



Pairing word order with headedness in the history of English: a corpus-based analysis

Javier Pérez-Guerra (jperez@uvigo.es)

Language Variation and Textual Categorisation Research Group
University of Vigo

lvtc

ISLE 2011, Boston, June 20

2



Goal

- **Goal:** check the distributional tendencies affecting the placement of modifiers/adjuncts and complements/arguments in noun phrases, adjectives phrases and verb phrases
- **Examples:**
 - (1) Now I will deal [with the construction] [in a way which will lead to odd results].
 - (1') Now I will deal [in a way which will lead to odd results] [with the construction].
 - (2) the author [of this book] [from London]
[N.A. the author is from London]
 - (2') the author [from London] [of this book]
 - (3) keen [on music] [to a large extent]
 - (3') keen [to a large extent] [on music]

lvtc

3



Outline

- **goal**
- **background**
- description of the **corpora**: complementation and adjunction strategies
- **analysis** of the data:
 - Complements-first
 - End-weight
- **concluding remarks**

lvtc




Background

- **Complements**
 - semantically selected or subcategorised
 - 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)
 - exclusion when the pattern is saturated
 - syntactic dependencies:
 - 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

lvtc

4




Background


- **Distribution of complements and adjuncts:**
 - **syntactic explanation:**
 - Quirk et al. (1985: 49-50): ‘Complements first’
 - Hawkins (2007): ‘Arguments precede X’
 - **processing explanation:**
 - Quirk et al. (1985: 1398): End-weight
 - Hawkins’ (2004) ‘Minimize Domains’ or MiD:

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 2000: 234)

Hawkins (2007) hypothesises that MiD is relevant especially to examples of complementation.



5



The corpora


- **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, Ann, Anthony Warner, Susan Pintzuk and Frank Beths (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, Anthony and Ann Taylor (2000) *Penn-Helsinki Parsed Corpus of Middle English, second edition*.
- **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, Anthony, Beatrice Santorini and Lauren Delfs (2004) *Penn-Helsinki Parsed Corpus of Early Modern English*.
- **(Late) Modern English:** 948,895 words (1700–1914)

Kroch, Anthony, Beatrice Santorini and Ariel Dierani (2010) *Penn Parsed Corpus of Modern British English*.




7




Background

- **Examples:**
 - (1) Now I will deal [with the construction] [in a way which will lead to odd results].
 - (1') Now I will deal [in a way which will lead to odd results] [with the construction].

(1) is claimed to be a better performance solution, on processing grounds (MiD, end-weight), than (1') because of the amount of structure which has to be processed (between the head category and the second constituent in the (local) phrase).




6




The corpora

- periods: OE, ME, EModE and (L)ModE
- parsed corpora, with (almost) identical similar parsing conventions
- parsed files (.psd), using P&P-based part-of-speech and syntactic tags
- retrieval by means of CorpusSearch
- (extensive) manual revision

./..



8



but, if you approve of this, if you please to lett 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=)

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

(.,.)

(NP-SBJ (PRO I))

(MD will)

(VB tell)


(NP-OB1 (PRO it))

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

(.,.)

lvtc

9



The data

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

PP + 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,IV.22)


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

ADV + 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)

NP-ADV + OBJ:
sends *this way one*, That way another, (MIDDLET-E2-H,21.513)

lvtc

11




The data

Examples of queries

- Verb-governed contexts: V + adjunct + complement**
verb group immediately precedes an adjunct, and the adjunct immediately precedes an **object**
OE:
node: IP*
query: (VB* iprecedes
W* | NUMP* | QP* | PP* | ADJ* | ADV* | INTJP | XP |
NP-ADT* | NP-DIR* | NP-EXT* | NP-LOC* | NP-TMP* |
NP-VOC*)
AND (W* | NUMP* | QP* | PP* | ADJ* | ADV* | INTJP | XP |
NP-ADT* | NP-DIR* | NP-EXT* | NP-LOC* | NP-TMP* |
NP-VOC* iprecedes NP-ACC* | NP-DAT*)
- ME, EModE and ModE:**
node: IP*
query: (VB* iprecedes W* | QP | PP | RRC | ADJ* | ADV* |
CP-* | IP-SUB | NP-*)
AND (W* | QP | PP | RRC | ADJ* | ADV* | CP-* | IP-SUB |
NP-* iprecedes NP-OB*)

lvtc

10




The data

- Verb-governed contexts: V + complement + adjunct**
verb group immediately precedes an **object**, and the **object** immediately precedes an adjunct
OE:
node: IP*
query: (VB* iprecedes NP-ACC* | NP-DAT*)
AND (NP-ACC* | NP-DAT* iprecedes
W* | NUMP* | QP* | PP* | ADJ* | ADV* | INTJP | XP |
NP-ADT* | NP-DIR* | NP-EXT* | NP-LOC* |
NP-TMP* | NP-VOC*)
- ME, EModE and ModE:**
node: IP*
query: (VB* iprecedes NP-OB*)
AND (NP-OB* iprecedes
W* | QP | PP | RRC | ADJ* | ADV* | CP-* | IP-SUB | NP-*)

lvtc

12



The data

OBJ + QP
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 + PP
I thoughte I wolde take *some spendyng money wyth me (MERRY TAL-E1-H,31.148)*

OBJ + RRC
If you have provided *those Stockings before spoken of* I pray send them. (STRYPE-E3-H,183.69)

OBJ + ADJ
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)

lvtc

13



The data


- Adjective-/Noun-governed contexts:**
A/N + adjunct + complement
head (adjective/noun) immediately precedes an adjunct, and the adjunct immediately precedes a complement (*that-* or *infinitive clause*)

EModE and ModE:
node: NP*
query: (NP* iDoms CP-THT)
AND (NP* iDoms *P*)
AND (*P* iprecedes CP-THT)

node: NP*
query: (NP* iDoms IP-INF)
AND (NP* iDoms *P*)
AND (*P* iprecedes IP-INF)

lvtc

15



The data


OBJ + ADV
But my Brother understood *the matter aright* (HOXINDEN-1660-E3-H,280.162)

OBJ + IP-SUB
Contrariwise, there be others againe, that will not endure to read *a short composition, bee it neuer so well approoued:* (CLOWES-E2-H,34.315)

OBJ + NP-ADV
I 'll ply *him that way,* (FARQUHAR-E3-H,9.326)

lvtc

14




The data

PP + CP-THT/IP-INF
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)
The master shewyng us that by neglygence *of some to belay the haylers,* the mayn yerde had fawln down (MADOX-E2-P1,112.434)
hee is able *without great care, study and much diligence, to labour commendably,* (CLOWES-E2-P1,8.69)

RC + CP-THT/IP-INF
[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)
Ond sona fylgde +t+at weorc, +t+at him
And immediately carried out the(neu.) work(neu.) rel(neu) him
gewunelic w+as, +t+at he godspellice lare l+arde;
accustomed was that he evangelical teaching taught
in mind of the great Obligation *that lies on them to live sutably to their Profession:* (BURNETROC-E3-P2,122.170)

lvtc

16




The data


adjunct + CP-THT

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)

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



17



The data

IP-INF + *than/that(so...that)*

for I hadde levyr *hafe bene in that payne to domysdaye*
for I had better have been in the pain to Doomsday
than hafe comen to hevene (CMJULNOR,55.163)
than have come to heaven


euerie man is more readie *to argue, than to obey and execute.*
(BACON-E2-P1,1,7R.73)

but the Bassa here was in such haste *to see us, that he dispatched the courier back which Mr. W- had sent to know the time he would send the convoy to meet us without suffering him to pull off his boots.* (MONTAGU-1718,86.111)


CP-THT + *than*

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)

& beon geornran *+t+at we Godes bebodu healdan, +tonne we*
& be more eager *that we God's command keep than we*
urne teonan gewrecan. (coblick,HomS_10_[BiHom_3]:33.123.444)
our harm avenge



19




The data


- Adjective-/Noun-governed contexts:**
A/N + complement + adjunct
head (adjective/noun) immediately precedes a complement (*that- or infinitive clause*), and the complement immediately precedes an adjunct

EModE and ModE:
node: NP*
query: (NP* iDoms CP-THT)
AND (NP* iDoms *P*)
AND (CP-THT iprecedes *P*)

node: NP*
query: (NP* iDoms IP-INF)
AND (NP* iDoms *P*)
AND (IP-INF iprecedes *P*)



18



The data


CP-THT/IP-INF + *sentential relative clause*

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)


[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)

CP-THT + *apposition*

he also expressed an opinion *that mulattoes inherited the vices of both races - a maxim which I had heard often enough before,*
(READE-1863,225.616)



20




Analysis of the data

(i)

Syntactic principle 'Complements-first'

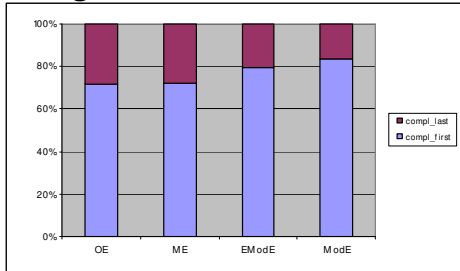
lvtc

21



Analysis of the data


- Verb-governed contexts:



- 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>ModE: yes ($P < .0001$)

lvtc

23



Analysis of the data


- Verb-governed contexts:

(based on Pérez-Guerra and Martínez-Insua 2010a,b)

	OE	ME	EModE	ModE
compl_first	8270	17803	7859	13084
compl_last	3290	6790	1995	2579

lvtc

22



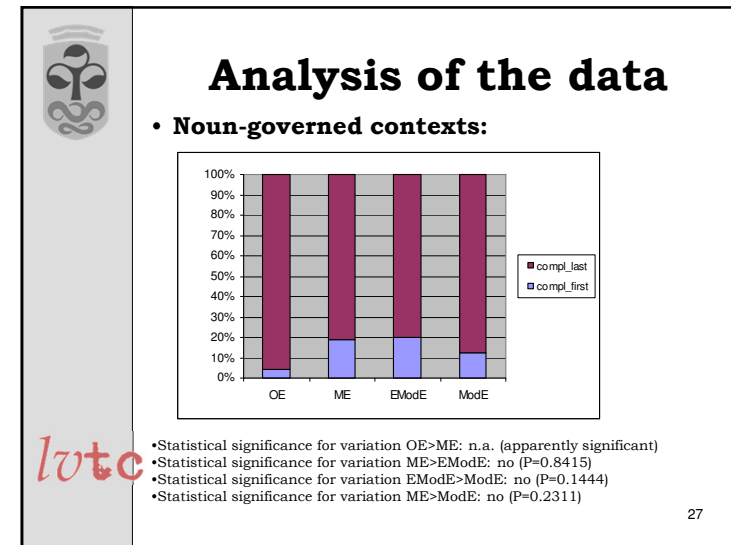
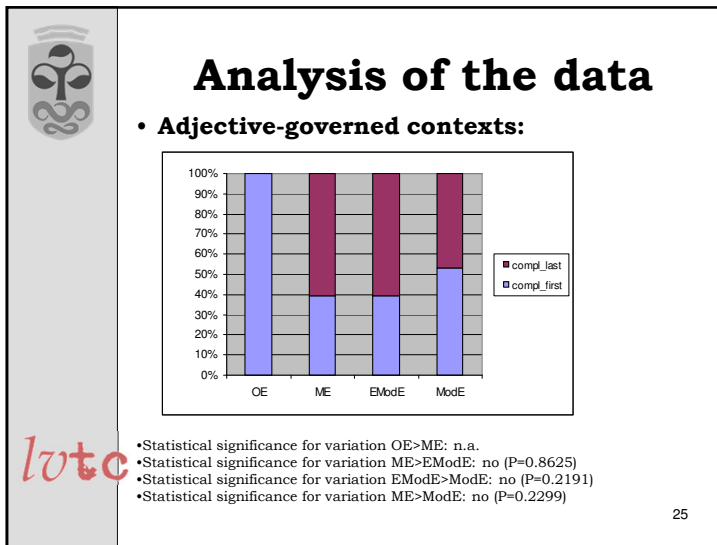
Analysis of the data

- Adjective-governed contexts:

	OE	ME	EModE	ModE
compl_first	10	20	19	32
compl_last	0	31	29	28

lvtc

24

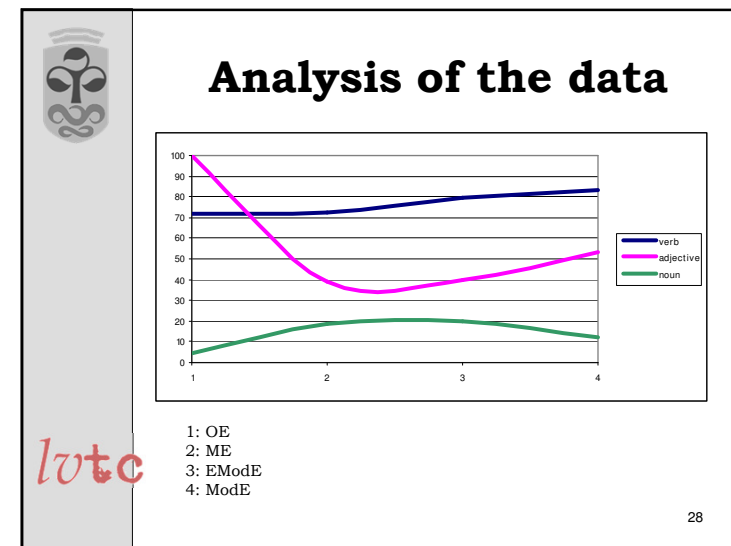



Analysis of the data

- Noun-governed contexts:

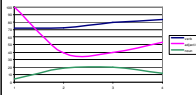
	OE	ME	EModE	ModE
compl first	1	20	26	15
compl last	22	86	105	107

26







Analysis of the data




- **VPs:**
 - most are **complement-first**
 - ME>ModE: (statistically significant) **increase of complement-first** VPs
 - pivotal period: **ME>EModE**
- **APs:**
 - OE: all the examples are complement-first (few data)
 - after ME: **half** are **complement-first**
 - pivotal period: **ME**
- **NPs:**
 - most are **complement-last** (ME>ModE variation is not significant)
 - pivotal period: **ME**



29



Analysis of the data

(ii) Processing principle 'End-weight'



31




Analysis of the data

So...

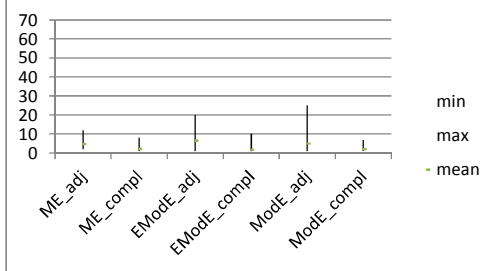
- Heads and compliance with Complements-first:
 - VPs clearly comply with Complements-first.
 - NPs do not comply with Complements-first.
 - APs occupy an intermediate position since half comply with Complements-first.
 - **connection head status ↔ Complements-first**
- Chronology of variation:
 - VPs: increase of complement-first from ME onwards
 - NPs: complement-last in all periods; quantitative stability from ME onwards
 - APs: intermediate situation from ME onwards


30




Analysis of the data

Complement-first VPs: (no. of words, outliers excl.)



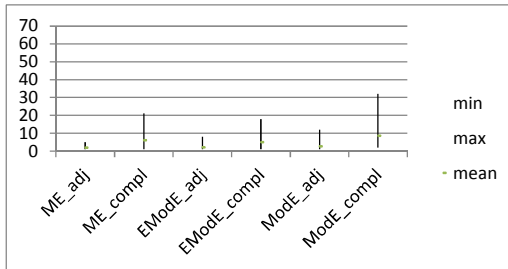
– End weight: × in all periods


32



Analysis of the data

Complement-last VPs: (no. of words, outliers excl.)



- End-weight not evident, more noticeable in ModE

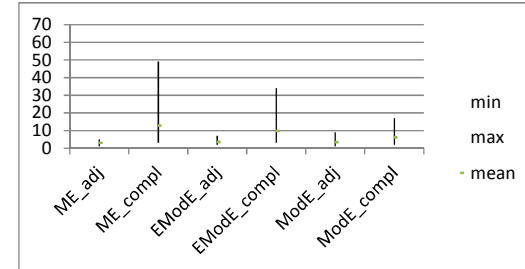
lvtc

33



Analysis of the data

Complement-last APs: (no. of words, outliers excl.)



- End-weight: ✓, less noticeable across time

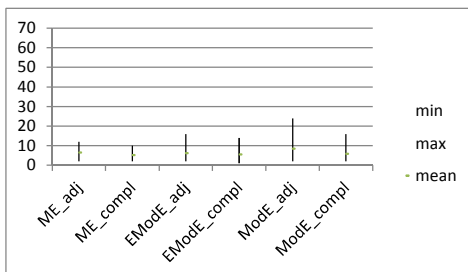
lvtc

35



Analysis of the data

Complement-first APs: (no. of words, outliers excl.)



- End-weight is not clear in the periods investigated

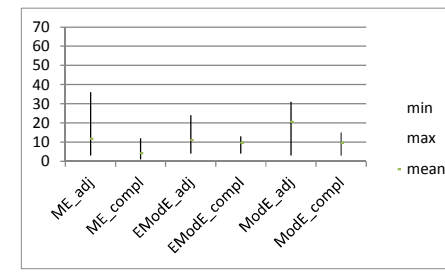
lvtc

34



Analysis of the data

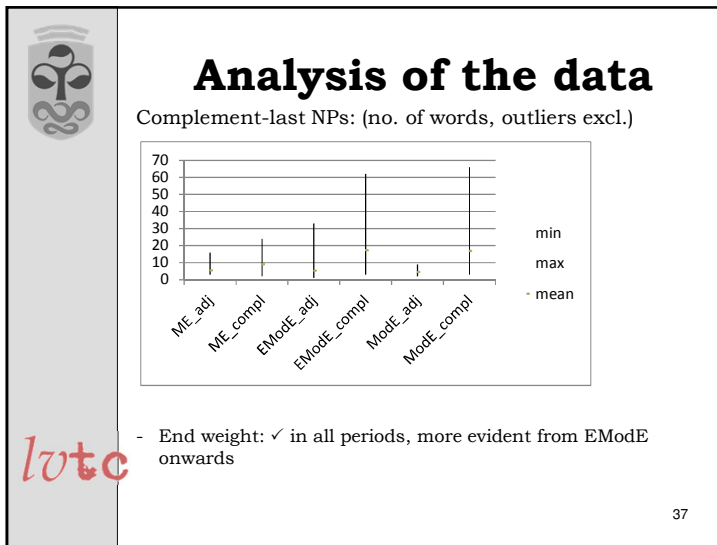
Complement-first NPs: (no. of words, outliers excl.)



- End weight: ✓ in all periods; more noticeable in ModE

lvtc

36



-
- Analysis of the data**
- So...
- general End-weight accomplishment hierarchy: $V < A < N$
 - When syntax is not the rule, ie when the construction is complement-last, End-weight is needed (clearly with N and A, and more noticeable with V).
 - **connection head status ↔ End-weight**
- lvtc
- 39

-
- Analysis of the data**
- **VPs:**
 - complement-first: do not observe End-weight
 - complement-last: do not clearly observe End-weight
 - **APs:**
 - complement-first: do not clearly observe End-weight
 - complement-last: observe End-weight (in general)
 - **NPs:**
 - complement-first: observe End-weight, and this is more evident across time
 - complement-last: observe End-weight, and this is more evident across time
- lvtc
- 38

-
- Concluding remarks**
- Competing forces: syntax (Complements-first) and processing (End-weight):
 - connection complement distribution and head category: **the more verbal the head is, the more likely its compliance with Complements-first**
 - Headedness prototypicality hierarchy: $V > A > N$:
 - frequency: the number of bare (intransitive) nouns in NPs outweighs the number of bare (intransitive) VPs
 - paradigmatic functionality: verbs take part in a wider range of complementation patterns than nouns
 - ellipsed nouns in NPs are more common than ellipsed verbs in VPs
 - morphological possibilities (Givón 1993: morphology implies syntactic integration)
 - APs: intermediate position since adjectives are more verbal than nouns
- lvtc
- 40



Concluding remarks

- Competing forces: syntax (Complements-first) and processing (End-weight):
 - connection complement distribution and head category: **the more verbal the head is, the less likely its compliance with End-weight**
 - **when syntax fails, End-weight is crucial** and justifies complement-last examples with the three categories (radically with nouns), especially in ModE
 - **when syntax is at work (Complements-first), End-weight is not responsible** for the distribution of complements and adjuncts since End-weight is not necessary

lvte

41



Concluding remarks

- **So...**

“the biggest single predictor of relative orderings (...) is syntactic weight” (Hawkins 2000: 232)

is **not** strictly correct since, according to the data, syntax (Complements-first) seems to be a bigger predictor.

[“in general the light-heavy distribution [end-weight] is no longer a major factor in English word order” (Traugott 1992: 276)]

lvte

43



Concluding remarks

- Competing forces: syntax (Complements-first) and processing (End-weight): The findings are **more noticeable from EModE onwards**.
- **word order:**
 - OE:** “word order regulated by pragmatic tendencies [authors: path “towards one determined predominantly by syntactic rules]” (Van Hoorick 1994: 53), with some trends:
 - V2 (several possibilities in the pre-V slot)
 - light-heavy distribution (Traugott 1992: 276)
 - ME:** fixation of the distribution of constituents in the clause in progress (Fischer 1992: 371).
 - EModE:** fixed (syntacticised) word order

lvte

42



References

- Fischer, Olga (1992) “Syntax”. Ed. Norman Blake *The Cambridge history of the English language. Volume II 1066-1476*, 207-408. Cambridge: Cambridge University Press.
- Givón, Talmy (1993) *English grammar II*. Amsterdam: John Benjamins.
- Greenbaum, Sidney, Gerald Nelson and Michael Weitzman (1996) “Complement clauses in English”. Eds. Jenny Thomas and Mick Short *Using corpora for language research: Studies in honour of Geoffrey Leech*. London: Longman, 76-91.
- Hawkins, John A. (1994) *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. (2000) “The relative order of prepositional phrases in English: going beyond manner-place-time”. *Language Variation and Change* 11: 231-266.
- Hawkins, John A. (2001) “Why are categories adjacent?” *Journal of Linguistics* 37:1-34
- Hawkins, John A. (2004) *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hawkins, John A. (2007) “Performance and grammatical variation in the ordering of verb, direct objects and obliques”. Delivered at DGS, Siegen.
- Matthews, P.H. (1981) *Syntax*. Cambridge: Cambridge University Press.

lvte

44



lvto

References

- Matthews, P.H. (2007) *Syntactic relations. A critical survey*. Cambridge: Cambridge University Press.
- Pérez-Guerra, Javier and Ana E. Martínez-Insua (2010a) "Diachrony and word order hand in hand: on complementation/adjunction performance solutions in English". Delivered at ICAME 31, Giessen.
- Pérez-Guerra, Javier and Ana E. Martínez-Insua (2010b) "Word order and performance optimality get on well in the history of English: on complement+adjunct and adjunct+complement solutions". Delivered at SLE 43, Vilnius.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985) *A comprehensive grammar of the English language*. London: Longman.
- Rohdenburg, Günther (1996) "Cognitive complexity and grammatical explicitness in English". *Cognitive Linguistics* 7: 149-182.
- Traugott, Elizabeth Closs (1992) "Syntax". Ed. Richard M. Hogg *The Cambridge history of the English language. Volume I: The beginnings to 1066*. Cambridge: Cambridge University Press, 168-289.
- Van Hoorick, Bart (1994) *Pragmatic positions and the history of English word order. Working Papers in Functional Grammar* 56. Amsterdam: University of Amsterdam.