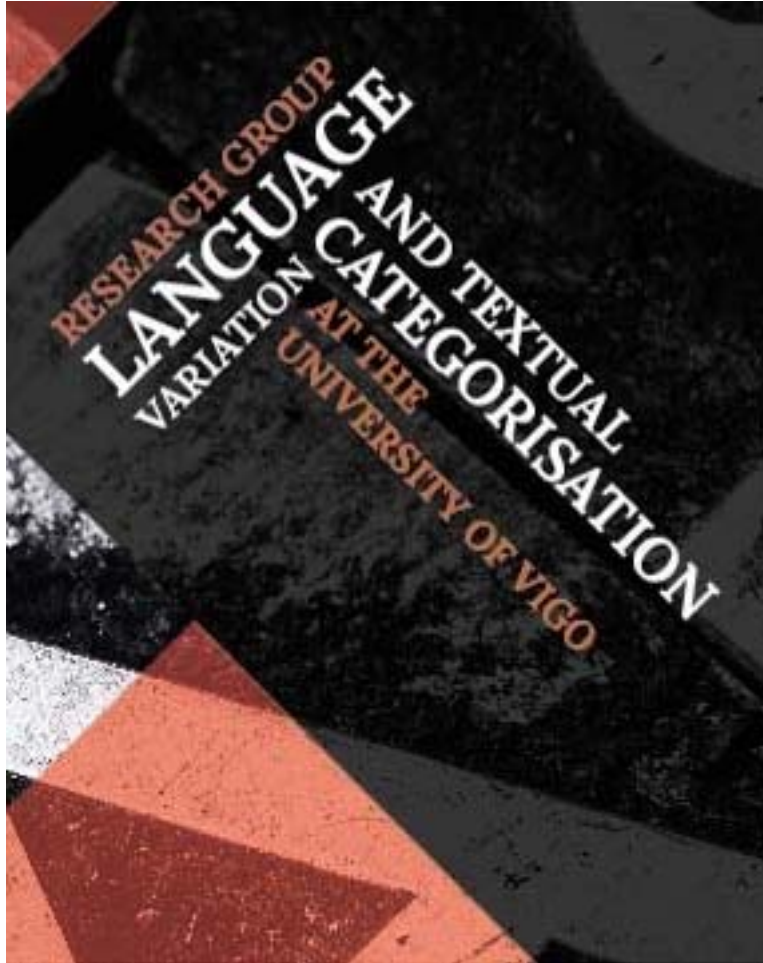


lvtc



Word order in the recent history of English: syntax and processing on the move

Javier Pérez-Guerra (jperez@uvigo.es)

Language Variation and Textual Categorisation Research Group
University of Vigo

UCREL Corpus Research Seminar (CRS)

Lancaster University, 19 Jan 2017

UniversidadeVigo



Research group **LVTC (Language Variation and Textual Categorisation)**:

- diachronic variation (mainly, syntax): EModE>PDE
- diatopic variation (World Englishes)
- diachronic text-type characterisation (speech-based/purposed vs written text types)
- textual linguistics (Systemic Functional Grammar)
- linguistic complexity: across time, L2 English
- empirical (corpus-based/driven) approach

Today

- Two pieces of research on the order of constituents in the clause (time permitting):
 - verb-object vs object-verb in the recent history of English:
 - People love British coffee.
 - *?People British coffee love.
 - complement-adjunct vs adjunct-complement in the history of English:
 - People love British coffee in the morning.
 - People love in the morning British coffee.

Verb-object vs object-verb in the recent history of English

Goal

- **OV** [Object-Verb] in (the recent history of) English:

The old men [young girls]_{obj} married. (READE-1863,219.452)

- Kayne (1994):

- VO is the basic (underlying) word order in English.
- OV surfaces as the result of leftward movement.
- Light elements (pronouns and particles), and not full NPs, can undergo leftward movement.
- So... OV is a **marked** configuration of the clause

Outline

- Some history
- Goal
- Data
- Analysis of the data
- Conclusions

Some history

Old English (OE) (Pintzuk 1991, Moerenhout and van der Wurff 2010):

- Both OV and VO in OE (Fischer and van der Wurff 2006: 185: ‘OV with V2’ grammar).

OV₁: OvV:

þe æfre on gefeohte his handa wolde afylan

who ever in battle his hands would defile

‘whoever would defile his hands in battle’

(*Ælfric’s Lives of Saints* 25.858; Pintzuk 1999: 102)

OV₂: vOV:

He ne mæg his agne aberan

he not can his own support

‘He cannot support his own’ (*CP* 7.53.1; Moerenhout and van der Wurff 2005: 85)





Some history

VO:

Ælfric munuc gret ÆDelwærd ealdormann eadmodlice.



Ælfric monk greets Æthelweard nobleman humbly

'The monk Ælfric humbly greets the nobleman Aethelweard.' (*ÆGenPref* 1)

- Fischer and van der Wurff (2006: 185): “OE verbs are usually in clause-final position”, so VO would be a “complication” (“a finite verb is moved to second position in main clauses”)
- OV was frequent:
 - with pronominal objects 
 - with ‘particles’ 
 - in subordinate clauses 
 - in main clauses with auxiliaries 




Some history



Early Middle English (EME) (Allen 2000, Kroch and Taylor 2000, Koopman 2005):

- OV and VO:
 - Trips (2002): almost rigid VO
 - Fischer and van der Wurff (2006: 187): “steady decline” of OV
 - Moerenhout and van der Wurff (2000): OV is less frequent but it does not disappear
- Kroch and Taylor (2000):
 - end-weight role: postverbal objects tend to be somewhat longer than preverbal objects => pronominal objects tend to be preverbal 
 - quantified objects tend to be preverbal 

Some history

Late Middle English (LME) (van der Wurff 1997, Moerenhout and van der Wurff 2000, Ingham 2002):


- OV and VO, the former limited in non-literary English  exclusively to these patterns:
 - clauses with auxiliaries, ie. vOV (Ingham's 2002 'embraciated') 
 - with negated/quantified objects: 

Ingham (2002): 90% of OV clauses have negated objects, so Neg movement of the object to SpecNegP (between Infl and VP), a type of movement which is no longer available in PDE (Ingham 2000: 34: Neg movement is a form of A'-movement and thus optional)
 - (coordinated clauses 
 - nonfinite clauses) 

./..

Some history

../..

- van der Wurff and Foster (1997a): OV survived “much more tenaciously than suggested”; van der Wurff and Foster (1997b: 147): not merely a survival or an archaism but fulfilled an information-packaging given-new function – “OV in late ME prose is anti-triggered by new objects”. 

Some history

Early Modern English (EModE) (van der Wurff and Foster 1997, Fischer and van der Wurff 2006, Moerenhout and van der Wurff 2005: 187):

- 1500–1550: “OV survives productively” (van der Wurff and Foster 1997a: 84): 0.37/1,000w
- 1550–:
 - OV dwindles away outside poetry (Rissanen 1999: 267: “exceptional”) 💡
 - van der Wurff and Foster (1997a): only 42% with pronominal objects, so... *given-new strategy (“the association between OV and pronominal objects seem to be lost in the course of time”, p.451) 💡

Some history

Present-Day English (PDE):

- van der Wurff and Foster (1997b): OV is an archaism
- Takizawa (2012): OV (only with *make*): 79 examples in the Bank of English (520 mio words)

Goal

- (initially:) OV in the recent history of English: EModE, LModE (and PDE)
- data from larger balanced multi-genre corpora:
 - previous studies were based on genre-specific corpora (eg. letters) or on small corpora
 - importance of balance since the distribution of OV is very different across genres – eg. in prose and in poetry in 14th and 15th century English:
 - Foster and van der Wurff (1995):
 - ~1340: OV is 6 times more frequent in poetry
 - ~1400: OV is 10 times more frequent in poetry
 - ~1470: OV is 20 times more frequent in poetry
- application of a widely accepted statistical model

Data

- Corpora:
 - for Early Modern English (EModE; 1500-1710), the **Penn-Helsinki Parsed Corpus of Early Modern English** or **PPCEME** – 1,737,853 words from the Helsinki directories of the Penn-Helsinki Parsed Corpus of Early Modern English, plus two supplements (Kroch et al. 2004)
 - for (Late) Modern English (LModE; 1700-1914), the **Penn Parsed Corpus of Modern British English** or **PPCMBE** – 948,895 words (Kroch et al. 2010)

Data

node: IP*

query: ((IP* idoms *SBJ)

AND (IP* idoms *OB* | CP-THT | CP-QUE)

AND (IP* idoms VA* | VB* | BA* | BE* | DA* | DO* | HA* | HV*)

AND (*SBJ precedes
VA* | VB* | BA* | BE* | DA* | DO* | HA* | HV*)

AND (*SBJ precedes *OB* | CP-THT | CP-QUE)

AND (*OB* | CP-THT | CP-QUE precedes
VA* | VB* | BA* | BE* | DA* | DO* | HA* | HV*))

- CP-THT (eg. *Craig (that) it was going to rain in Lancaster announced*), not bracketed as OB
- CP-QUE (eg. *Craig when it is going to rain asked*), not bracketed as OB
- participles: BA (of *be*), DA (of *do*), HA (of *have*), VA (of other verbs)
- verbs other than participles: BE, DO, HV, VB

Data

- OV frequencies

		examples	words	nf/1,000w
EModE1	1500-1569	165	567,795	0.29
EModE2	1570-1639	60	628,463	0.10
EModE3	1640-1710	9	541,595	0.02
LModE1	1700-1769	2	298,764	0.01
LModE2	1770-1839		368,804	0.00
LModE3	1840-1914	1	281,327	0.00

Data

- OV frequencies

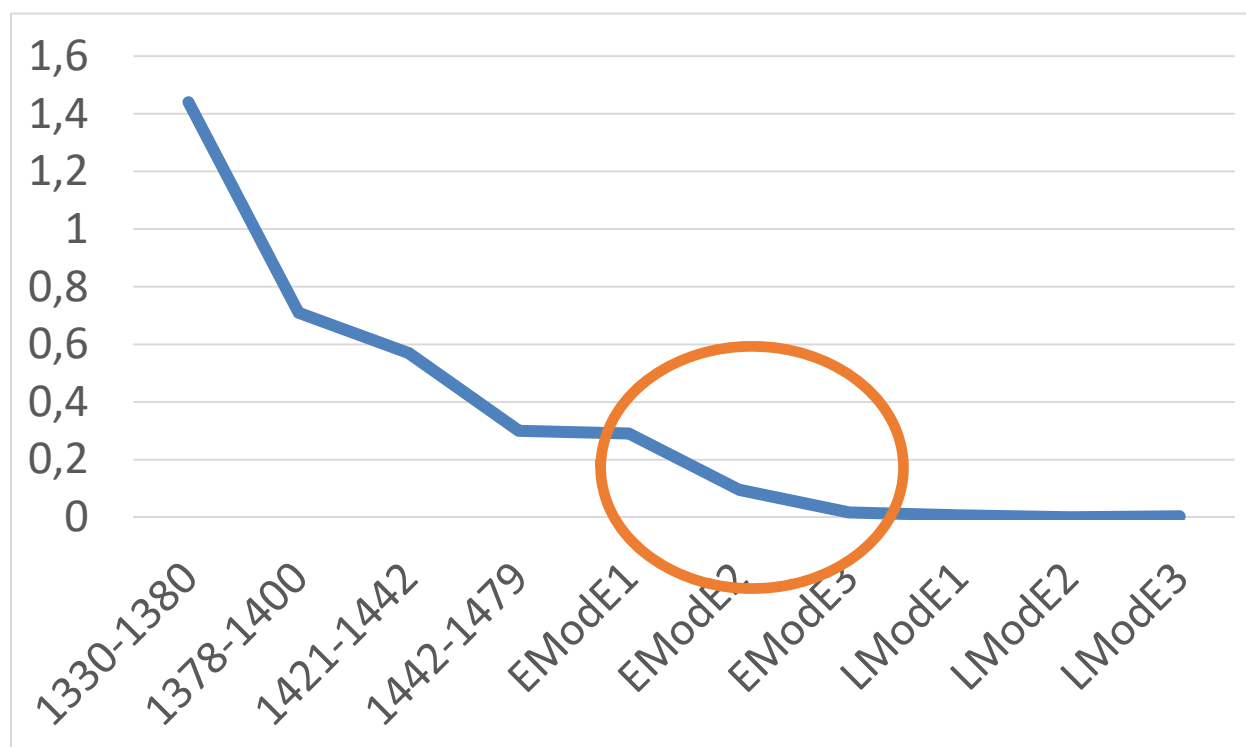
	nf/1,000w	source
1330-1380	1.44	[2]
1378-1400	0.71	[1]
1421-1442	0.57	[1]
1442-1479	0.30	[1]
EModE1	0.29	
EModE2	0.10	
EModE3	0.02	
LModE1	0.01	
LModE2	0.00	
LModE3	0.00	

[1] Moerenhout and van der Wurff (2000), *Paston Letters*

[2] Foster and van der Wurff (1995)

Data

- OV frequencies



n.f./1,000w

Data

So... (definitive goal:) focus on **EModE**.

RQ: forces shaping OV in EModE

- Determining the EModE database size:
 - examples of OV in PPCEME: 234
 - examples of VO in PPCEME: 49,047
 - examples VO+OV in PPCEME: 49,281
 - R (The R Project for Statistical Computing, <https://www.r-project.org>): function 'n.for.survey' (library epiDisplay) to determine the min. database size:
`n.for.survey(p=.08, delta=.02, popsize=49281, alpha=0.05)`
Sample size = 697 (min.)

Analysis of the data

- Determining the (initial) variables:
 - textual:
 - genre
 - object length
 - category of object
 - linguistic:
 - patterns
 - co-occurrence with auxiliaries
 - discontinuity
 - particles
 - finiteness
 - main/subordinate clause
 - (c/)overt subject
 - subject length
 - semantic, discourse-related:
 - quantified objects
 - negated objects

Analysis of the data

Genre (based on Culpeper and Kytö 2010):

writ	writing-based/purposed/like	educ-treatise
		history
		law
		science-medicine
		science-other
		travelogue
		biography-auto
		biography-other
		fiction
		handbook-other
speech	speech-based/purposed/related	diary-priv
		drama-comedy
		letters-non-priv
		letters-priv
		proceeding-trials
		sermon
phil	philosophy	

Analysis of the data

VO patterns

- SVO:
 - **SVO**: *but the Trinity keep you.* (APLUMPT-E1-H,185.85)
 - **SvVO**: *when he was building that admirable worke of his tombe* (ARMIN-E2-H,46.410)
 - **SVXO**: *He had no sooner the liberty of his tongue, but that he curst and swore like a diuel:* (DELONEY-E2-P2,51.297)
 - **SvVXO**: *but by her cheeks you might find guilty Gilbert* (ARMIN-E2-P2,39.298)
 - **SvXVO**: *the middle letter doth alwayes signifie the Angle propounded,* (BLUNDEV-E2-P2,57V.18)
 - **SvXvVO**: *that I shoulde thus haue refused the oth.* (MORELET2-E1-H,506.44)
 - **SvXVXO**: *And if any one shall throughly weigh in his Mind the Force and Energy of the one and of the other,* (BOETHPR-E3-H,191.376)

Analysis of the data

VO patterns

- SVO:
 - **SXvVO**: *I truly can accuse you of none.* (THOWARD2-E2-P2,101.55)
 - **SXVXO**: *And in this yere the kynge at the Request of the duke of Orleauce sent ouer the foresayd duke his sone* (FABYAN-E1-H,174V.C2.196)
- inverted subjects:
 - **VSO**: *Ford. Has Page any braines?* (SHAKESP-E2-P1,49,C1.876)
 - **vSVO**: *And thus do the best Divines expound the Place.* (JUDALL-E2-P2,1,175.312)
 - **vSVXO**: *L. C. J. Did my Lady Lisle ask you that Question?* (LISLE-E3-P2,4.118.337)
 - **vSXVO**: *should we therefore judg those who retain their Sight to be blind also?* (BOETHPR-E3-H,183.330)

Analysis of the data

VO **patterns**

- subjectless:
 - **OVO**: and *O* saw great danger on both hands: (BURNETCHA-E3-P1,2,171.260)
 - **OvVO**: and *O* will emploie all other meanes possible, (EDMONDES-E2-H,394.23)
 - **OvXO**: and *O* kepe close such matters. (LATIMER-E1-H,38L.351)
 - **OvXVO**: and would eat as much at one time as *O* might very well serve four or five ordinary men, (PENNY-E3-P1,33.196)

Analysis of the data

OV **patterns**

- OV:
 - **SOV**: This profe I trow may serue, though *I no word spoke*. (STEVENSO-E1-H,54.218)
 - **SOXV**: *God all Rules by goodnes order* (BOETHEL-E2-P2,71.256)
 - **SXOV**: *who for like faulte out of the citie the name of kings abolisshed*. (BOETHEL-E2-P1,34.464)
 - **SXOXV**: *And Goodlucke I dare sweare, your witte therin would low*. (UDALL-E1-P2,L1563.786)

Analysis of the data

OV patterns

- vOV:
 - **SvOV**: alledging that *he hath nothing done*, (WOLSEY-E1-H,2.2,21.17)
 - **SvOXV**: *I shall hir no more see*. (UDALL-E1-H,L.1111.442)
 - **SvXOV**: *We should therat such a sporte and pastime haue founde*, (UDALL-E1-P2,L1563.780)
 - **SXvOV**: Here *Martin luther for his shrewed brayne wyll some thyng wrastell agaynst vs*. (FISHER-E1-P2,337.68)
- vOV_inversion:
 - **vSOV**: C. Cust. *Will ye my tale breake?* (UDALL-E1-P2,L1469.671)
 - **vSOXV**: T. Trusty. *Do you that part wel play* (UDALL-E1-P2,L1594.797)
 - **vSXOV**: *So shall we pleasantly bothe the tyme beguile now, And eke dispatche all our workes ere we can tell how*. (UDALL-E1-H,L.297.196)

Analysis of the data

OV patterns

- subjectless:
 - **OOV**: nor also *O none can haue*. (MORERIC-E1-P1,32.135)
 - **OOXV**: and *O hym myserably in his Chaumbre slewe* (FABYAN-E1-H,170R.C1.85)
 - **0vOV**: But I woulde be auenged in the meane space, On that vile scribler, that *O did my wowyng disgrace*. (UDALL-E1-H,L.1145.493)
 - **OXOV**: And *O by and by them opened*, euen as they were before, (STEVENSO-E1-H,14.147)
 - **OXvOV**: ich trust *O soone shalt it see* (STEVENSO-E1-P1,33.539)

Analysis of the data

Pattern simplification

- $vV > V$, to avoid interaction with **auxiliary/no_aux**
- SXv or $SXV > S$, since we are focusing on $[(v)V...O]$
- No difference is made between subjectless examples and those with subjects to avoid interaction with **subj/subjectless**
- verb-first examples will not be considered specific patterns (interrogatives, exclamatives, inversions) to avoid interaction with **v_first/non-v_first**
- $VXO > VO$, to avoid interaction with **continuous/discontinuous**

Analysis of the data

Pattern simplification

- OV: collinearity with response variable (ov)
- VO: collinearity with response variable (vo)
- vOV: collinearity with response variable (ov) and auxiliary
- vXOV: only partial collinearity with response variable (ov)
- vXVO: only partial collinearity with response variable (vo)

To avoid collinearity with the response variable (ov/vo) and the variable auxiliary, the list of patterns were replaced with the variable:

- **intervening material** following v (mat): vXVO, vXOV
- no intervening material following v (no_mat)

Analysis of the data

Auxiliary (v)

- auxiliary
- no_aux

Continuous (X, between V and O [VXO], or O and V [OXV])

- continuous
- discontinuous

Verb-first:

- v_first
- non-v_first

Analysis of the data

Particles

- And there was a Justice of peace had taken *away* much of friends goods: (FOX-E3-P2,109.140)

Finiteness

- finite
- infinitive: And thus I desyre *our Lorde to have you in his moste gracious tuytion*. (GCROMW-E1-P1,209.9)
- *ing* clause: *The Priest and the Tanner seeing the Taylor*, mused what hee made there: (DELONEY-E2-P1,16.253)
- (no examples of *ed* clauses in the corpus)

Analysis of the data

Main/Subordinate/Coordinated clause

- main
- subordinate: for I thinke *so God me mende*, This will proue some foolishe matter in the ende. (UDALL-E1-P2,L751.17)
- coordination: “Then that is the top of felicitie, that stowtly rules & *0 gently all disposith.*” (BOETHEL-E2-P2,71.264)

(C/)Overt subject

- with overt subject
- subjectless

Analysis of the data

Subject length (ordinalisation>factorising)

- average: 0-2 words (771 examples)
- long: 3-6 words (89 examples)
- very long: 7-22 words (13 examples)

Object length (ordinalisation>factorising)

- average: 1-3 words (628 examples)
- long: 4-9 words (187 examples)
- very long: 10-32 words (45 examples)

Analysis of the data

Quantified object

- definite
- indefinite (inc. zero)
- cardinal
- ordinal

Negated object

- non-negated
- negated: M. Mery. Nay fayth ye shall promise that he shall *no harme* haue, (UDALL-E1-H,L.1179.505)

Analysis of the data

Category of object

- pronominal: only a non-*wh* pronoun (*me, I, mine*)
- NP: NP including a noun
- other: eg. clauses (, *wh*-elements)

Analysis of the data

- response variable: ov/vo
- (definitive) variables:
 - textual:
 - genre (simplified)
 - linguistic:
 - intervening material
 - co-occurrence with auxiliaries
 - discontinuity
 - particles
 - finiteness
 - main/subordinate clause
 - (c/)overt subject
 - subject length (ordinal)
- object length (ordinal)
- category of object
- semantic, discourse-related:
 - quantified objects
 - negated objects

Analysis of the data

Logistic regression analysis: R, functions glm and lmr

	Estimate	Std. Error	z value	Pr(> z)	
auxiliary [T.no_aux]	-1.064e+00	2.342e-01	-4.543	5.55e-06	***
continuous [T.discontinuous]	3.563e-01	3.702e-01	0.962	0.335903	
finiteness [T.finite]	-1.596e+01	1.075e+04	-0.001	0.998816	
finiteness [T.inf]	-1.884e+01	1.075e+04	-0.002	0.998602	
finiteness [T.ing]	-1.748e+01	1.075e+04	-0.002	0.998703	
genre2 [T.speech]	-1.506e+00	3.452e-01	-4.363	1.28e-05	***
genre2 [T.writ]	-3.211e+00	3.955e-01	-8.118	4.72e-16	***
main_sub [T.main]	4.049e-01	3.430e-01	1.180	0.237833	
main_sub [T.sub]	8.972e-01	3.332e-01	2.693	0.007090	**
mat [T.no_mat]	1.935e+00	5.832e-01	3.317	0.000908	***
neg_obj [T.non-neg]	-2.473e+00	5.320e-01	-4.648	3.34e-06	***
obj_length	-2.698e-01	8.251e-02	-3.270	0.001076	**
object [T.other]	-1.602e+01	9.188e+02	-0.017	0.986089	
object [T.pro]	8.371e-01	2.684e-01	3.119	0.001818	**
particles [T.particles]	-2.378e+00	1.114e+00	-2.135	0.032736	*
quantif_obj2 [T.definite]	1.851e+01	1.960e+03	0.009	0.992468	
quantif_obj2 [T.indefinite]	1.650e+01	1.960e+03	0.008	0.993283	
subj_length	2.294e-01	8.593e-02	2.670	0.007590	**
subjectless [T.subjectless]	2.110e+00	3.746e-01	5.631	1.79e-08	***
v_first [T.v_first]	-1.806e+01	1.789e+03	-0.010	0.991946	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Analysis of the data

Logistic regression analysis: R, functions glm and lmr

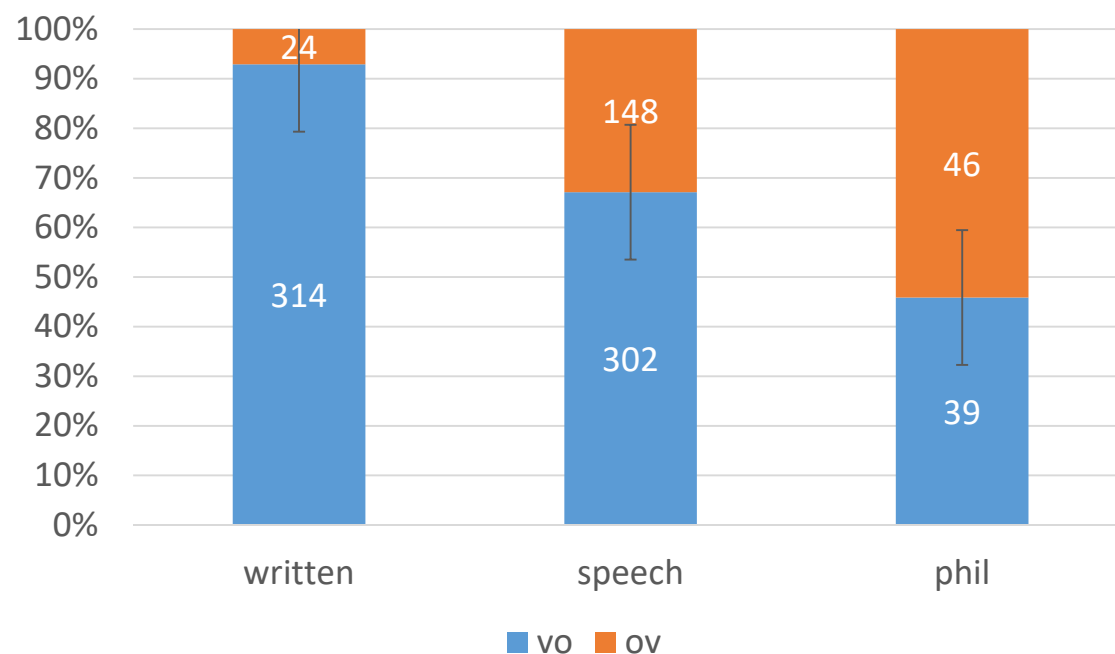
	Pr(> Z)
genre	<0.0001
neg_obj	<0.0001
auxiliary	0.0001
interv. material	0.0005
obj_length	0.0015
object	0.0046
particles	0.0327
main_subord	0.4818
subj_length	0.5118
continuous	0.5132
verb_first	0.8505
quantif_obj	0.8570
finiteness	0.9796

Discrimination indexes:

- (Nagelkerke) $R^2=0.540$
(very good if $>.5$)
- C (Concordance)=0.903
(outstanding if $>.9$)

Analysis of the data

Variable **genre**

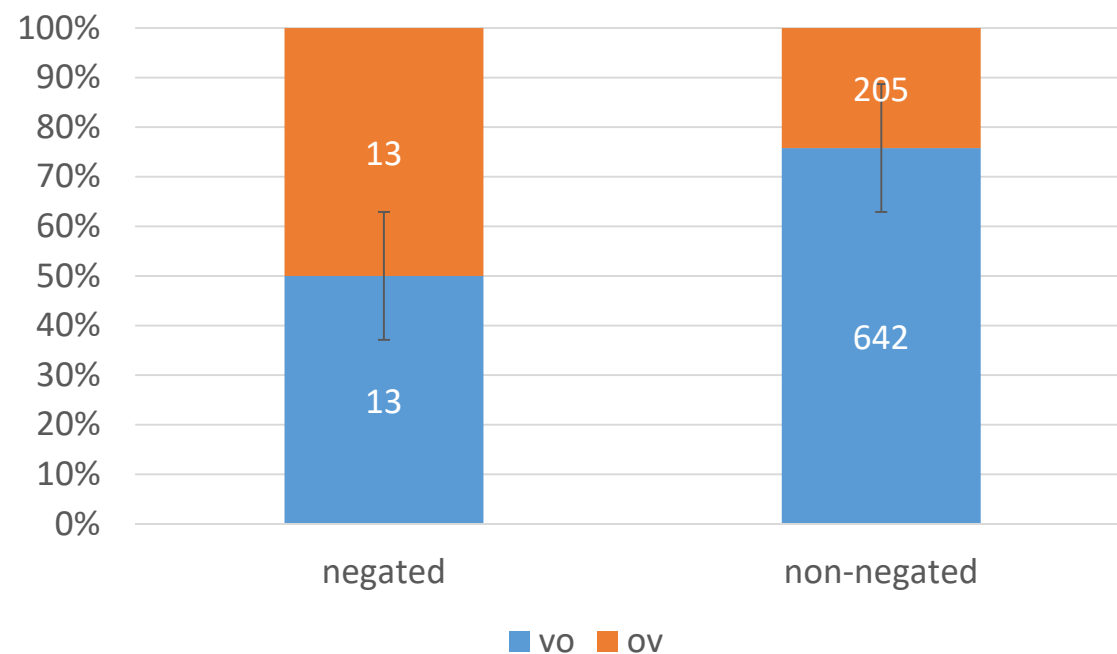


written vs speech: $\chi^2(1)=73.73$, $p<.0001$

speech vs phil: $\chi^2(1)=12.04$, $p=.0003$

Analysis of the data

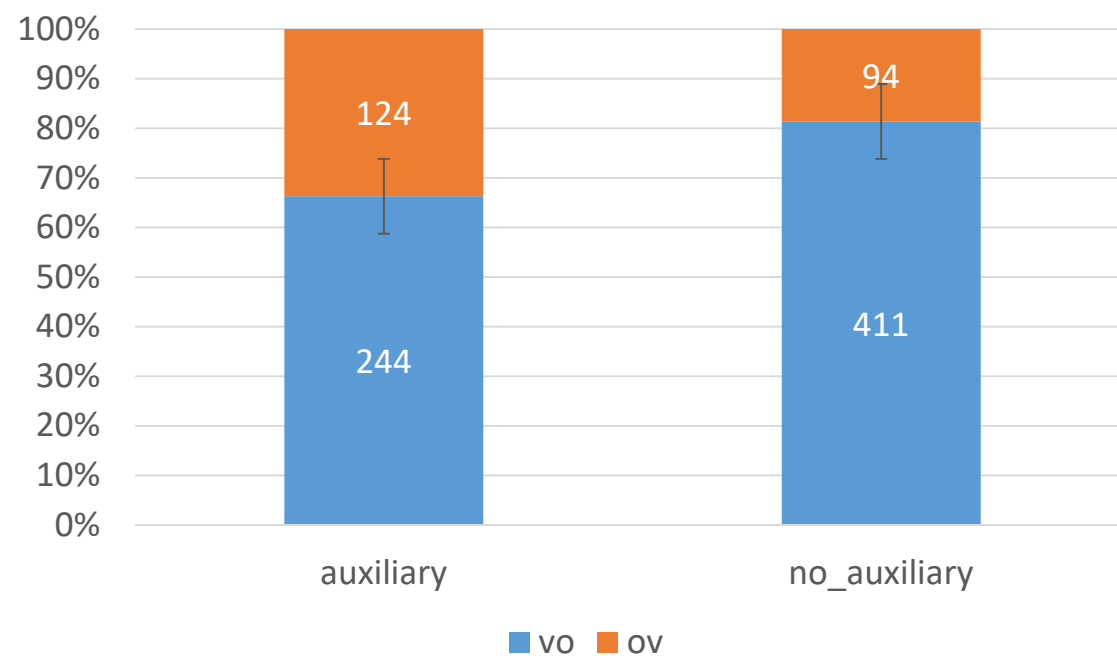
Variable **negated object**



$\chi^2(1)=7.64, p=.0057$

Analysis of the data

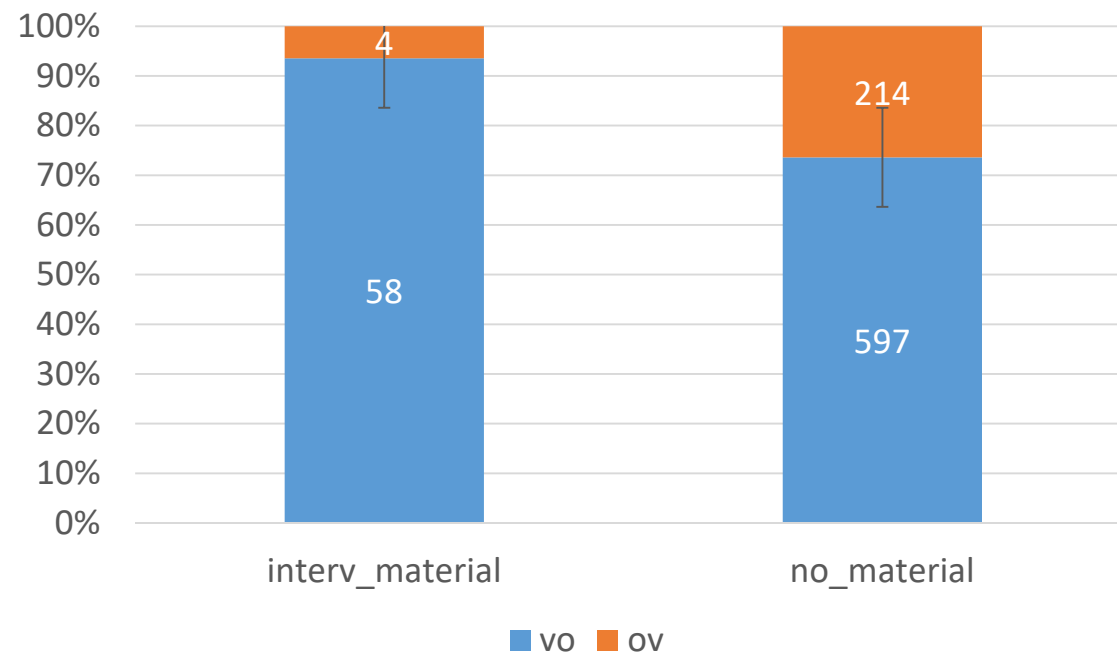
Variable **auxiliary**



$\chi^2(1)=25.05, p<.0001$

Analysis of the data

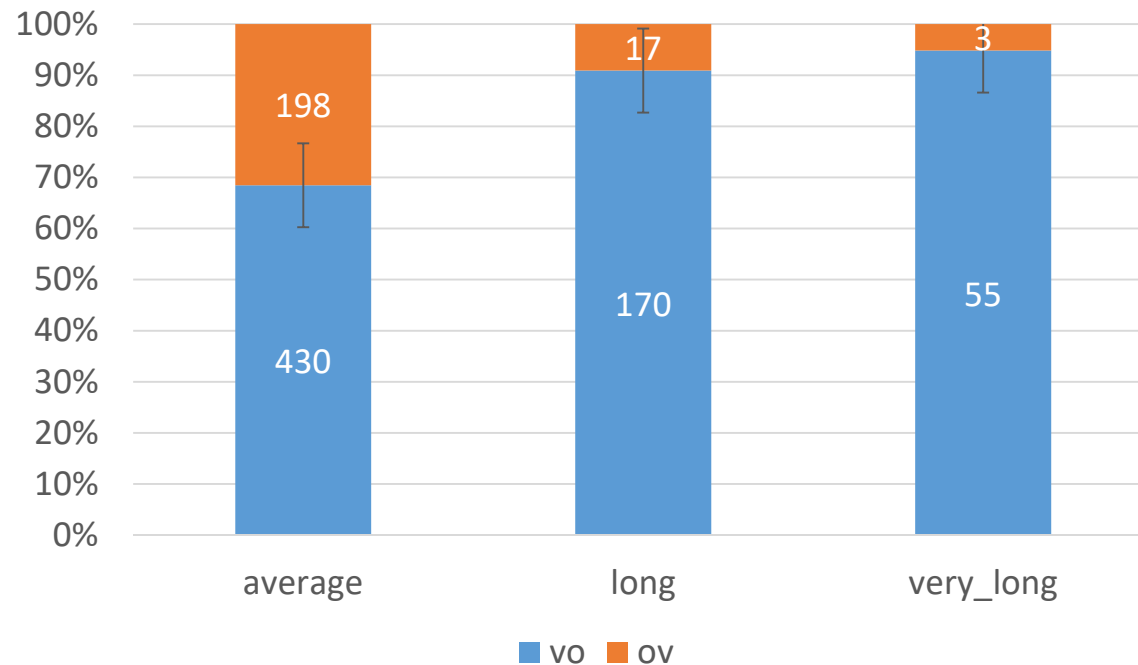
Variable **intervening material**



$\chi^2(1)=11.18, p=.0008$

Analysis of the data

Variable **object length**

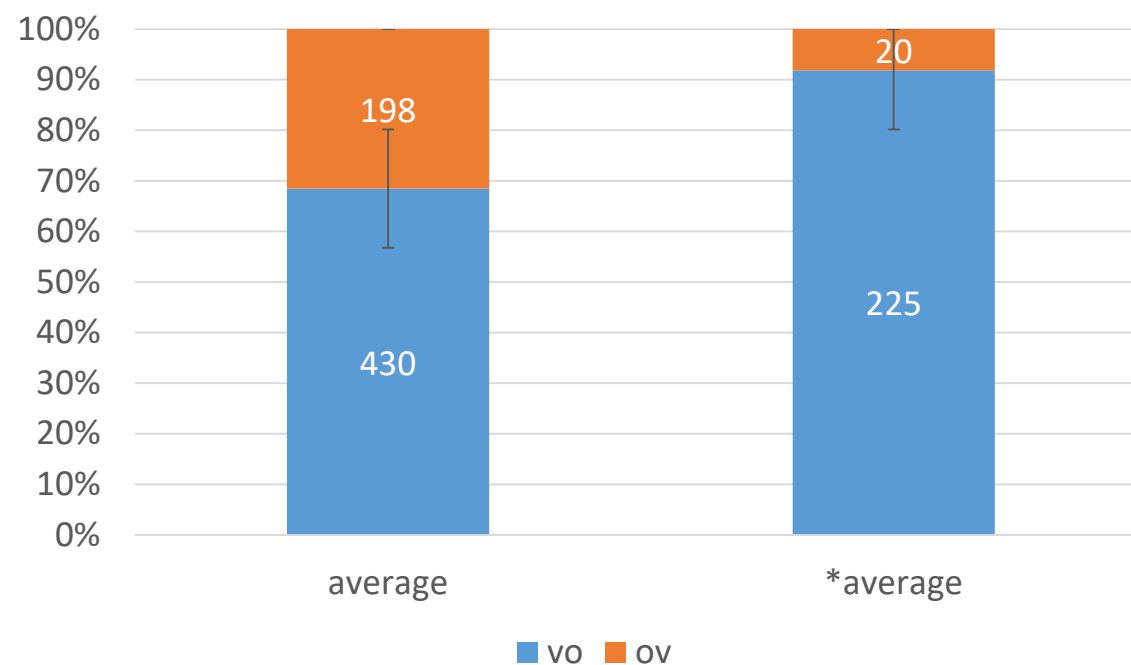


average vs long: $\chi^2(1)=36.21$, $p<.0001$

long vs very_long: Fischer(1), $p(\text{two-tailed})=.4221$

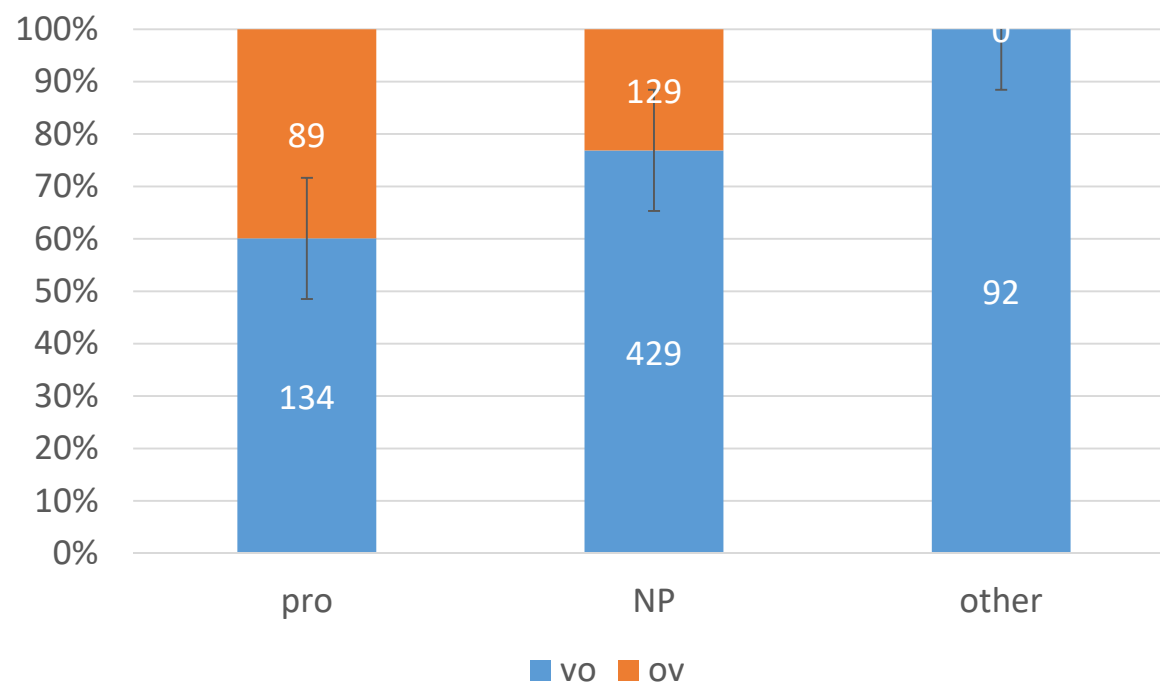
Analysis of the data

Variable **object length** (recodified)



Analysis of the data

Variable **object type**




pro vs NP: $\chi^2(1)=21.5$, $p<.0001$

Summary and conclusions

- Goal (I): study of OV in the recent history of English
- Frequency: statistically marginal in EModE =>
=> lack of evidence in LModE
- Goal (II): statistical analysis of the forces favouring OV in EModE
- Data: PPCEME
 - OV: 218 examples (234 inc. Bible)
 - VO: 655 randomised examples
- Analysis of 13 variables
- Logistic regression analysis: 6 sufficiently explanatory variables


Summary and conclusions

- OV is favoured in speech-based/related/purposed and 'speechy' (inc. Philosophy) text types.
- OV is favoured by negated objects.
- OV is favoured by auxiliaries in the verbal group.
- OV is disfavoured by lexical material between v and V/O (vXVO, vXOV). 
- OV is favoured by short and average (in length) objects.
- OV is favoured by pronominal objects.

Summary and conclusions

(i) Textual (performance) issue as a trigger of OV: the speechier, the greater the frequency of OV

(ii) Prominence of end-weight as the triggering force of OV: preference for reduced lexical complexity of the object:

- objects:
 - shorter objects
 - pronominal objects
- verbal groups:
 - with auxiliaries (short objects and ‘expanded’ verbal groups)
 - without intervening material between auxiliary and rest of the predicate (vXOV)  (maybe reinforces the desired effect of shortening of the object or the non-verbal part of the predicate)
- So... OV in EModE already accommodated within the principles ruling performance in Modern English (end-weight).

Complement-adjunct vs adjunct-complement in the history of English

Outline

- Assumptions
- Goals
- Data
- Analysis of the data:
 - complements-first
 - end-weight
- Conclusions and further research
- References

Assumptions

- Dependents in phrases: complements vs adjuncts
 - Complements:
 - reserved positions in the clause:
Huddleston and Pullum (2002: 225): “[c]omplements are more restricted than most adjuncts as to what positions they can occupy in the clause. In general, there is a basic or default position for a given kind of complement”
 - semantically selected or subcategorized:
Matthews (2007: 187): “unit in a construction either required or specifically taken by an individual member of a lexical category”
Matthews (1981: 124-127): impossibility of dropping (if dropped, then latent)

Assumptions

- Dependents in phrases: complements vs adjuncts
 - Complements:
 - exclusion when the pattern is saturated
 - syntactic dependencies; eg. lexical restrictions or formal determination (Greenbaum et al. 1996: 76): {*deal, compliance*} + *with-PP*; {*assume, certain, hypothesis*} + *that-clause*
 - Adjuncts:
 - loose semantic connection between the adjunct and the head => not required

Assumptions

- Distribution of complements and adjuncts is governed by:
 - syntactic rule: complements precede non-complements (**complements-first**)
 - Quirk et al. (1985: 49-50): 'Complements first'
 - Hawkins (2007): 'Arguments precede X'

Assumptions

- processing: incremental constructionalisation of constituents (**end-weight**):
 - Quirk et al. (1985: 1398): End-weight
 - Hawkins' (2004, 2007) 'Minimize Domains (MiD)': preference for short-long designs:

“Given two or more categories A, B, [...] related by a grammatical rule R of combination and/or dependency, the human processor prefers to minimize the distance between them within the smallest surface structure domain sufficient for the processing of R.” (Hawkins 2004: 234)

“[g]iven a structure {A, X, B} (...), the more relations of combinations or dependency that link B to A, the smaller will be the size and complexity of X” (Hawkins 2004: 37)

Assumptions

- processing: incremental constructionalisation of constituents (**end-weight**):
 - Temperly (2007: 315): “If a word has multiple dependent constituents and there is a choice as to their ordering, the shorter one(s) should be placed closer to the parent head”
 - Psycholinguistic argument:
 - Hawkins (2001: 7): “Less demands are made on working memory and there is less expenditure of effort in reaching these structural definitions” (similarly Wasow 2002: 32)
 - Gibson: “syntactic predictions held in memory over longer distances are more expensive (...), and longer distance head-dependents integrations are more expensive” (1998: 8); “each lexical item in a structure has an activation level (...). The lexical activation decays as additional words are integrated” (2000: 11)

Assumptions

- Examples:

(1) I would take [some spending money] [with me].

(2) I would take [with me] [some spending money].

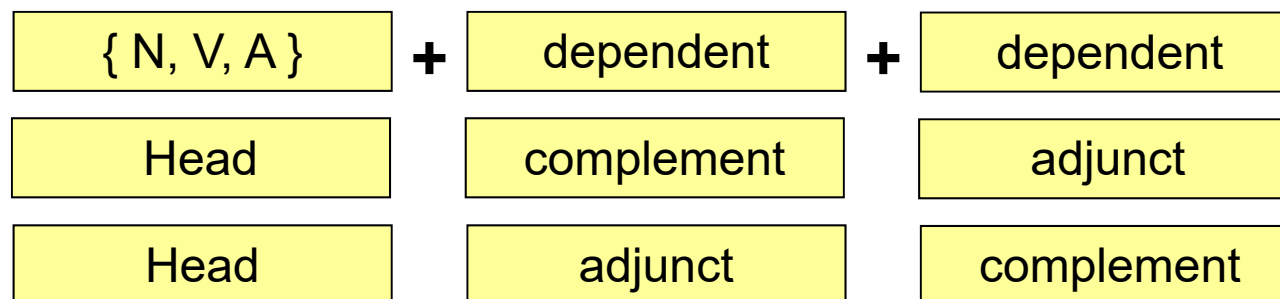
[‘Heavy NP Shift’; see Wasow (2002: 5)]

(1) is claimed to be a better performance solution than (2) on syntactic grounds (complements-first).

(2) is claimed to be a better performance solution than (1) on processing grounds (MiD, end-weight).

Goals

- Account of the distribution of complements and adjuncts in phrases by using a corpus-driven methodology
- Connection between the distribution of complements and adjuncts in phrases and the process of word-order syntacticisation



Data

Connection between the distribution of complements and adjuncts and the process of syntacticisation of English word order:

“loose, paratactic, ‘pragmatic’ discourse structure develop -- over time -- into tight, ‘grammaticalized’ syntactic structures” (Givón (1979: 208-209))

So.. focus on post-ME (EModE, LModE and PDE)

Data

- [Old English: 1.5+ million words (Old English section of the Diachronic Part of the Helsinki Corpus of English Texts, with certain additions, c750–): Taylor et al. (2003) *The York-Toronto-Helsinki Parsed Corpus of Old English Prose.*]
- [Middle English: 1,155,965 words (Middle English section of the Diachronic Part of the Helsinki Corpus of English Texts, with certain additions and deletions, 1150–1500): Kroch and Taylor (2000) *Penn-Helsinki Parsed Corpus of Middle English, second edition.*]

Data

- Early Modern English: 1,737,853 words (the Helsinki directories of the Penn-Helsinki Parsed Corpus of Early Modern English plus two supplements; 1500–1710): Kroch et al. (2004) *Penn-Helsinki Parsed Corpus of Early Modern English*.
- Late Modern English: 948,895 words (1700–1914): Kroch et al. (2010) *Penn Parsed Corpus of Modern British English*.

Data

- Present-Day English: approx. 2 mio words (1961–1989): *The Penn Treebank 3* (1 mio words of The Brown Corpus plus 1 mio words from 1989 Wall Street Journal; Switchboard corpus excluded)

Data

- parsed corpora, with (almost) identical similar parsing conventions
- parsed files (.psd/.mrg), using P&P-based part-of-speech and syntactic tags
- retrieval by means of CorpusSearch (differences among corpora):
 - node: IP***
 - query: (VB* iprecedes W* | QP | PP | RRC | ADJ* | ADV* | CP-* | IP-SUB)**
 - AND (W* | QP | PP | RRC | ADJ* | ADV* | CP-* | IP-SUB iprecedes NP-OB*)**
- (extensive) manual revision

Data

but, if you approve of this, if you please to let me know $y=r=$
pleasure, I will tell it $M=r=$ Isaac.
(ANHATTON-E3-H,2,214.41)

((IP-MAT (CONJ but)

(, ,)

(PP

(P if)

(CP-ADV

(C 0)

(IP-SUB

(NP-SBJ

(PRO you))

(VBP approve)

(PP

(P of)

(NP (D this))))))

(, ,)

(PP

(P if)

(CP-ADV

(C 0)

(IP-SUB

(NP-SBJ (PRO you))

(VBP please)

(IP-INF

(TO to)

(VB lett)

(IP-INF

(NP-SBJ

(PRO me))

(VB know)

(NP-OB1

(PRO\$ y=r=)

(Npleasure))))))

(, ,)

(NP-SBJ (PRO I))

(MD will)

(VB tell)

(NP-OB1 (PRO it))

(NP-OB2 (NPR M=r=) (NPR Isaac))

(. .))

Data

VPs (see also Pérez-Guerra 2016)

- verb group immediately precedes an adjunct, and the adjunct immediately precedes a complement (object)

neither will I againe smite {any more} {euery thing liuing}, as I haue done. (AUTHOLD-E2-H,VIII,20G.466) [QP + OBJ]

and sitting in some place, where no man shall prompe him, by him self, let him translate {into Englishe} {his former lesson}. (ASCH-E1-H,1V.22) [PP + OBJ]

Lisle. My Lord, this Fellow that now speaks against me, broke {open} {my Trunk}, (LISLE-E3-H,IV,120C1.203) [Adjective + OBJ]

Moreouer, there is no one thing, that hath more, either dulled the wittes, or taken {awaye} {the will of children from learning}, then the care they haue, to satisfie their masters, in making of latines. (ASCH-E1-H,1R.9) [Adverb + OBJ]

Data

VPs

- verb group immediately precedes a complement (object), and the complement (object) immediately precedes an adjunct

Will tels {the king} {how Terrils Frith was inclosed}. (ARMIN-E2-H,44.338)
[OBJ + W*]

so this time will trouble {y=r= Losp} {no more} w=th= y=r= most
obedient, duttyful daughter, A. Nottingham. (ANHATTON-E3-H,2,212.29)
[OBJ + QP]

I thoughte I wolde take {some spendyng money} {wyth me} (MERRYTAL-
E1-H,31.148) [OBJ + PP]

and cut {it} {not so close to the Body as to hurt it}, nor yet so long that it
be a Stump, (LANGF-E3-H,122.269) [OBJ + AdjectiveP]

Data

VPs

But my Brother understood {the matter} {aright} (HOXINDEN-1660-E3-H,280.162) [OBJ + Adverb]

The post served {me} {just as it did y=r= Losp}. (ANHATTON-E3-H,2,211.4) [OBJ + CP]

\$I \$'ll ply {him} {that way}, (FARQUHAR-E3-H,9.326) [OBJ + NP-Adverb]

Data

VPs

Beda writes {that he was dead long before}, {although if the time of his sitting Archbishop be right computed sixteen years, he must survive this action}. (MILTON-E3-H,X,150.77, 1670) [*that* cl + concessive adjunct]

Also I read {in Iohannes Libaulty, his Booke Intituled Le Meson Rustick, and also in other Learned Writers}, {that the dung of a Cow heated vnder the Ashes, betwixt Wine or Colwort leaues, & mingled with vineger, hath the property to bring Scrophulous swellings to ripenes, &c}. (CLOWES-E2-H,26.212, 1602) [place adjunct + *that* cl]

Data

NPs (see Pérez-Guerra 2016)

- noun immediately precedes an adjunct, and the adjunct immediately precedes a complement ((*that-* or) infinitive clause)

[The master shewyng us that by] neglygence {of some} {to belay the haylers}, (MADOX-E2-P1,112.434) [PP + IP]

in mind of the great Obligation {that lies on them} {to live sutably to their Profession:} (BURNETROC-E3-P2,122.170) [rel cl + IP]

Data

NPs

- noun immediately precedes a complement ((*that*- or) infinitive clause), and the complement immediately precedes an adjunct

[King James sent a Person down to him, with] Offers {to mitigate his Fine upon Conditions of ready Payment}, {to which his Lordship reply'd, that if his Majesty pleas'd to allow him a little longer time, he would rather chuse to play double or quit with him}: (CIBBER-1740,44.134) [IP + rel cl]

Data

NPs

the duke had got a solemn promise {of the king} {that he would never speak to him of religion}. (BURNETCHA-E3-P2,2,180.98) [PP + *that* cl]

[He would not hearken to this, which made me inclined to believe] a report {I had heard}, {that the duke had got a solemn promise of the king that he would never speak to him of religion}. (BURNETCHA-E3-P2,2,180.98) [rel cl + *that* cl]

And there was a feeling {by no means uncommon, and very deadly}, {that India would be lost for ever, and with it all the glory of England}. (TROLLOPE-1882,177.356) [AdjectiveP + *that* cl]

There is a wise saying {that nine-tenths of the noble work done in the world is drudgery}, {which is often misused as if it meant that nine-tenths of the drudgery done in the world is noble work}. (BENSON-1908,46.109) [*that* cl + rel cl]

Data

APs (see Pérez-Guerra 2016)

- adjective immediately precedes an adjunct, and the adjunct immediately precedes a complement ((*that-* or) infinitive clause)

[And therefore the quickest wittes commonlie may proue the best Poetes, but not the wisest Orators:] readie {of tonge} {to speak boldlie}, (ASCH-E1-P1,4V.34) [PP + IP]

Data

APs

- adjective immediately precedes a complement (*that-* or infinitive clause), and the complement immediately precedes an adjunct

[none was] more willing {to resign} {than she}. (BEHN-E3-P1,163.135)
[IP + *than* cl]

[I haue beene as] careful {to please her} {as euer I was to please mine own mother}, (GIFFORD-E2-H,B1R.60) [IP + *as* cl]

[He told him they were] fully resolv'd {to dye for their Country}, and ready {to fight it out to the last Man, if Occasion requir'd,} {at which Xerxes derided him, as he did before when he spake of the Valour of his Country-men; (HIND-1707,310.144) [IP + rel cl]

Data

APs

[yea I am] sorie, {with all my harte}, {that they be giuen no more to riding, then they be}: (ASCH-E1-P1,10R.186) [PP + *that* cl]

For we are no less certain {that there is a great Town called Constantinople, the seat of the Ottoman Empire}, {than that there is another called London}. (BURNETROC-E3-P1,79.231) [*that* cl + *than* cl]

Analysis of the data: complements-first

Analysis of the data: complements-first

- Pérez-Guerra (2016):

- object + adjunct:

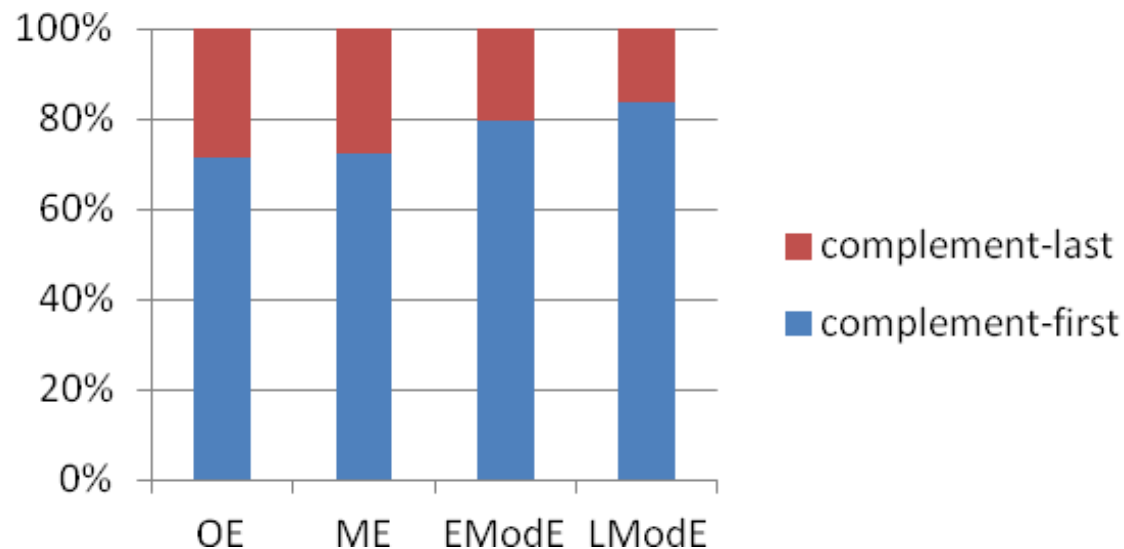
[I thoughte] I wolde take [some spendyng money]_{object} [wyth me]_{adjunct} (MERRY TAL-E1-H,31.148) [complement plus adjunct in a VP]

- adjunct + object:

[and sitting in some place, where no man shall prompe him, by him self,] let him translate [into Englishe]_{adjunct} [his former lesson]_{object}. (ASCH-E1-H,1V.22) [adjunct plus complement in a VP]

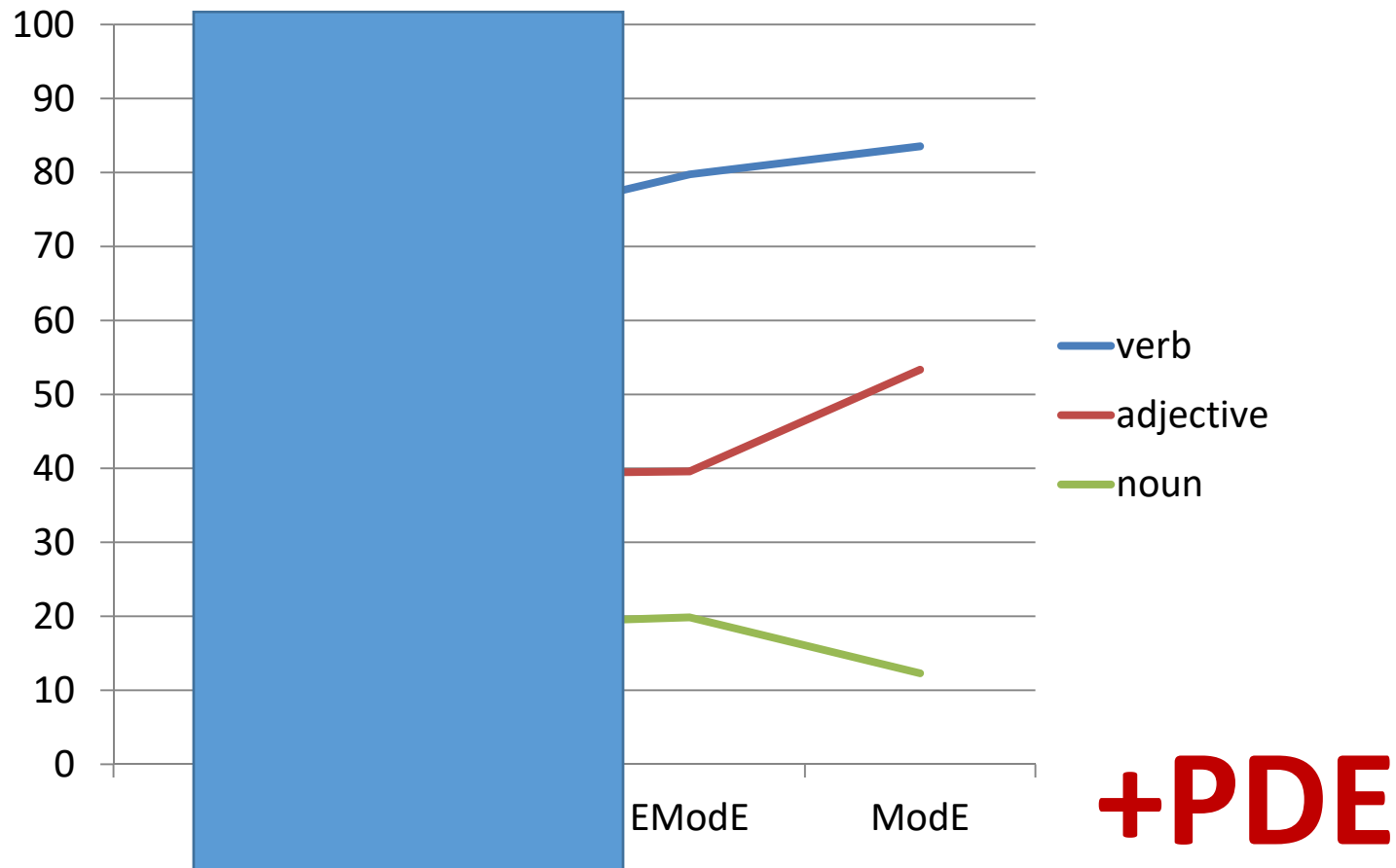
Analysis of the data: complements-first

- Pérez-Guerra (2016):



- Statistical significance for **full variation**: **yes** ($P < .0001$)
- Statistical significance for variation **OE > ME**: **no** ($P = 0.0949$)
- Statistical significance for variation **ME > EModE**: **yes** ($P < .0001$)
- Statistical significance for variation **EModE > LModE**: **yes** ($P < .0001$)

Analysis of the data: complements-first



Analysis of the data: complements-first

- Pérez-Guerra (2016):

- ME>EModE seems to be the pivotal period as far as compliance with complements-first is concerned
- Connection type of head and compliance with complements-first:

VP > **AP** > NP

- VPs:
 - Most VPs are complement-first
 - Statistically significant increase of complement-first VPs from ME to LModE
- Half of the APs are complement-first in LModE
- Most NPs are complement-last

Analysis of the data: complements-first

- Another experiment:
 - also focuses on complements/adjuncts but only after word-order syntacticisation, that is, after ME (ME>EModE as the pivotal period)
 - focuses on only VPs
 - challenges the supremacy of complements-first by investigating its plausibility with structurally long and syntactically complex complements: *that* clauses

Analysis of the data: complements-first

Query example:

node: IP-MAT

query: (IP-MAT iDoms VBP)
AND (IP-MAT iDoms CP-THT)
AND (IP-MAT iDoms *P*)
AND (VBP iprecedes CP-THT)
AND (CP-THT iprecedes *P*)

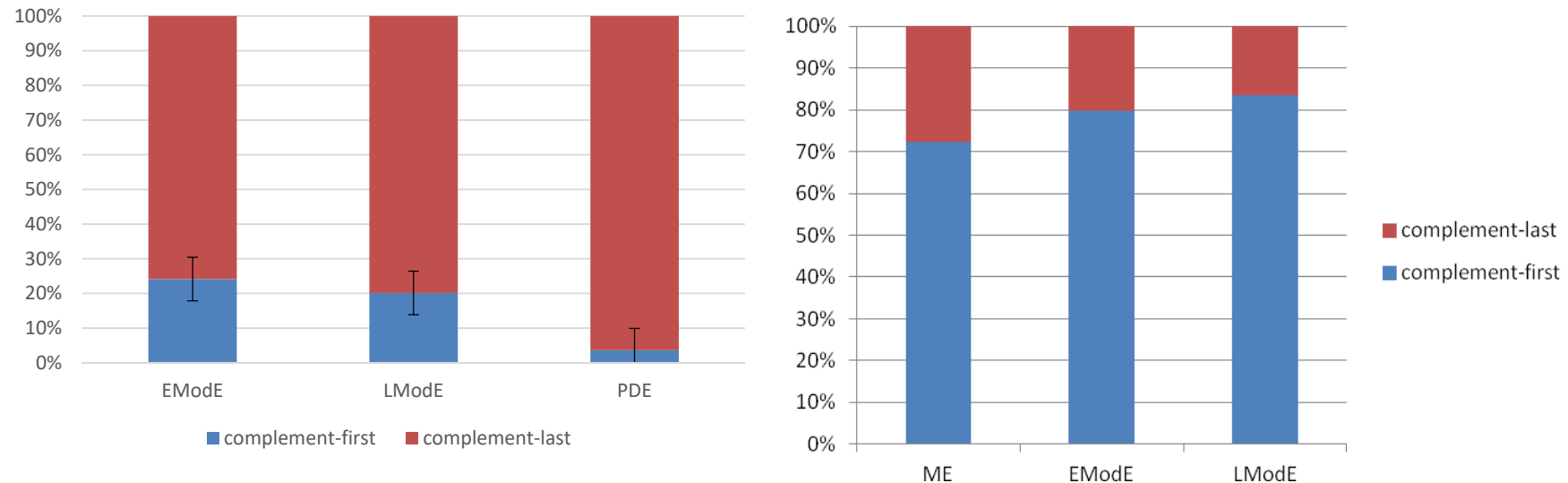
with parsing differences among corpora

Analysis of the data: complements-first

- Examples:
 - Beda writes_v [that he was dead long before,]_{that-cl} [although if the time of his sitting Archbishop be right computed sixteen years, he must survive this action.]_{adjunct} (MILTON-E3-H,X,150.77, 1670) [complement plus adjunct in a VP]
 - Also I read_v [in Iohannes Libaulty, his Booke Intituled Le Meson Rustick, and also in other Learned Writers,]_{adjunct} [that the dung of a Cow heated vnder the Ashes, betwixt Wine or Colwort leaues, & mingled with vineger, hath the property to bring Scrophulous swellings to ripenes, &c.]_{that-cl} (CLOWES-E2-H,26.212, 1602) [adjunct plus complement in a VP]

Analysis of the data: VPs

- Incidence of the type of complement:



only *that*-clauses (this experiment) all types of complements (objects) and adjuncts (Pérez-Guerra 2016)

- So... end-weight is a crucial factor

Analysis of the data

- So... **tension between end-weight and compl-first**
 - VPs:
 - with non-clausal objects, complements-first is the leading force in VPs, and increasing (70%+80% are compl-first)
 - with clausal (*that* cl) objects, complement-last is the leading design in VPs, and decreasing (<4% are compl-first in PDE)
 - NPs:
 - with clausal (*that* and infinitive cl) complements, complement-last is the leading design (0% are complement-first in PDE)

Analysis of the data: end-weight

Analysis of the data: end-weight

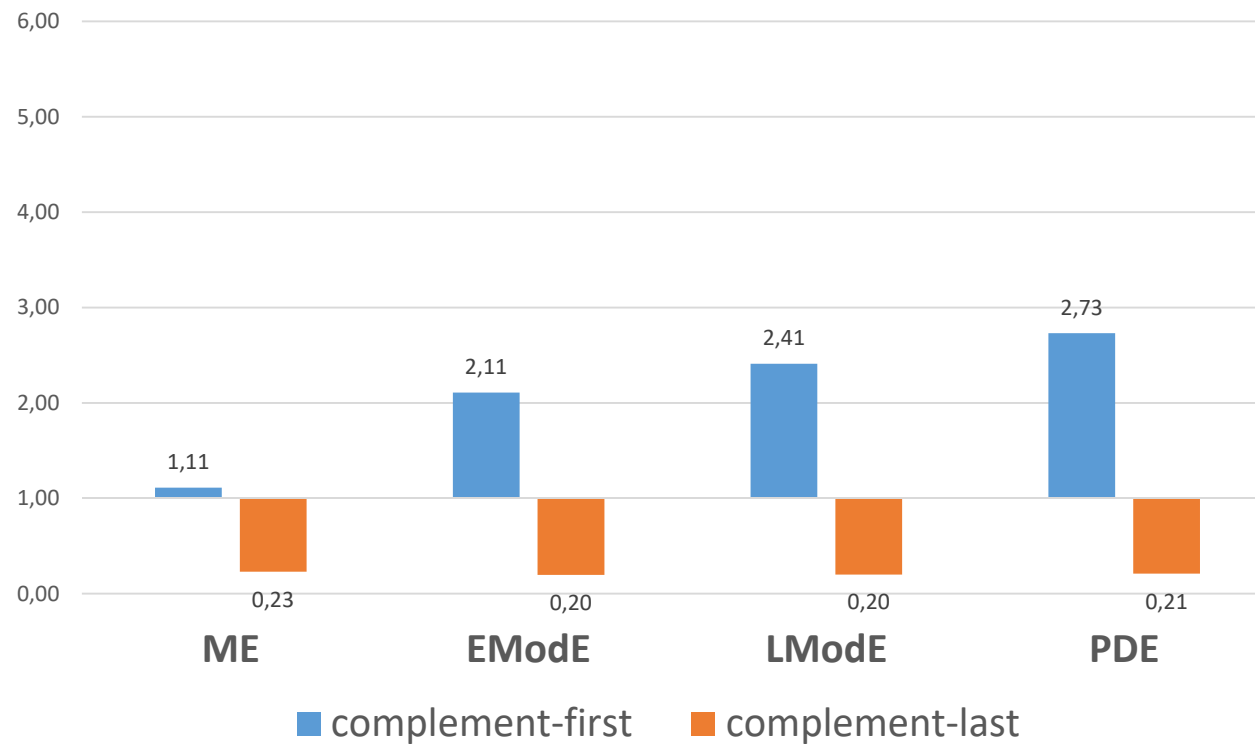
- Stowell (2006: 239):
 - “it has consistently proved to be virtually impossible to define ‘heaviness’ in a satisfactory way”
- For summaries of proposals, see Wasow (1997) and Pérez-Guerra and Martínez-Insua (2010).

Analysis of the data: end-weight

- Metric:
 - Gries (2003: 83-84): no. of syllables, no. of words, no. of morphemes, with very similar results
 - Yaruss (1999: 339): “very strong, positive, significant correlations (...) among measures of length in words, syllables, morphemes, and clausal constituents”
 - Szmrecsányi (2004: 1038): “determining length in words (...) is by all means (...) nearly as accurate as the most sophisticated and cognitively, conceptually, or even psychologically ‘more real’ methods”
 - Shih and Grafmiller (2011): no. of words is a sufficient proxy for weight

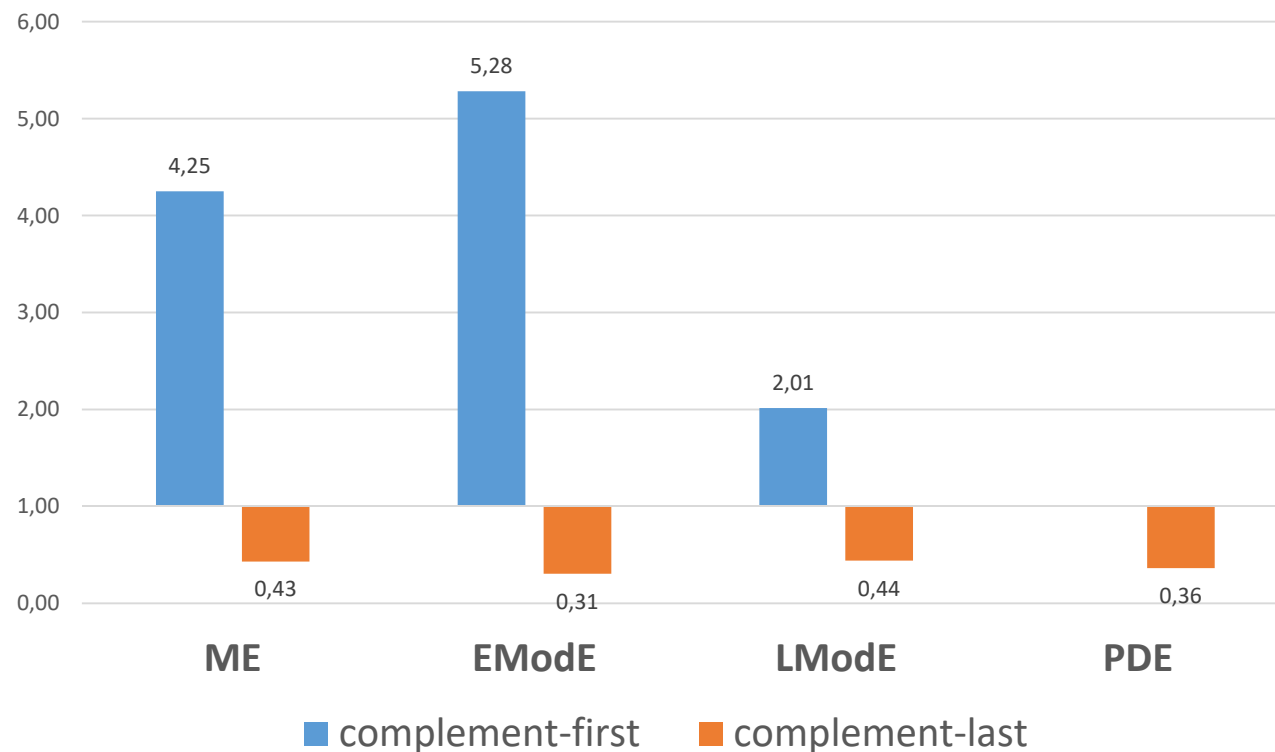
Analysis of the data: end-weight

- Times 1st dependent is longer than 2nd in VPs



Analysis of the data: end-weight

- Times 1st dependent is longer than 2nd in NPs



Analysis of the data: end-weight

- End-weight is a major factor only in complement-last constructions in VPs and NPs: the 1st dependent is notoriously shorter than the 2nd dependent only in complement-last constructions.
- Most complement-first constructions do not comply with end-weight:
 - VPs: 1st dependents are progressively longer across time
 - NPs: 1st dependents are progressively shorter across time

Conclusions

Conclusions

- Two forces:
 - complements-first: complement as the first dependent
 - end-weight: second dependent is longer
- Application to phrases: VPs, NPs and APs
- This study:
 - (ME -) EModE - LModE – PDE, after the syntacticisation of word order in English
 - extreme scenario: (long, complex) *that*-clauses as complements

Conclusions

- Most patterns comply with end-weight (and increasing across time):
 - Hawkins (2000: 232): “the biggest single predictor of relative orderings (...) is (...) weight”
- We cannot argue in favour of:
 - Traugott (1992: 276): “in general the light-heavy distribution [end-weight] is no longer a major factor in English word order”
- Complements-first is still a significant force in VPs:
 - evidence from other complements (all types of objects)

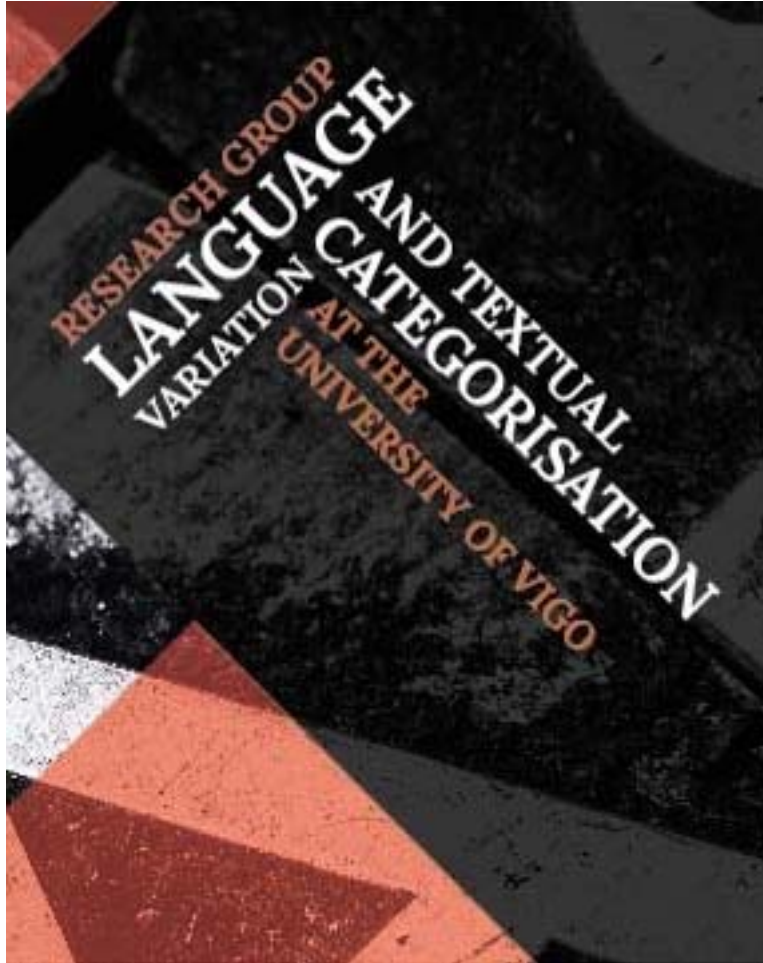
Conclusions

- Complements-first is more influential in VPs (than in APs) than in NPs=> connection type of Head / complements-first (the more verbal the head is, the more likely the structure of the phrase is governed by specifically the syntactic principle of complements-first).

Conclusions

- VERBS ARE MORE PROTOTYPICAL HEADS THAN NOUNS
 - frequency: fewer intransitive Vs (23.29% in PPCMBE) than intransitive Ns (56.04%)
 - paradigmatic versatility: wider with Vs (complementation options: monotransitive, intensive, ditransitive, complex-transitive, transitive-adverbial)
 - ellipsis: 4.09% of verbless VPs vs. 52.98% nounless NPs
 - morphological choices: number/person/tense/aspect in V; morphology contributes to syntactic integration, a feature of headedness (Givón 1993: 23,26; Noonan 2007: 101)

lvtc



UniversidadeVigo

Word order
in the recent history
of English:
syntax and processing on
the move

Javier Pérez-Guerra (jperez@uvigo.es)

Thanks!