Forces that Shape Post-Auxiliary Ellipsis in Late Modern English

Evelyn Gandón Chapela

evelyn.gandon@uvigo.es

Journée d'études sur l'ellipse

Université Paris Diderot-Paris 7

July 9, 2015

Universida de Vigo

Outline

1. **Post-Auxiliary Ellipsis**
2. Goals
3. Methodology
4. Data analysis
5. Conclusions

1. Post-Auxiliary Ellipsis (1)

Post-Auxiliary Ellipsis (Sag 1976; Warner 1993; Miller 2011; Miller & Pullum 2013).

Missing material after an auxiliary

Ellipsis of a VP, PP, DP or AP after:

modal auxiliaries, auxiliaries *be*, *have* and *do* and infinitival marker *to* (defective non-finite auxiliary verb; Pullum 1982; Gazdar 1985; Levine 2012; Miller & Pullum 2013).

1. Post-Auxiliary Ellipsis (2)

1.1. Two Subtypes:

-VP Ellipsis:

(1) I have written a squib but he hasn't [~~written a squib~~]VP.

(2) John is tall but Sara is not [~~tall~~] AP

*(3) John is a doctor and Anne is [~~a doctor~~]DP **too.***

(4) Bill's son is on the beach, although he shouldn't be [~~on the beach~~]PP because he's allergic to the sun.

-Pseudogapping:

(5) John invited Sarah, and Mary did ~~invite~~ Jane.

(Adapted from Gengel 2013: 23)

1. Post-Auxiliary Ellipsis (3)

1.2. VP Ellipsis (VPE)

(6) *Will likes coke and I think Karen does ~~like coke~~ too.*

→ Ellipsis of a VP

→ Possible in subordination

→ It applies across sentence boundaries

(7) *A: Can you pass me the salt?*

   B: Yes, of course I can ~~pass you the salt~~.

(8) *A: Did you phone Alice?*

   B: No I didn't ~~phone~~. Sorry, I forgot.

1. Post-Auxiliary Ellipsis (5)

1.4. Pseudogapping

(9) *Does that make you mad? It* **would** ~~*make*~~ *me* ~~*mad*~~*!*

(adapted from Levin 1986: vii)

→ Very similar to VPE, but in this case, a remnant is left.

→ More constrained acceptability → preference for coreferential subjects and comparative contexts.

(10) *He realized that he could make more money in some other position than he* **could** ~~*make money*~~ *farming.*

(adapted from Levin 1978:229)

(11) *Some brought roses and others* **did** ~~*bring*~~ *lilies.*

(adapted from Gengel 2007)

Outline

1. Post-Auxiliary Ellipsis

2. Goals

3. Methodology

4. Doubts and Queries

5. To-do list

2. Goals (1)

2.1. Other corpus studies on PAE:

Hardt 1993, Hardt & Rambow 2001; Nielsen 2003, Nielsen 2005; Hendriks & Spenader 2005;   Bos & Spenader 2011; Miller 2011; Miller & Pullum 2013; Miller 2014 (for Present Day English).

2.2. Aim of my PhD dissertation:

Empirical study of Post-Auxiliary Ellipsis (both quantitative and qualitative) from Late Modern English to Present-Day English based on the Penn Corpora of Historical English (1700-1914).

2. Goals (2)

2.3. Focus of my PhD Dissertation:

Syntactic variation, genre distribution and discourse variables (type of anaphora, mismatches in polarity, aspect, voice, modality, tense; comparison of clause types; distance, linking, type of focus; etc).

Outline

1. Post-Auxiliary Ellipsis
2. Goals
3. **Methodology**
4. Data analysis
5. Conclusions

3. Methodology (1)

-Corpus-based methodology: manual and automatic analysis of texts.

-Manual analysis of **12** raw texts out of 102 files (12.24% =**112,347 words** analysed out of 948,895 words) all belonging to different genres and periods of time of the PPCMBE from the Penn Corpora.

-Automatic analysis with the programme CorpusSearch 2.

3. Methodology (2)

-The patterns of PAE were examined in order to draw some generalizations:

(12) *He did.*

(IP-SUB (NP-OB1 *T*-1)

(NP-SBJ (PRO he))

(DOD did)

(**VB ***)))))

(. .))

-The tag *VB** indicates that the verb has been elided. Simply searching for the tag VB indicating that it immediately dominates (iDoms) the * symbol returns a fair amount of VPE instances.

3. Methodology (3)

CorpusSearch algorithm

```
node:    *
query:  (VB* iDoms  \*)
OR  (HV* iDoms  \*)
OR  (MD* hasSister  !VB*|BE*|DO*|HV*)
OR  ((MD* iPrecedes  HV*)
AND  (HV* iPrecedes  [.,]))
OR ((MD* iPrecedes NEG)
AND (NEG iPrecedes  HV*)
AND  (HV* iPrecedes  [.,]))
OR  ((MD* iPrecedes  HV*)
AND  (HV* iPrecedes  BE*)
AND  (BE* iPrecedes  [.,]))
OR ((MD* iPrecedes  NEG)
AND  (NEG iPrecedes  HV*)
AND  (HV* iPrecedes  BE*)
AND (BE* iPrecedes [.,]))
OR  (BE* iPrecedes  [.,])
OR  ((BE* iPrecedes  NEG)
AND  (NEG iPrecedes  [.,]))
OR  (HV* iPrecedes  [.,])
OR  ((HV* iPrecedes  NEG)
AND  (NEG iPrecedes  [.,]))
OR  ((HV* iPrecedes  NP-SBJ)
AND  (NP-SBJ iPrecedes  [?]))
OR  ((DO* iPrecedes  NEG)
AND  (NEG iPrecedes  NP-SBJ)
AND  (NP-SBJ iPrecedes  [.,?]))
OR  (DOI iPrecedes  [.,])
OR  (CP* hasLabel  CP-QUE-TAG*)
```

3. Methodology (4)

Penn Corpora of Historical English

→Text samples of British English prose across its history, from the earliest Middle English documents up to the First World War.

→The texts come in three forms: **raw text**, **part-of-speech tagged text** and **syntactically annotated text**.

→The syntactic annotation (parsing) **permits searching** not only for words and word sequences, but also for **syntactic structure**.

3. Methodology (5)

Genres (18)

-Bible

-Biography,autobiography

-Biography, other

-Diary

-Drama, comedy

-Educational treatise

-Fiction

-Handbook, other

-History

-Law

-Letters, non-private

-Letters, private

-Philosophy

- Trial proceedings

-Science, medicine

-Science, other

-Sermon

-Travelogue

Todas las tablas

Post-Auxiliary Ellipsis

Field	Value	Field	Value
Id	ANON-1711.11.105	Polarity source	Positive
Period	1700	Polarity target	Positive
Genre	Educational treatise	Voice source	active
Example	we cou'd not fail of learning the Latin Language, as well as we do the Modern Languages.	Voice target	active
Type	Pseudogapping	Aspect source	non-perfective_non_progressive
Licensor	do	Aspect target	non-perfective_non_progressive
Aux Before Licensor	none	Modality source	None
Linking	Subord_comparative	Modality target	None
Connector	as as	Tense source	-ing
Distance	0	Tense target	Present
Turn	☐	Tag question	☐
Domain	IP-SUB	Clause type source	Declarative
Type of anaphora	anaphoric	Clause type target	Declarative
Category source	VP	Focus	Obj-aux-choice
Category target	VP	Remnants:	NP
Vehicle change	☐		
Split Antecedents	☐	Comments	Polarity issues

Outline

1. Post-Auxiliary Ellipsis

2. Goals

3. Methodology

4. Data analysis

5. Conclusions

4. Data analysis (1)

1. Outline

I.Core defining variables

I.1. **Grammatical variables:**

    (i) Licensors

    (ii) Auxiliary before licensor

    (iii) Syntactic linking

    (iv) Domain

    (v) Category of source vs. Category of target of ellipsis

    (vi) Split antecedents

    (vii) Remnants Pseudogapping (category & function)

    (viii) Auxiliary-related variables: polarity, voice, aspect, modality and tense mismatches.

4. Data analysis (2)

1. Outline

I.2. Semantic, discursive variables:

(i) Type of clause (declarative, interrogative, imperative and tag questions)

(ii) Type of anaphora: exophoric, anaphoric and cataphoric.

(iii) Type of focus: subject-choice, auxiliary-choice, subject-auxiliary choice, object-choice, etc.

(iv) Vehicle change

(v) Turn

4. Data analysis (3)

1. Outline

II. Usage variables

II.1. Diachronic hypothesis: there is no statistically significant variation in the frequency of PAE in Late Modern English.

II.2. Textual hypothesis: there is no statistically significant variation in the frequency of PAE by text-types in Late Modern English.

4. Data analysis (4)

1. Outline

III. Processing variables:
  - distance: IP- (syntactic) and word- (lexical) distance

IV. Qualitative Analysis of VPE and Pseudogapping

V. Concluding remarks

4. Data analysis (5)

2. Licensors

Licensors	Pseudogapping		VPE		Total	
	Abs.freq.	%	Abs.freq.	%	Abs.freq.	%
Be	29	33.72	280	26.92	309	27.44
Have	13	15.12	56	5.38	69	6.13
Do	33	**38.37**	372	**35.77**	405	35.97
Modals	11	12.79	324	31.15	335	29.75
To	0	0	8	0.77	8	0.71
TOTAL	86		1040		1126	

4. Data analysis (6)

Licensors	Total	Total nr. corpus	N.f. /1,000
Be	309	37,727	8.19
Have	69	11,642	5.92
Modals	335	16,995	**19.71**
To	8	15,335	0.52

*DO: Number of main verbs in past/present tense→72,254 = 5.61
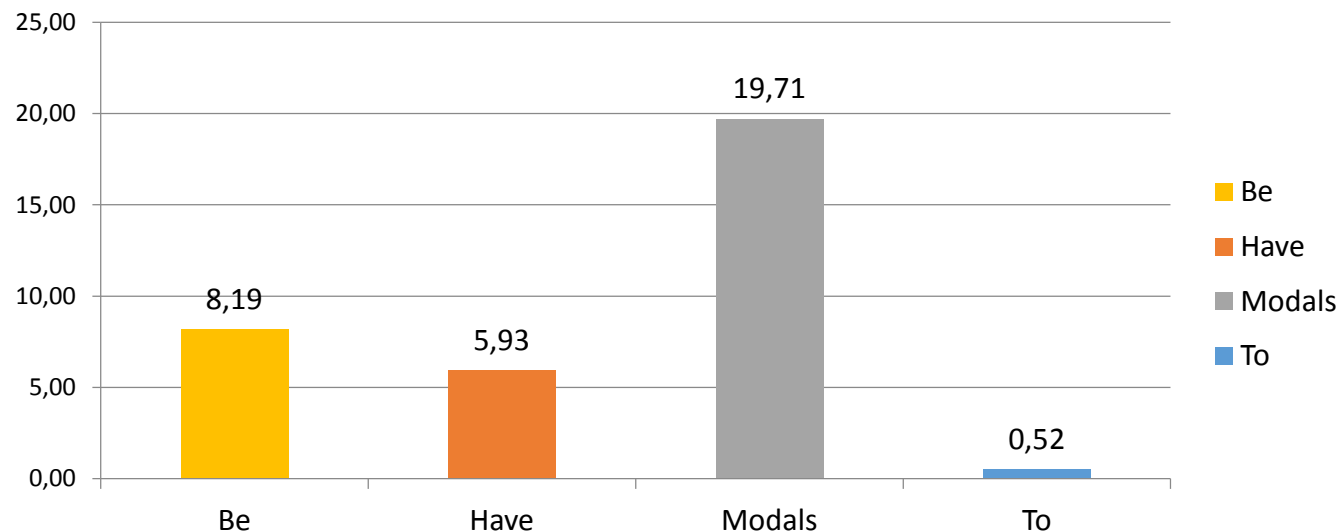
4. Data analysis (7)

3. Auxiliary before licensor

2.1. Pseudogapping

(13) *he was more cruel and treacherous to his poor Hungarian subjects, than ever the Turk **has been** to the Christians.*

(14) *Even the delicate associations with words can be expounded through our own language; just as they **must be** to the pupil who is studying the original.*

2.2. VPE

(15) *I am connected with a public newspaper, and **have been** for some years.*

(16) *but by and by it was all broke off on a sudden, as the other **had been**.*

4. Data analysis (8)

3. Auxiliary before licensor

Auxiliary	Pseudogapping	
	Abs. Freq.	%
none	79	91.86
have	3	3.49
must	2	2.33
would	2	2.33
TOTAL	**86**	

8.14% cases

4. Data analysis (9)

Auxiliary	VPE	
	Abs. Freq.	%
none	990	95.19
have	8	0.77
could	7	0.67
ought to	6	0.58
may	4	0.38
can	3	0.29
shall	3	0.29
should	3	0.29
would	3	0.29
would have	3	0.29
must	2	0.19
need	2	0.19
should have	2	0.19
could have	1	0.10
might	1	0.10
might have	1	0.10
will	1	0.10
TOTAL	1040	

4.81% cases

4. Data analysis (10)

4. Syntactic linking

Linking	Pseudogapping		VPE		Total	
	Abs.freq.	%	Abs.freq.	%	Abs.freq.	%
None	15	17.44	532	**51.15**	547	**48.58**
Subord_adverbial	37	**43.02**	193	18.56	230	20.43
Subord_comparative	29	33.72	191	18.37	220	19.54
Coordination	4	4.65	62	5.96	66	5.86
Subord_relative	1	1.16	62	5.96	63	5.60
Total	**86**		**1040**		**1126**	

4. Data analysis (11)

5. Syntactic domain

(17) **IP-MAT:** *If the Church of Rome authorizes the worship of the Virgin Mary, or of Angels, or of the Spirits of men departed; the Church of Christ **does not.***

(18) **IP-SUB:** *she fries better than she **did**, but not like Jenny.*

(19) **CP-QUE-TAG:** *I think, **don't you**? that there may be another attraction.*

(20)   **CP-QUE:** *Phipps: I don't observe an alternation in your lordship's appearance .*

*Lord Goring: **You don't**, Phipps?*

(21) **IP-IMP:** *Make haste after me, **do**, now!*

4. Data analysis (12)

5. Syntactic domain

DOMAIN	PG		VPE		Total	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
IP-SUB	72	**83.72**	656	63.08	728	**64.65**
IP-MAT	14	16.28	310	29.81	324	28.77
CP-QUE	0	0	18	1.73	18	1.60
CP-QUE-TAG	0	0	50	4.81	50	4.44
IP-IMP	0	0	6	0.58	6	0.53
TOTAL	**86**		**1040**		**1126**	

4. Data analysis (13)

6. Category of the source vs. Category of target of ellipsis

Categories	PG		VPE		Total	
	Abs. Freq.	%	Abs.freq.	%	Abs.freq.	%
VP (source)-VP (target)	62	72.09	809	77.79	871	77.35
AP (source)-AP (target)	12	13.95	97	9.33	109	9.68
NP (source)-NP (target)	9	10.47	94	9.04	103	9.15
PP (source)-PP (target)	3	3.49	32	3.08	35	3.11
AdP (source)-AdP(target)	0	0	5	0.48	5	0.44
Non-constituent (source)-non-const. (target)	0	0	2	0.19	2	0.18
Non-constituent (source)-VP (target)	0	0	1	0.10	1	0.09
TOTAL	86		1040		1126	

(22) Q. **Did you make any answer to it, or did you give it him**? A. No, I did not;

4. Data analysis (14)

I.2. Semantic, discursive variables: Turn

Turn	PG		VPE		Total	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
NO	78	**90.70**	698	**67.12**	776	**68.92**
YES	8	9.30	342	32.88	350	31.08
Total	86		1040		1126	

(23) "I did not think you would have held out against her".
" Not when I **had against you**?"

(24) Mr. Serjeant Copley. How did you read it over? A. The same as I **would any thing else**.

(25) Mabel Chiltern: Don't forget. Lord Goring: Of course I **shan't**.

4. Data analysis (15)

I.2. Semantic, discursive variables: Type of anaphora

Type of anaphora	Pseudogapping		VPE		Total	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
anaphoric	86	100	1025	98.56	1111	98.67
cataphoric	0	0	15	1.44	15	1.33
TOTAL	86		1040		1126	

(26) My lord replied, "If you **had not**, we should have done it ourselves".

(27) **Do**, send the cloaths if you send them in a wheelbarrow.

4. Data analysis (16)

II. Usage: Period

PERIOD					
	18th cent.	**N.f. 1,000 IPs**	**19th cent.**	**N.f. 1,000 IPs**	**Total**
Pseudogapping	50	0.78	36	0.49	86
VPE	391	6.11	649	8.86	1040
Total	441	63980	685	73228	1126

Diachronic hypothesis: there is no statistically significant variation in the frequency of PAE in Late Modern English. P= <.0001

4. Data analysis (17)

II. Usage: Period
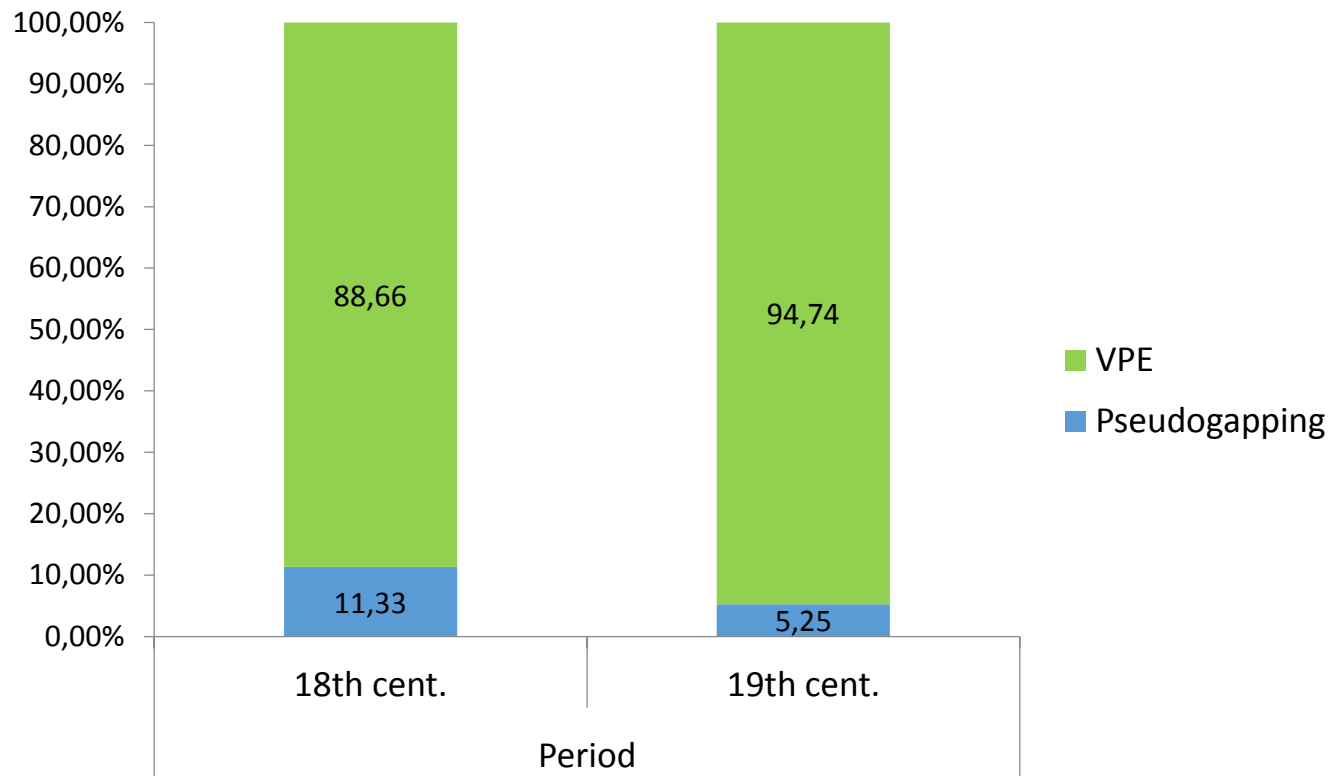
Distribution of Pseudogapping and VPE statistically significant. P=0.0003

4. Data analysis (18)

II. Usage: Genre

Genre	Pseudogapping	VPE	Number IPs	N.f. 1,000 IPs
Speech-related	49	765	59,876	13.59
Writing-related	21	140	54,266	2.96
Fiction oral	2	54	10,355	5.40
Fiction narrative	6	38	10,355	4.24
Fiction total	8	92	10,355	9.65
Philosophy	3	28	2,881	10.76

Distribution of PAE by genre

4. Data analysis (19)

II. Usage: Genre

Genre	Pseudogapping	VPE	Nr. IPs	N.f. 1,000 IPs
Speech-related	52	793	62,757	13.46
Writing-related	21	140	54,266	2.96
Total	**73**	**933**	**117,023**	**8.59**

Distribution of PAE in speech and writing-related genres

4. Data analysis (20)

II. Usage: Genre

	PG	VPE
Speech-related	0.83	12.64
Writing-related	0.39	2.58

Normalised frequency of PAE in speech and writing-related genres/ 1,000 IPs.

4. Data analysis (21)

II. Usage: Genre
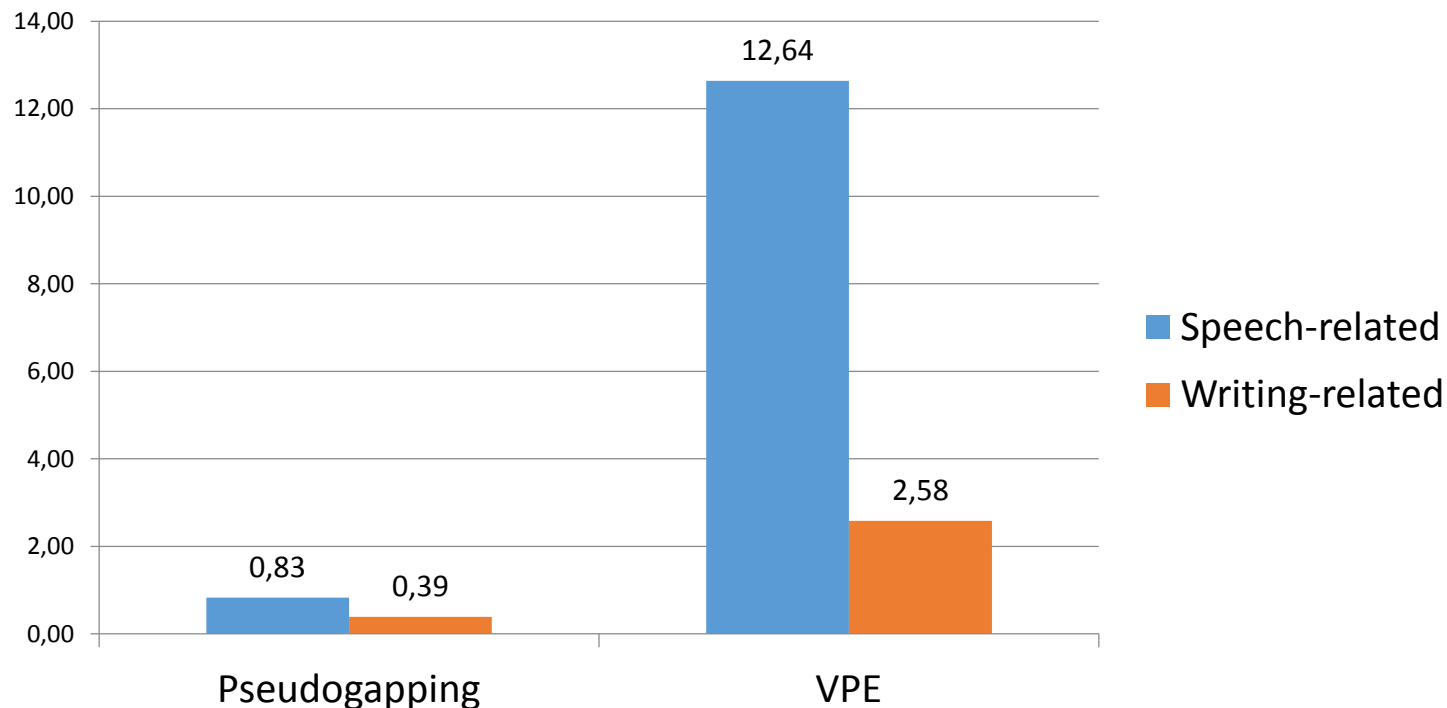
Textual hypothesis: there is no statistically significant variation in the frequency of PAE by text-types in Late Modern English. P= <.0001

4. Data analysis (22)

III. Processing Variables: Syntactic distance

Distance Nr. IPs	Pseudogapping		VPE		Total	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
0	79	91.86	804	77.31	883	78.42
1	5	5.81	179	17.21	184	16.34
2	2	2.33	38	3.65	40	3.55
3	0	0	10	0.96	10	0.89
4	0	0	4	0.38	4	0.36
>5	0	0	5	0.48	5	0.44
TOTAL	86		1040		1126	

4.Data analysis (23)

III. Processing Variables: Lexical distance

Distance Nr. words	Pseudogapping		VPE		Total	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
0-5	35	40.70	809	77.79	844	74.96
6-10	36	41.86	169	16.25	205	18.21
11-15	15	17.44	45	4.33	60	5.33
>15	0	0	17	1.63	17	1.51
TOTAL	86		1040		1126	

Outline

1. Post-Auxiliary Ellipsis
2. Goals
3. Methodology
4. Data analysis
5. **Conclusions**

5. Conclusions (1)

- Licensors: **DO** is the **most frequent licensor** of PAE. But the relative frequency of licensors shows that **modal auxiliaries** are over **three times more frequent** than the rest.

- The frequency of the **presence of auxiliaries before licensors** is slightly **higher in PG** than in VPE.

- **Adverbial and comparative subordination** are the most **frequent** types of **syntactic linking** in **PG**, whereas **adverbial subordination** or **no linking** at all predominate in **VPE**.

5. Conclusions (2)

- Over **80%** of examples of **PAE** take place within the **domain** of **subordinate** and **matrix clauses** respectively.

- **Categorial identity** between the **category** of the **source** and that of the **ellipsis site**.

- Almost **70%** of the examples of **PAE occur within the same turn**.

- Over **98%** of cases of **PAE** are **anaphoric**. **Cataphoric** ellipsis only found in **VPE**.

5. Conclusions (3)

- There is a **statistically significant variation** in the **frequency** of **PAE** in **Late Modern English**.

- **PAE** is by far **more frequent** in **speech-related genres** than in writing-related genres.
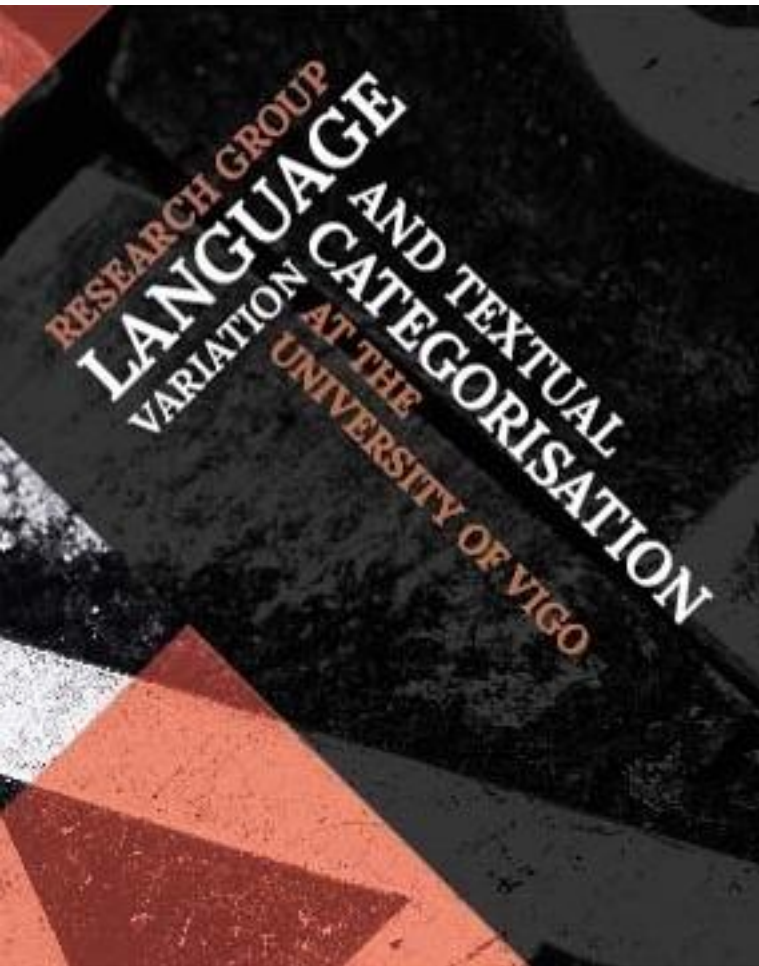
THANK YOU VERY MUCH FOR YOUR ATTENTION!

Acknowledgements

46

References

Bos, Johan and Jennifer Spenader 2011: *An annotated corpus for the analysis of VP ellipsis. Language Resources and Evaluation 45(4): 463–494.*

Carter, Ronald & McCarty, Michael. (2006). *Cambridge Grammar of English: A Comprehensive Guide. Spoken and Written English Grammar and Usage.* Cambridge: Cambridge University Press.

CorpusSearch Home. <http://corpussearch.sourceforge.net/> (Accessed 14 May 2012).

van Craenenbroeck, Jeroen and Jason Merchant 2011: 'Ellipsis phenomena'

    https://webshare.uchicago.edu/users/merchant/Public/pubs/ellipsis.cup.pdf (Accessed 14 May, 2012)

Hardt, Daniel 1993: *Verb phrase ellipsis: Form, meaning, and processing. PhD thesis,* University of Pennsylvania. (Distributed as IRCS Report 93-23.)

Hardt, Daniel and Owen Rambow 2001: 'Generation of VP Ellipsis: a Corpus-based Approach' <http://aclweb.org/anthology-new/P/P01/P01-1038.pdf> (Accessed 14 May, 2012)

Gengel, Kirsten. (2007). Focus and ellipsis. A Generative Analysis of Pseudogapping and other Elliptical Structures. PhD Dissertation, University of Stuttgart.

Gengel, Kirsten 2013: *Pseudogapping and Ellipsis*. Oxford: Oxford University Press.

Levin, Nancy 1986): *Main Verb Ellipsis in Spoken English*. New York NY: Garland.

Nielsen, Leif Arda 2003: 'A corpus-based study of verb phrase ellipsis'. *Proceedings of the 6th Annual CLUK Research Colloquium. Edinburgh. 109-115.*

Miller, Philip 2011: "The choice between verbal anaphors in discourse". In Hendrickx, I., Devi, S. L., Branco, A., and Mitkov, R., editors, *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2011*, volume 7099 of *Lecture Notes in Artificial Intelligence*, pages 82–95, Berlin. Springer.

Miller, Philip, and Geoffrey K. Pullum 2013: "Exophoric VP Ellipsis". In Philip Hofmeister and Elisabeth Norcliffe, eds., *The Core and the Periphery: Data-Driven Perspectives on Syntax Inspired by Ivan A. Sag*, 5-32. Stanford, CA: CSLI Publications.

Forces that Shape Post-Auxiliary Ellipsis in Late Modern English

Evelyn Gandón Chapela

evelyn.gandon@uvigo.es

Journée d'études sur l'ellipse

Université Paris Diderot-Paris 7

July 9, 2015