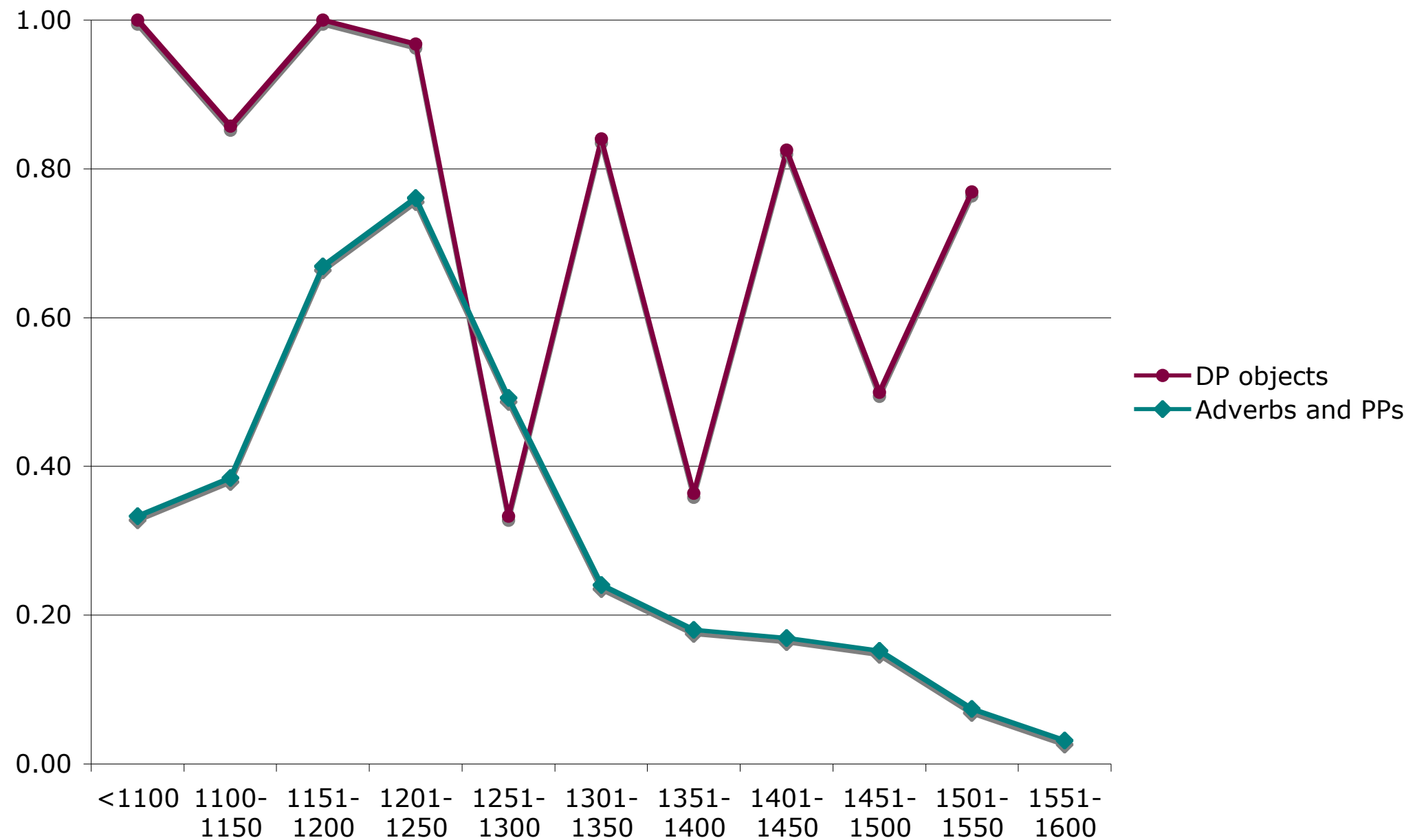
— Old French —+— Middle French +— Modern French

V2 frequency: sentences with full NP subjects



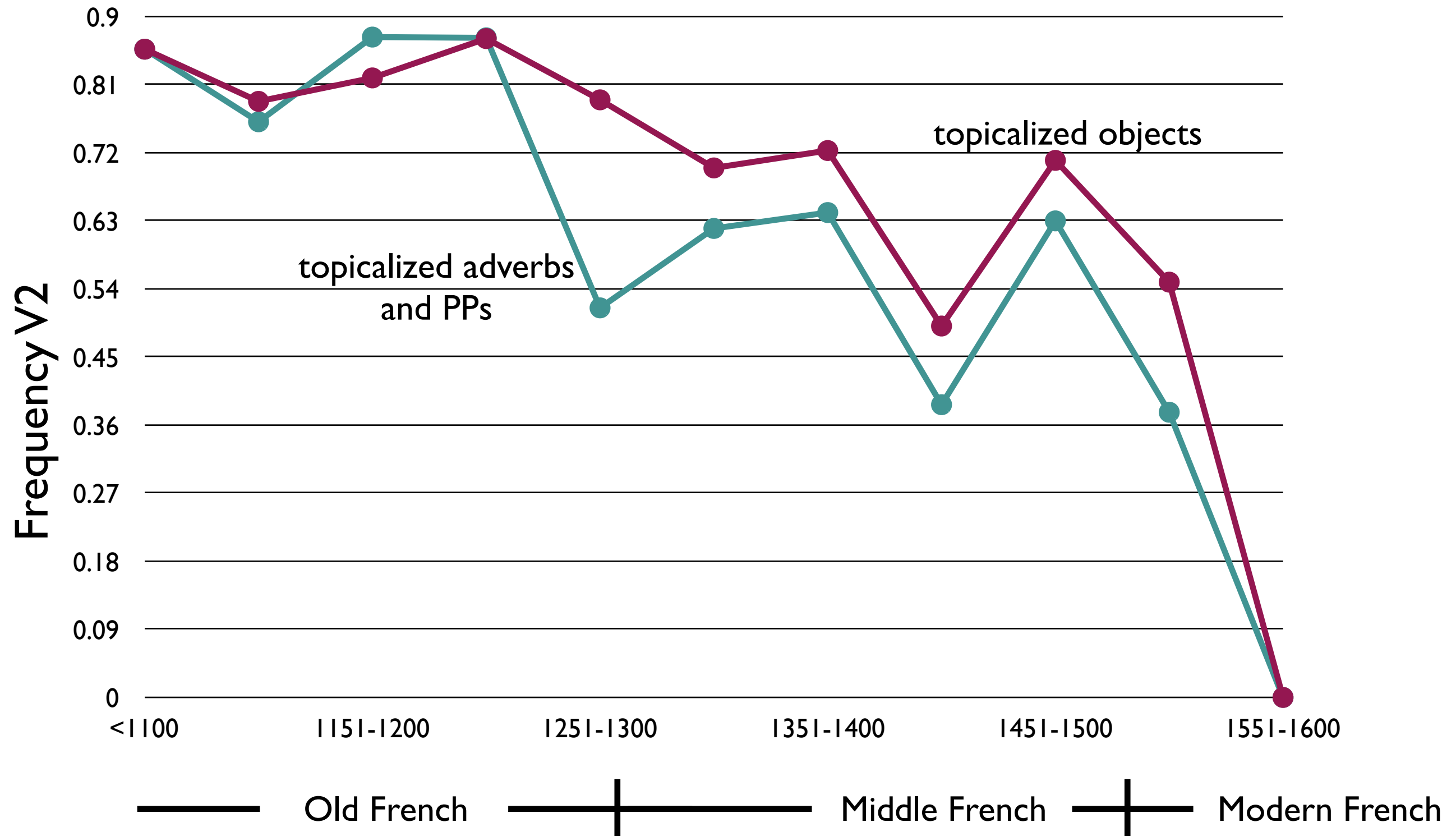
— Old French —+— Middle French +— Modern French

V2 frequency: sentences with overt pronoun subjects



— Old French —+— Middle French +— Modern French

Corrected evolution of V2 word order in French in sentences with full DP subjects



Germanic and Romance inversion in Old and Middle French

“Germanic” inversion in Old and Middle French

(1) messe e matines ad li reis escultet
mass and matins has the king heard
Roland 11.139

(2) chars avoient ils assés
meat had they enough
Froissart, 135.569

(3) une chose ont-ilz asez honneste
one thing have-they enough honest
Commynes, 120.1634

“Romance” inversion in Old French

(1) ... puis **si** chevalchet od sa grant ost **li** **ber**
then so rides with his great army the baron
Roland, 179.2438

(2) ... **ço** ad tut fait **Rollant**
that has all done Roland
Roland, 24.301

(3) **ceste parole** ot escoutee **li** **seneschax**
this speech has heard the seneschal
Yvain 134.4663

Ambiguous cases

- (1) **Après** parlat **ses filz** envers Marsilies
then spoke his son to Marsilies
Roland 37.466
- (2) **Bien** fiert **nostre guarent**
well fights our guardian
Roland 124.1665
- (3) **Mult fierement** chevalchet **li emperere**
very proudly rides the emperor
Roland 23.3296

Temporal evolution of V2 with full DP subjects for all types of preposed XP

	sentences with an auxiliary verb	sentences with a single verb
Old French	0.86 [218]	0.83 [2163]
Middle French	0.69 [402]	0.70 [3633]
Modern French	0.27 [33]	0.22 [160]

Temporal evolution of Germanic and Romance inversion
in V2 sentences with auxiliary verbs
(sentences with topicalized XPs and full DP subjects)

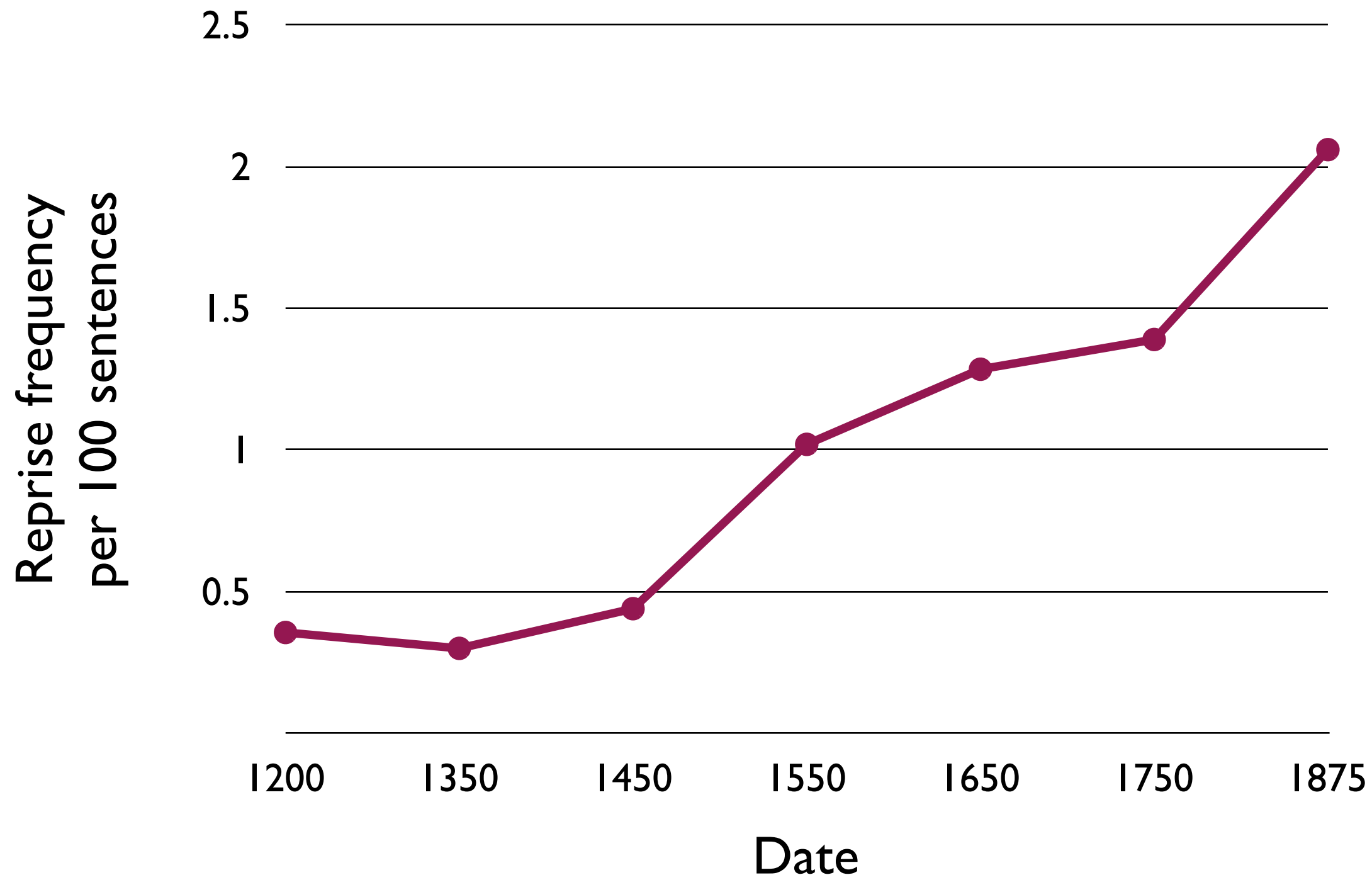
	frequency of Germanic inversion	frequency of Romance inversion	Romance + Germanic inversion
Old French	0.50	0.36	0.86
Middle French	0.32	0.37	0.69
Modern	0.03	0.24	0.27

An independence result

	Romance + Germanic inversion	sentences with a single verb
Old French	0.86	0.83
Middle French	0.69	0.70
Modern French	0.27	0.22

Why does French completely lose
object topicalization?

Rise of clitic left-dislocation and loss of topicalization (Priestley 1955)



Modern French clitic left dislocation

(1) *Le Figaro*_i, Jean ^{*}(*le*_i) lit tous les jours.
The Figaro John it reads every day

(2) *Ma femme*_i, *elle*_i travaille à la Bibliothèque Nationale.
My wife she works at the library national

Temporal evolution of subject and object left dislocation frequencies per thousand sentences

	frequency of subject left dislocation	frequency of object left dislocation	number of matrix clauses
Old French	2.6	2.2	12022
Middle French	3.8	1.8	24634
Early Modern	28	4.3	3514

Cleft sentences in Modern French

- (1) C'est *Le Figaro*_i que Jean lit _{t_i} tous les jours.
It's *The Figaro* that John reads every day
- (2) C'est *ma femme*_i qui _{t_i} travaille à la BN.
It's my wife that works at the BN
- (3) Il y a *un an*_i qu'elle travaille à la BN _{t_i}.
It's one year that-she works at the BN

Temporal evolution of cleft sentence frequencies per thousand sentences

	frequency of temporal clefts	frequency of subject and object clefts	number of matrix clauses
Old French	1.2	0.25	12022
Middle French	0.41	0.61	24634
Early Modern	0.56	5.4	3514

Finis

The decline of OV word order

AuxOV word order in early Middle English

- (1) for þa twelf kunreden **sculden** þer mide **heore þurst kelen**.
for the twelve kinfolk should therewith their thirst cool.
(CMLAMBI, 141.271)
- (2) Herode king **ma33** swiþe wel **þe laþe gast bitacnenn**;
Herod king may very well The Evil Spirit betoken;
(CMORM, I, 280.2285)
- (3) for no man ne **mai his sinnes bete**; er he habbe nume shrift þerof.
for no man may his sins atone; before he has taken shrift thereof.
(CMTRINIT, 59.796)
- (4) ... þat Frolle **wolde** mid faehte **France bitellen**.
... that Frolle would with violence France defend.
(LAYAMON, 618.2151)

OV_{fin} word order in early Middle English

for æuric rice man his castles made

for every rich man his castles made

(CMPETERB, 55.425)

and eft sone ech at shrifte þat ilke bihoteð.

and repeatedly each at shrift that same promises.

(CMTRINIT, 61.837)

and fele here lif fulliche lateð.

and many their life fully surrender.

(CMTRINIT, 181.2470)

His cun i þisse londe ure aelderne a-qualden.

His kin in this land our elders killed.

(LAYAMON, 522.706)

Cador his sweord an-hof

Cador his sword up-lifted

(LAYAMON, 562.1314)

AuxOV word order in early Old French

- (1) Li reis Marsilies **ad la culur muee**; (1100-ROLAND, 34.419)
the king Marsilies has the color changed
- (2) Quant la dame **ad la messe oïe**, Ariere vait,
when the lady has the mass heard back goes
(116X-M-D-F, 19.352)
- (3) Li sire en **ad Deu mercié**;
the lord for-it has God thanked
(116X-M-D-F, 45.886)
- (4) Chaucuns **ha la seue espousee** (1190-BORON-V, 103.1639)
each has the his married

AuxOV word order in later Old French

- (1) Quant li rois **ot ceste parole dite** si se tot,
when the king had this word said, SI REFL was-silent
(1225-QUESTE, 31.726)
- (2) La parole que cis dist **a moult les autres meüs**,
the word that this-one said has much the others moved
(127X-CASSIDORUS, 6.85)
- (3) Quant les jouvenceaulx **eurent leur ante emprisonnee....**
when the young-men had their aunt imprisoned
(133X-PERCEFOREST, 99.656)
- (3) Li honme de la ville **avoient le pont desfait**.
the men of the city had the bridge destroyed
(1373-FROISSART, 385.7693)

OV_{fin} word order in early Old French

- (1) Sainz Alexis **la sue** li **alascet**:
Saint Alexis the his him gave

(10XX-ALEXIS, 75.691)

- (2) Et Engelpers li Guasquinz de Burdele **Sun cheval brochet**,
and Engelpers the Gascon of Burdele his horse spurred
(1100-ROLAND, 100.1292)

- (3) Pois li prophetes Samuel **une pierre leva**,
Then the prophet Samuel a stone lifted
(1150-QUATRELIVRE, 15.473)

OV_{fin} word order in later Old French

- (1) se **ele** **cheval** li **done** bon et bel
if she horse him gives good and beautiful
(1225-QUESTE, 122.3246)
- (2) ...entendra que **qui** **eglise** **edifie** **lo** **regna** **daus** **ceus** s'**aparellia**.
...means that who church builds the reign of heaven acquires
(122X-PS-TURPIN, 332.2055)
- (3) ...dist ... que **qui** **peche** **encontre** **le** **Saint** **Esperit** ja **merci** n'en **aura**,
...says ... that who sins against the HS so mercy not EN will-have
(1279-SOMME I, 25.675)

The problem of poetry

(1) Li reis Marsilies ad **la culur mueue**;
De sun algeir ad **la hanste crollee**.

(1100-ROLAND, 34.419-20))

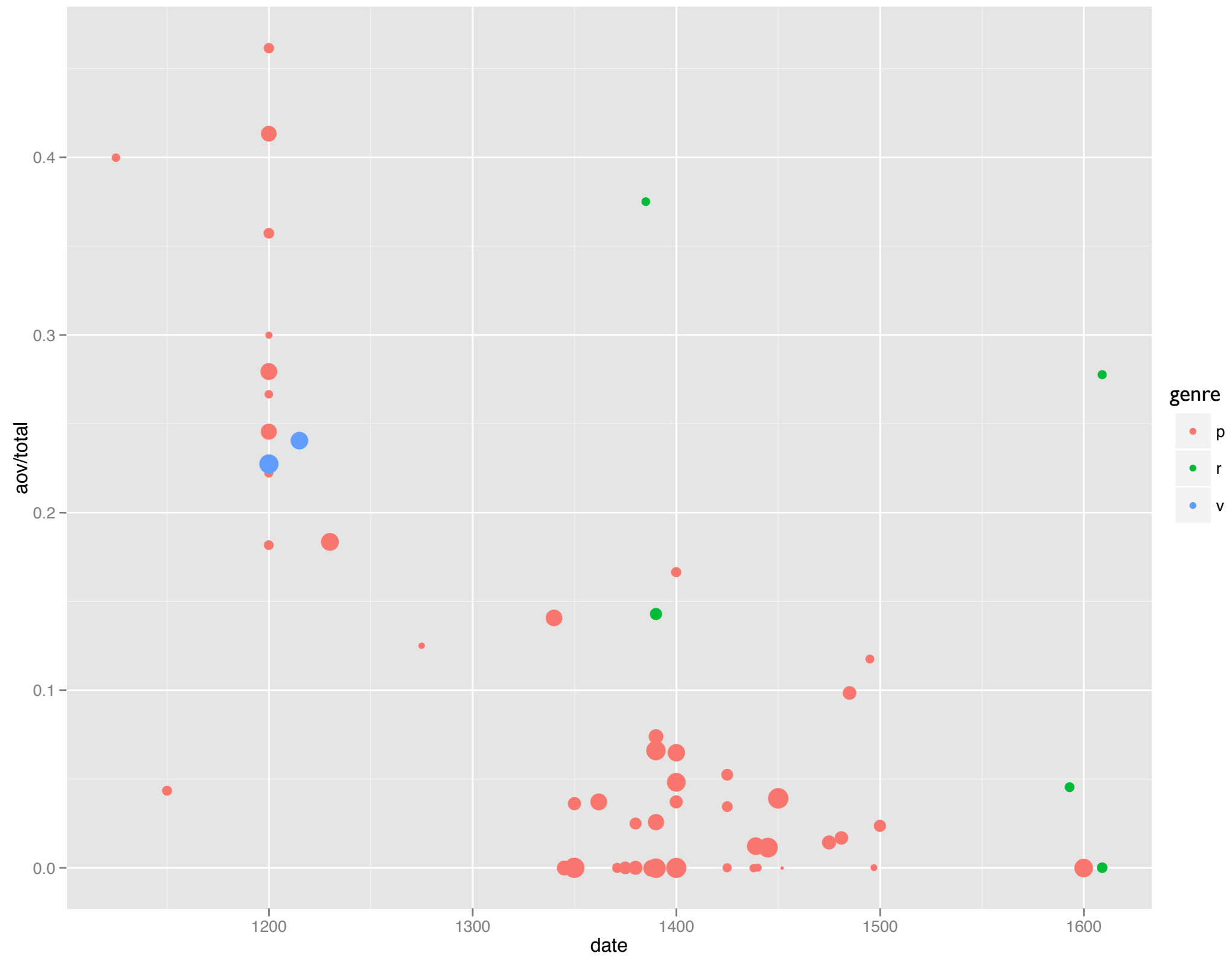
(2) It failleth nat whils it is in youre **hoold**
And whan this knyght hath thus **his tale toold**

(1390-CHAUCER-SQT, 58-59)

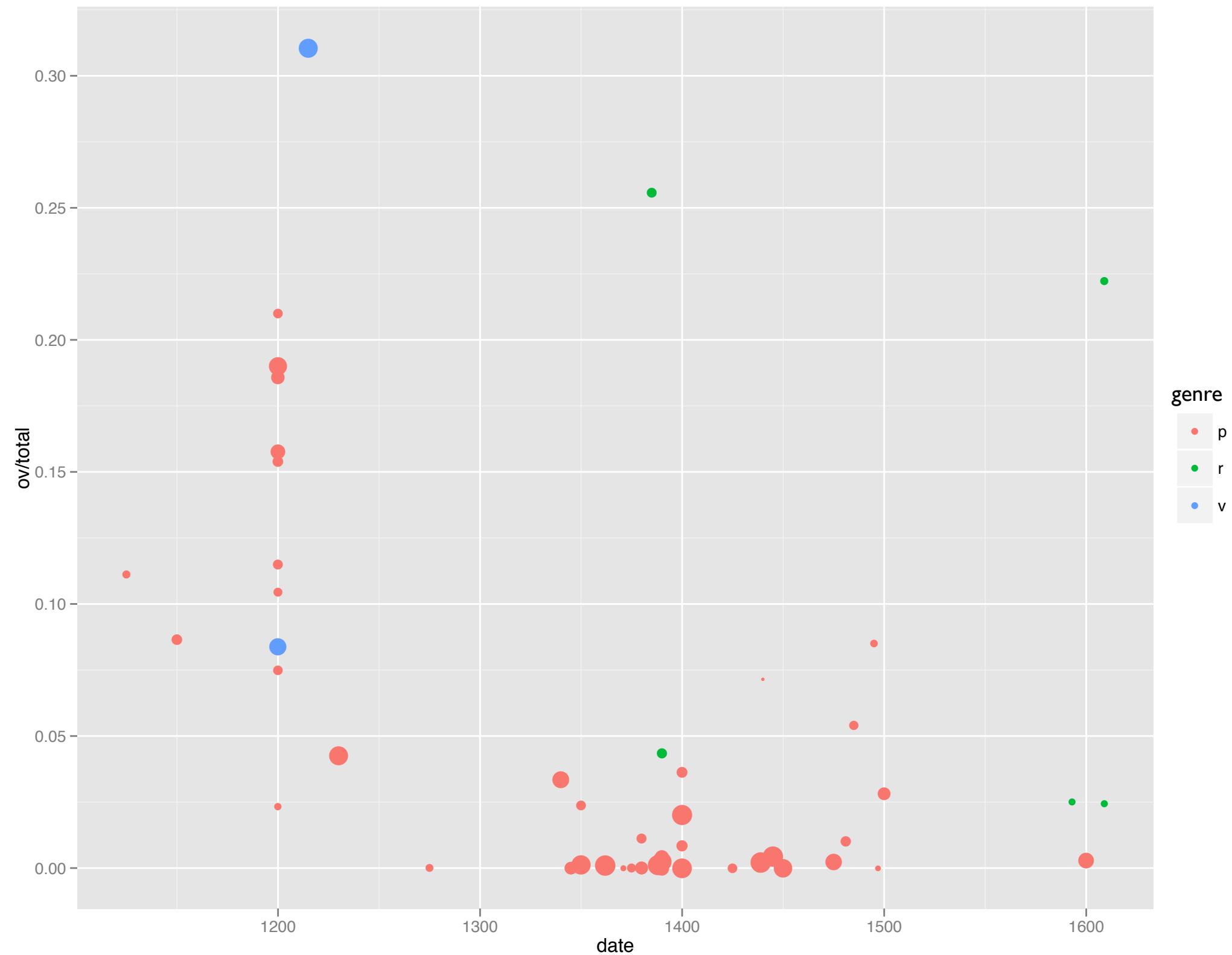
(3) They told her to return at once to **Steve**,
But she **her child** would never ever **leave**.

(Modern English doggerel)

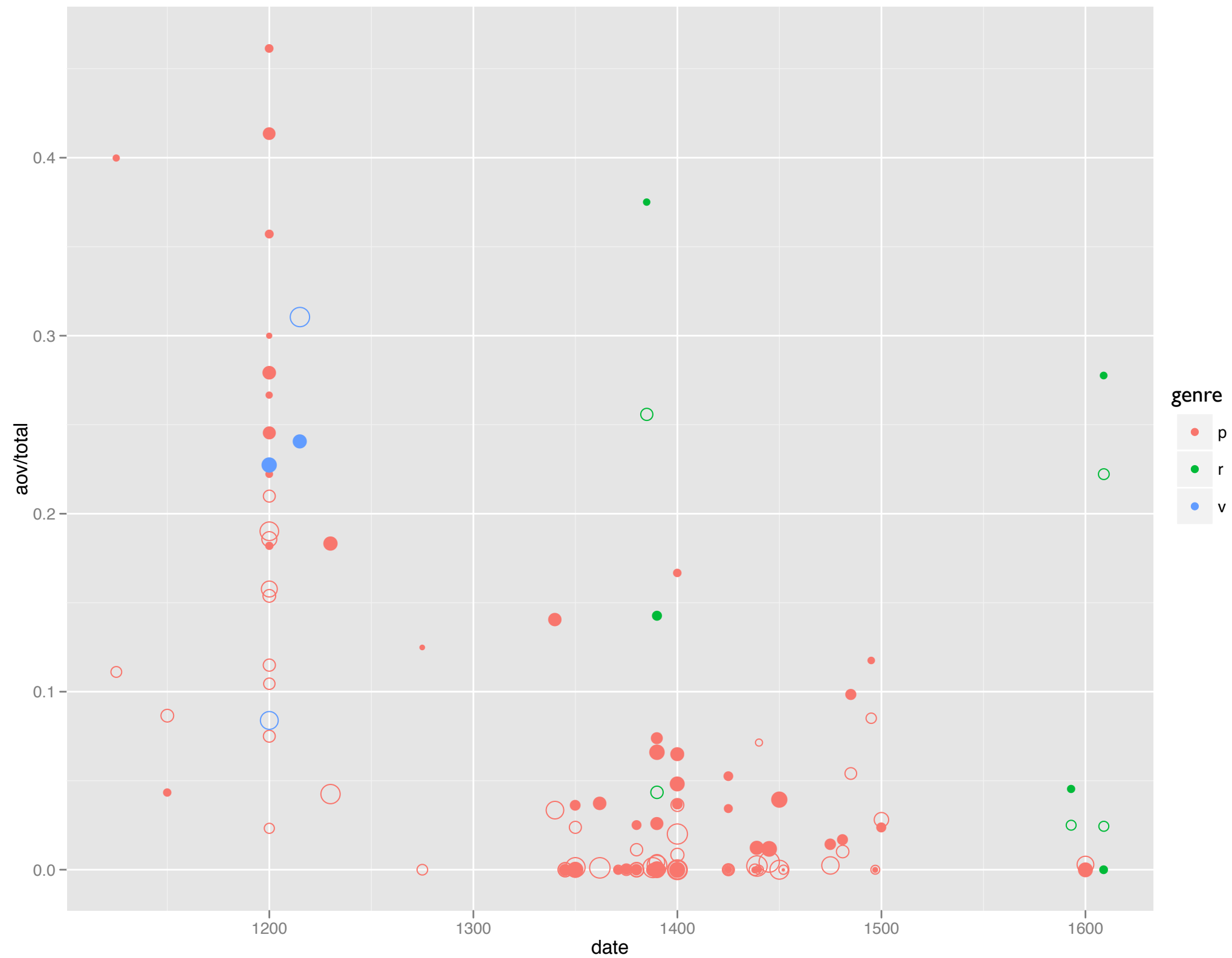
Frequency of AuxOV word order in English



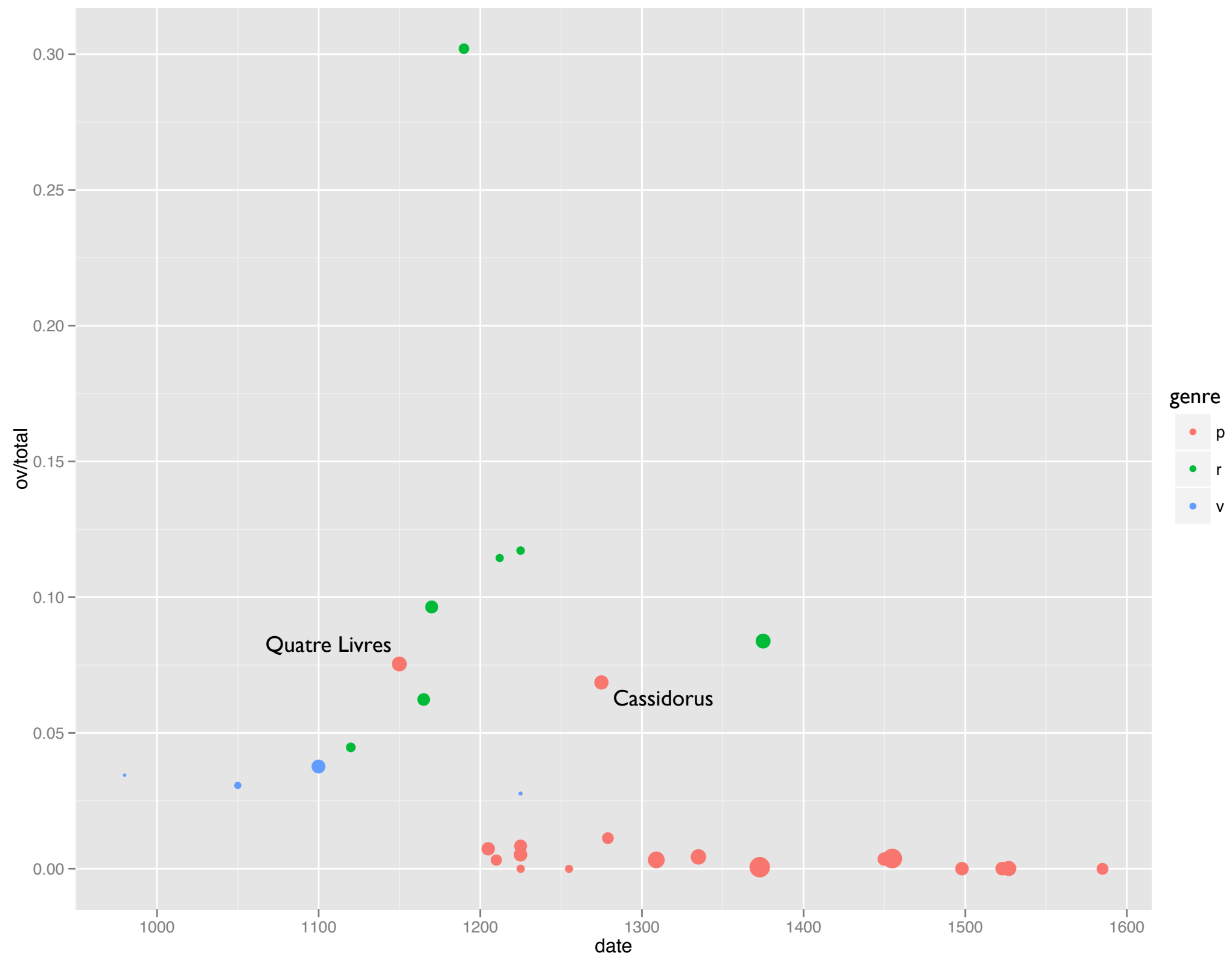
Frequency of OV_{fin} word order in English



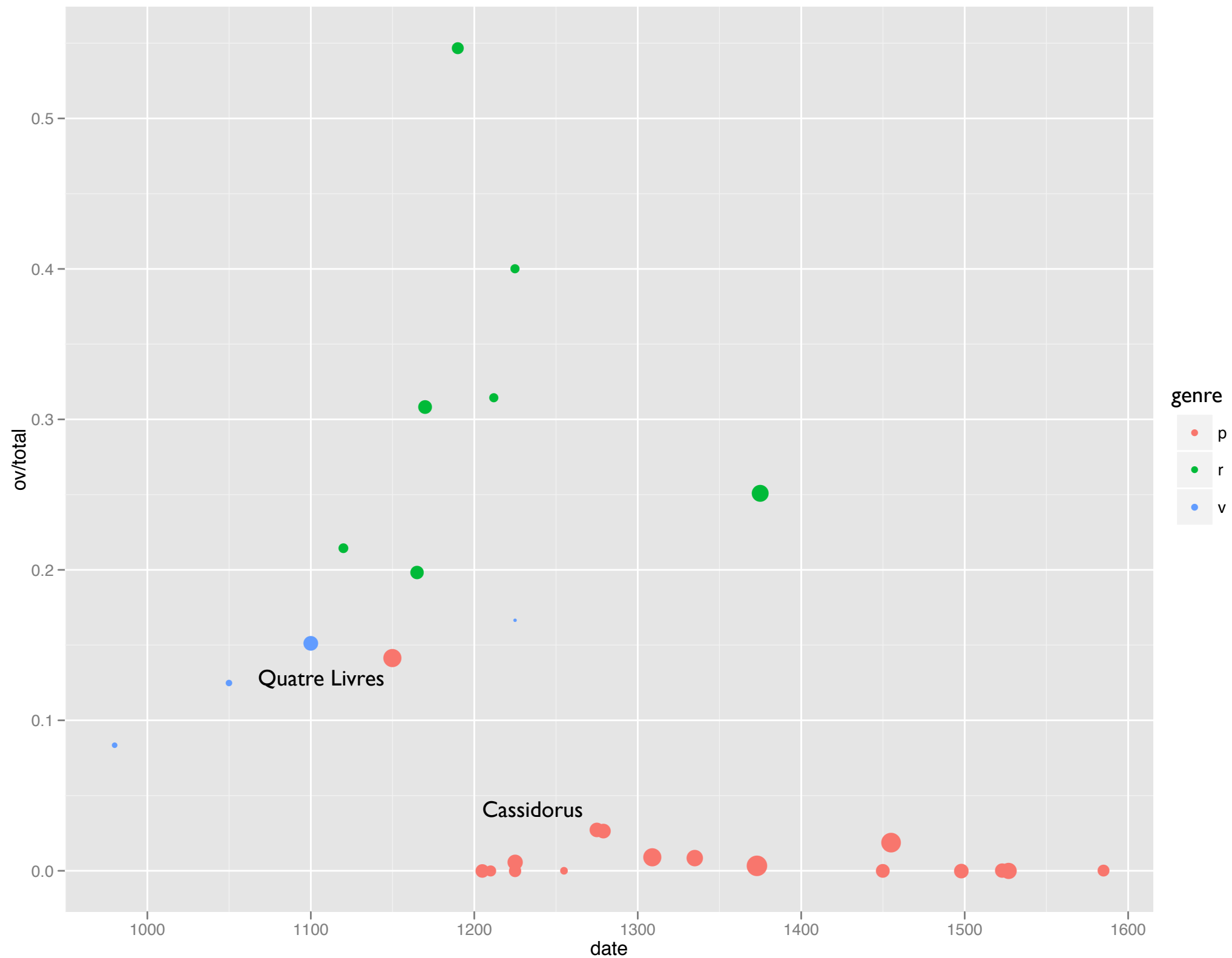
Frequency of AuxOV compared to OV_{fin} in English



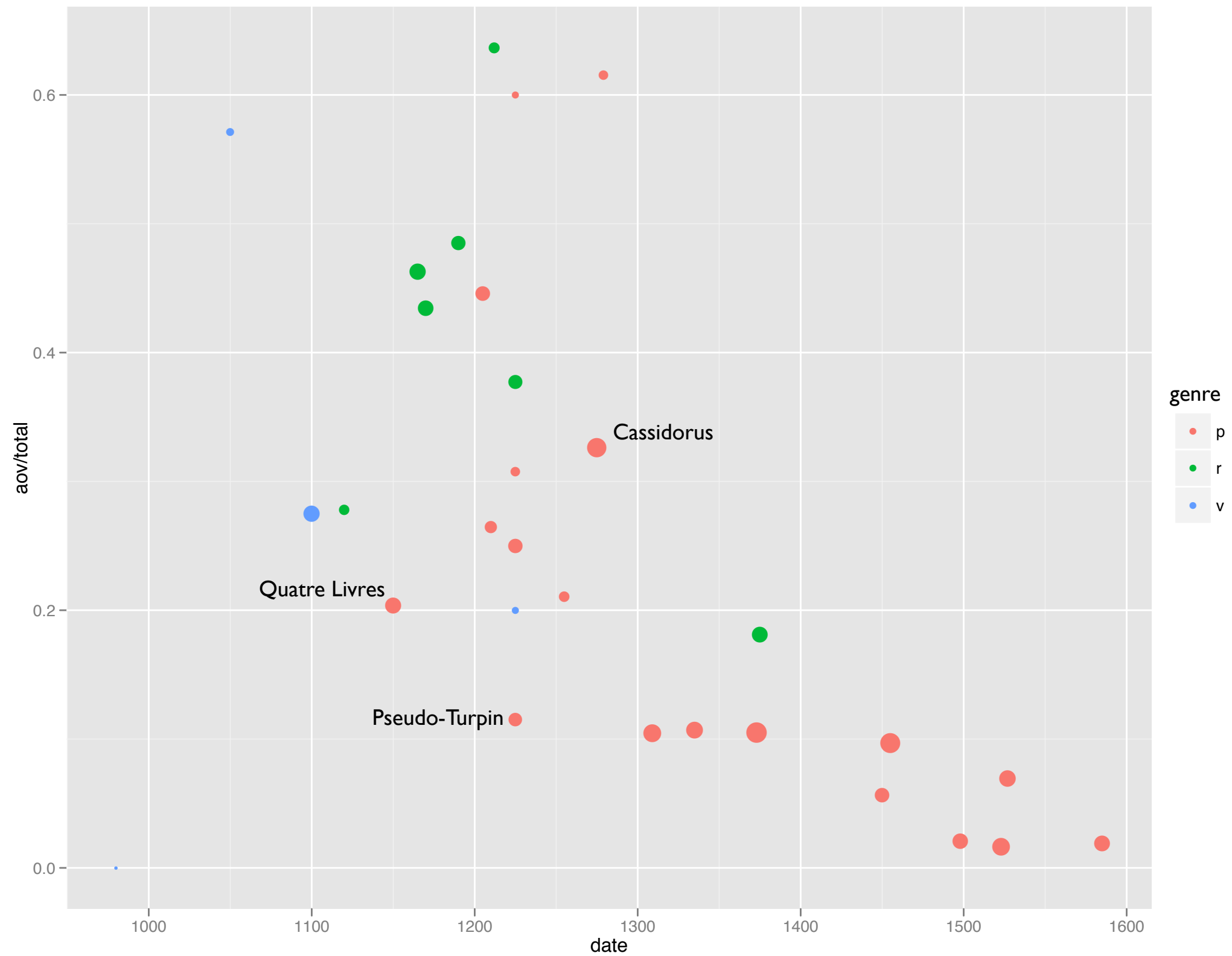
Frequency of OV_{fin} word order in French



Frequency of OV_{fin} word order in French, subject pronouns removed



Frequency of AuxOV word order in French



Frequency of AuxOV compared to OV_{fin} in French

