

# Data and Discourse – Three Observations, Three Case Studies

Prof. Tony McEnery



@TonyMcEnery



THE QUEEN'S  
ANNIVERSARY PRIZES  
FOR HIGHER AND FURTHER EDUCATION  
2015

# Investigating the past with corpora – what is in it for linguists

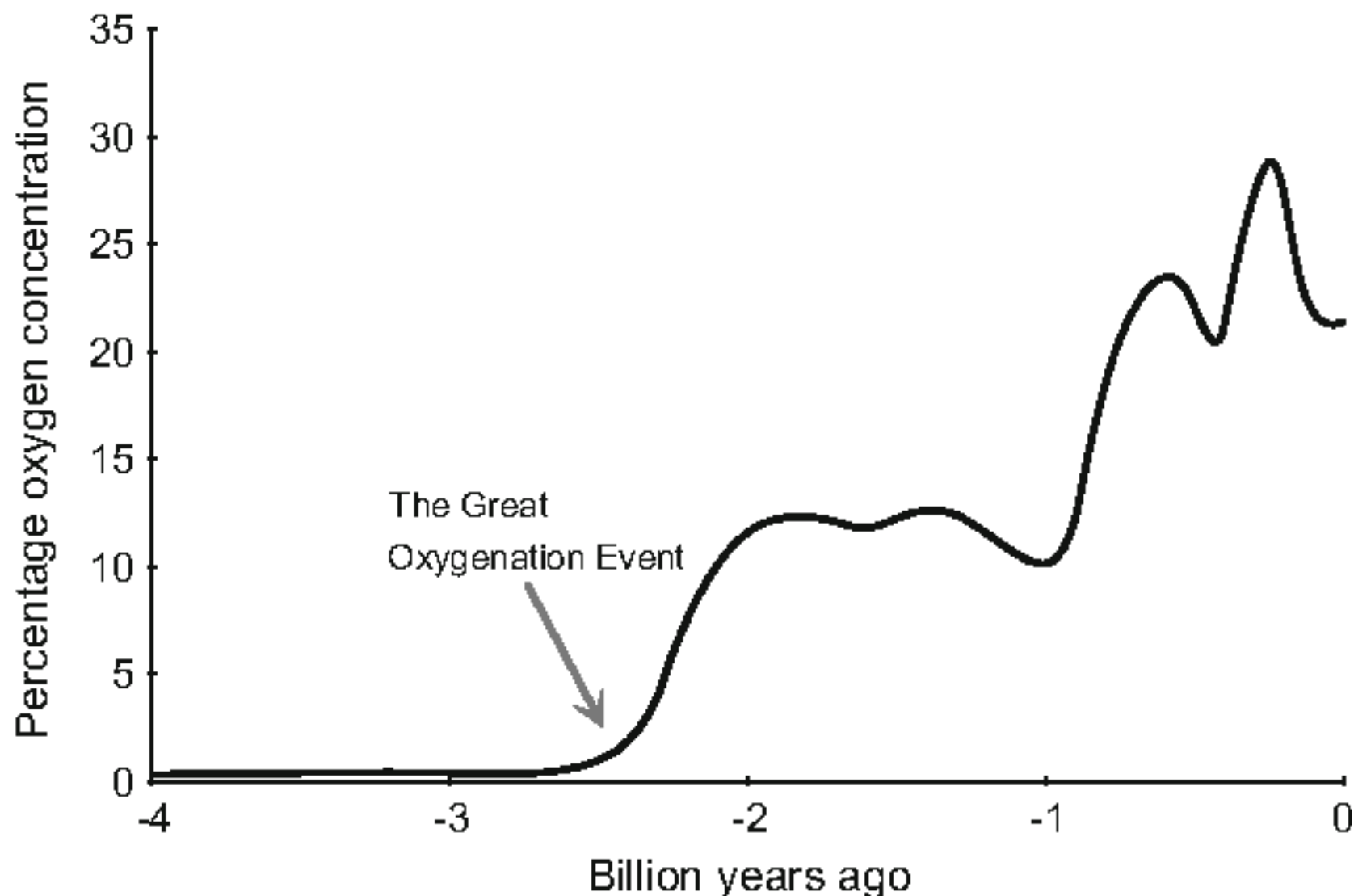
- Resources are enabling the study of discourse, especially in the past, in greater depth
- The shift from the frequent and grammatical
- Looking at change in lexis, especially content words
- Looking for evidence of suggested processes, e.g. shifts in semantic and discourse prosody
- Avoiding accidental novelty
- Methods/concepts challenged – and extended?

# Key Elements ...

- Pattern discovery
- Resource interpretation
- Discourse analysis
- Triangulation
- Close reading is key

# The Problem of Measurement - Data Points

- Static data point – measuring the same thing, in the same way, with the same meaning.
- Dynamic data point – measuring the same thing, in the same way, but with different meaning potentially.



# Google books Ngram Viewer

Graph these comma-separated phrases:  ☐ case-insensitive

between  and  from the corpus  with smoothing of

[Search lots of books](#)

[G+](#) Share

[Tweet](#)

[Embed Chart](#)



- For a better interface to Google Books see <https://googlebooks.byu.edu/>

# The Problem

- Unless you are interested in word form (and even then only in one specific wordform rather than variability) language does not provide a static datapoint – but an awful lot of research presumes it does
- Language provides a dynamic data point – meaning is changing. Grammar is changing. Pretty much everything is changing.
- Meaning change can and does occur – as this talk will show.
- Spelling also changes – but how good are our spelling standardizers?
- The conceptual map of the world also changes.
- Also, to search you need to know what you are looking for – how good is our lexicography for any given period? How reliable are current claims of word meaning in the past?



# Case study one – Availability.

## The great Mahometan Berry panic

- Researchers have written a great deal about the emergence of the coffee house in post-Restoration London and the accompanying fears that the *Mahometan berry* (the coffee bean) threatened Christianity itself

- “The coffee bean was called ‘The Mahometan Berry’” Suter, K. (2005) ‘The Rise and Fall of English Coffee Houses’ *Contemporary Review*, 286 (1669) p 107-110.
- “The introduction of coffee from Istanbul in the 1650s generated a vast literature about this “Mahometan berry”  
(<http://www.plainislam.com/in-depth/history-of-muslims-in-britain-from-queen-elizabeth-i-to-george-i/>)
- “Known as the ‘Mahometan berry’ the debates around coffee touched on a range of issues’ Lewis, P. (2003) ‘Christians and Muslims in the West: From isolation to shared citizenship’, *International Journal for the Study of the Christian Church*, 3 (2), pp 77-100)
- “Every English man, woman and child, whether interested in Islam or not, could not avoid hearing it reviled and appropriated into current controversies: when coffee was introduced to English drinkers in the second half of the seventeenth century, the ale sellers, an interest group that negatively affected the new beverage, denounced the ‘Mahometan Berry’, accusing it of being Islam’s secret weapon for the subversion of pious English customers’, Matar, B. (2009) ‘Britons and Muslims in the early modern period: from prejudice to (a theory of) toleration’, *Patterns of Prejudice*, 43 (3-4), pp 213-231.

- That food altered behaviour – even converted the consumer into a religious other – is perhaps no more explicitly found than in England's reception to the "Mahometan berry": coffee. While coffee was certainly a welcome new drink to some in early modern England (the first coffee house opened in 1652), others denounced it.<sup>11</sup> In 1605, Francis Bacon warned that coffee can "disturb the mind" as it functions like opiates (272). Other writers believed coffee threatened English religious culture. As Matar has argued, "as far as they were concerned, coffee ... was a 'Mahometan gruel' drunk by potential renegades from Christianity...For many writers, coffee-drinking was dangerous because it prepared Englishmen for apostasy to Islam" Ebrahim, F. (2015) *Turning to Food: Religious Contact and Conversion in Early Modern Drama*, PhD Thesis, University of Western Ontario.
- "Another worry in the minds of the English was that frequent contact with the 'unholy' Muslims would lead to spiritual contamination; certain writers even denounced the drinking of coffee (Matar, 1998: 111-114). Coffee was known as the "Mahometan berry" (Ibid: 111) and it was "dangerous because it prepared Englishmen for apostasy to Islam" (Ibid: 112)." Shah, N. (2015) *At sea, in text, and on stage: Islam and Muslims in Early Modern English drama*, M. Res Thesis, University of Birmingham.
- Matar, N. *Turks, Moors and Englishmen in the Age of Discovery*, Columbia University Press, New York.

# What do we find in 1 billion words?

- Mahometan berry – 0
- Turkish coffee – 3
- Arabic coffee – 1
- Coffee – 1,079 mentions
- Collocates focus on other drinks (e.g. *brandy*, *cider*, *tea*), other intoxicants (e.g. *tobacco*), where it is drunk (e.g. *coffee-houses*) and its form (e.g. *berry*). No religion collocates. Turkish collocates 8 times, ARAB not at all.
- Oh – and *gruel* is never, ever *mahometan*

No	Filename	Solution 1 to 13	Page 1 / 1
1	<a href="#">A36763</a>	this following commendation . The most excellent Virtues of the <b>Berry</b> called	<a href="#">Coffee</a> . COffee is a Berry which only grows in the desert of
2	<a href="#">A36763</a>	this following commendation . The most excellent Virtues of the Berry called	<a href="#">Coffee</a> . COffee is a <b>Berry</b> which only grows in the desert of
3	<a href="#">A36763</a>	commendation . The most excellent Virtues of the <b>Berry</b> called Coffee .	<a href="#">COffee</a> is a Berry which only grows in the desert of Arabia ,
4	<a href="#">A36763</a>	commendation . The most excellent Virtues of the Berry called Coffee .	<a href="#">COffee</a> is a <b>Berry</b> which only grows in the desert of Arabia ,
5	<a href="#">A42100</a>	Stalk . And so in the rest . TAB . LXXVII .	<a href="#">Coffee</a> <b>Berry</b> stone . 1 , The belly of the Stone . 2
6	<a href="#">A44230</a>	Be , the Myrtle , the Indian Date Plumb , and the	<a href="#">Coffee</a> <b>Berry</b> Tree , all smooth leaves . The Chesut is long and
7	<a href="#">A44230</a>	on foot-stalks in long clusters , and a winged leaf . The	<a href="#">Coffee</a> Bean , or <b>Berry</b> , is much like it . The Coffee-berry
8	<a href="#">A51275</a>	Country about Mecca produced abundance of that sort of <b>Berry</b> of which	<a href="#">Coffee</a> is made . Kufa , or Kalufa the Holy City , called
9	<a href="#">A56639</a>	Humane shape , where a Liquor made of an Arabian <b>Berry</b> called	<a href="#">Coffee</a> is drunk . Six or seven years ago was it first brought
10	<a href="#">A80707</a>	you make us merry . With the rare virtues , of the	<a href="#">Coffee</a> <b>Berry</b> ; But shall grow jealous of your Muse , and hate
11	<a href="#">A87472</a>	contrary course . And of themselves they sue out a Divorce .	<a href="#">Coffee</a> so brown as <b>berry</b> does appear , Too swarthy for a Nymph
12	<a href="#">B01792</a>	contrary course . And of themselves they sue out a Divorce .	<a href="#">Coffee</a> so brown as <b>berry</b> does appear , Too swarthy for a Nymph
13	<a href="#">B06461</a>	in England , by Pasqua Rosee . THE Grain or <b>Berry</b> called	<a href="#">Coffee</a> , grows upon little Trees , only in the Deserts of Arabia

- There is no evidence of the drink being linked to religion at all. The onus is on those who claim that there is to show the texts that they studied to produce the claim that what we may call ‘the great Mahometan berry panic’ ever occurred.

McEnery, T. and Baker, H. (2017) ‘A corpus-based investigation into English representations of Turks and Ottomans in the early modern period’, in M. Pace-Sigge and K. Patterson (eds) Lexical priming: Applications and Advances, John Benjamins, Amsterdam, pp 42-66.

# Case Study 2 – Dynamism in meaning.

## What did the word mean to the audience and the author?

- Diachronic
- He hath given his empire up to a whore (Shakespeare)
- Kisse the whore on her arse (Johnson)
- I co'ud never keep a whore to myself (Wycherley)

# OED

- **a.** A woman who prostitutes herself for hire; a prostitute, harlot.
- **b.** More generally: An unchaste or lewd woman; a fornicatress or adulteress. ***to play the whore*** (of a woman), to commit fornication or adultery.
- **c.** A male prostitute; any promiscuous or unprincipled person. (Esp. as a term of abuse.)
- *fig.; spec.* in biblical use, applied to a corrupt or idolatrous community (cf. [WHOREDOM n. 2](#)), and hence in controversial use, esp. in phr. ***the whore of Babylon***, to the Church of Rome (in allusion to Rev. xvii. 1, 5, etc.).
- Is the ordering of this dictionary entry right for the 17<sup>th</sup> century? Does it change across the century?

<b>All three decades</b>	criminality, gender identity, insult, near synonym, purchase, relative, religious, undesirable characteristic
<b>00 only</b>	location
<b>30 and 70</b>	age, acting, controllers, honesty, title
<b>70 only</b>	disease, illegitimacy, pity



# Definitions - 00

- Whore – noun a: A woman who sells sexual favours. She is associated with criminality, a range of undesirable character traits, e.g. impudence and slum living. To have a whore as a relative was viewed as a problem.
- Whore – noun b: a metaphorical usage, drawing on the underlying meaning of *whore* used to insult women.
- Whore – noun c: the Whore of Babylon provides a complex allegory which can be drawn upon either to supplement or to amplify usage b. Collocates indicating this meaning of *whore* include, *cup, fornication, beast* and *kings*.

# Definitions - 30

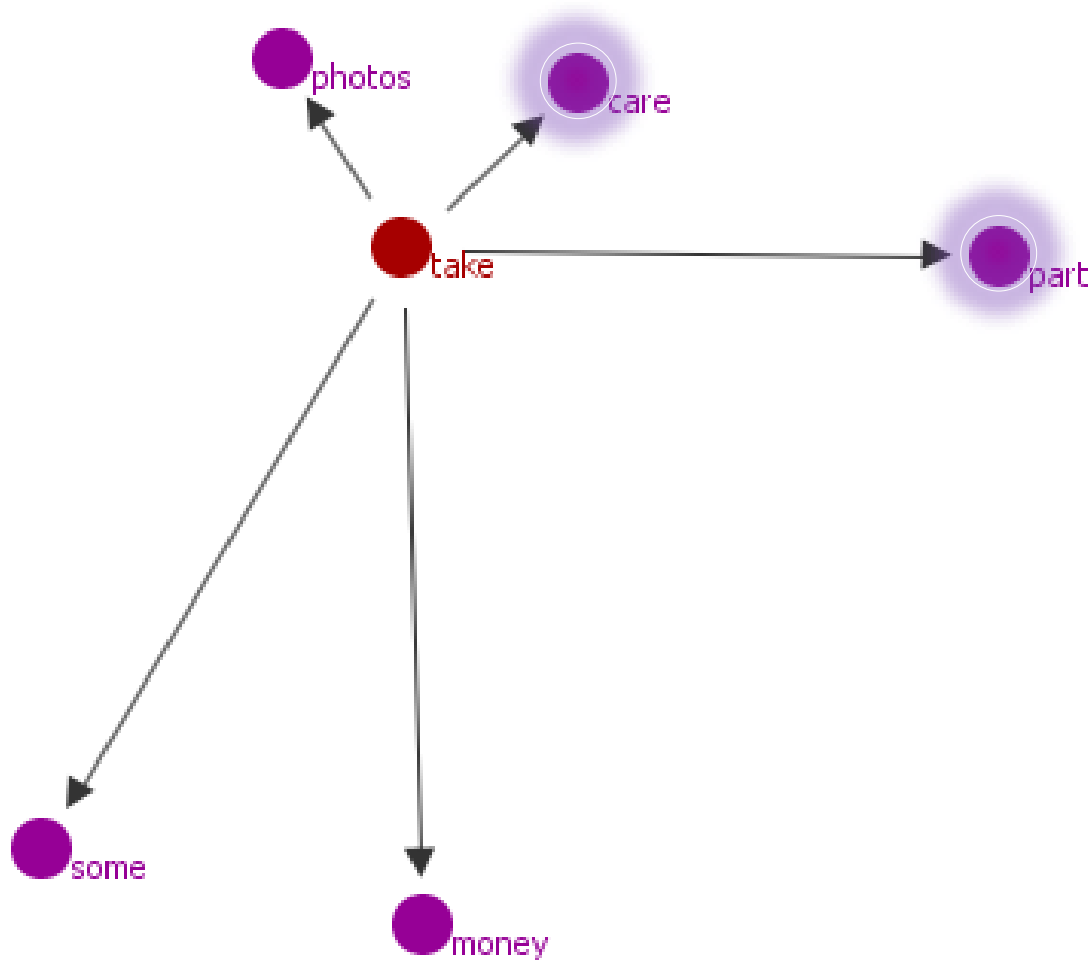
- Whore – noun a: A woman who **chooses to sell** sexual favours, **often in association with a bawd**. She is **dishonest and** associated with criminality and a range of undesirable character traits, e.g. impudence. To have a whore as a relative was viewed as a problem. **Nonetheless, people of quality associated with whores.**
- Whore – noun b: a metaphorical usage, drawing on the underlying meaning of *whore* used to insult women.
- Whore – noun c: the Whore of Babylon provides a complex allegory which can be drawn upon either to supplement or to amplify usage b. Collocates indicating this meaning of whore include, *cup*, *fornication*, *beast* and *kings*.

# Definitions - 70

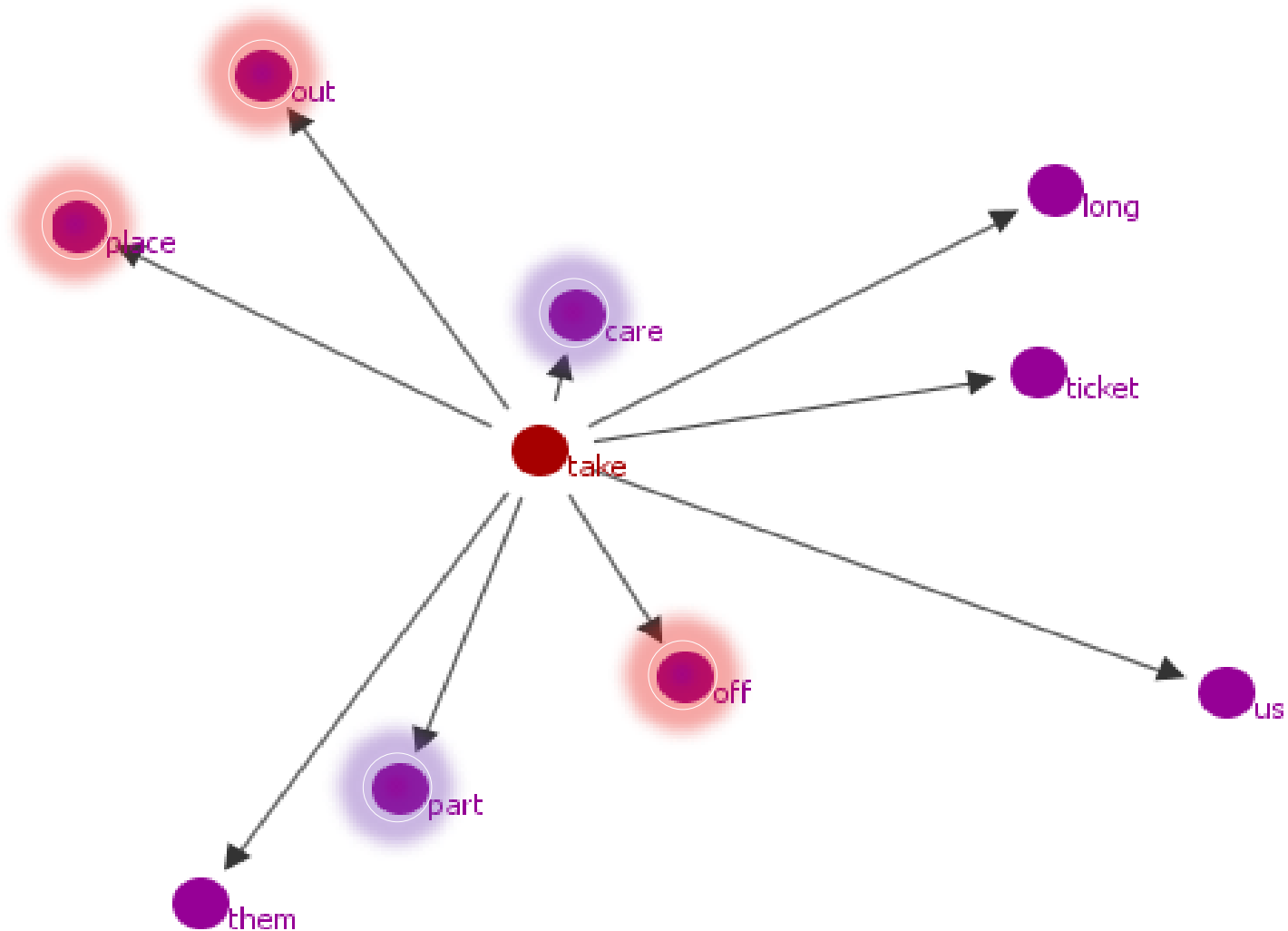
- Whore – noun a: A woman who **chooses to sell** sexual favours, **often in association with a bawd**. She is **dishonest and** associated with criminality, **disease**, a range of undesirable character traits, e.g. impudence, **and the production of illegitimate children**. To have a whore as a relative was viewed as a problem, **though whores generally may evoke pity**. **Nonetheless, people of quality associated with whores**.
- Whore – noun b: a metaphorical usage, drawing on the underlying meaning of *whore* used to insult women.
- Whore – noun c: the Whore of Babylon provides a complex allegory which can be drawn upon either to supplement or to amplify usage b. Collocates indicating this meaning of whore include, *cup, fornication, beast* and *kings*.
- What about synchronic?

See: <https://t.co/IGHvyJMyqg> McEnery, T and Baker, H (2017) *Corpus Linguistics and 17th-Century Prostitution*, London, Bloomsbury.

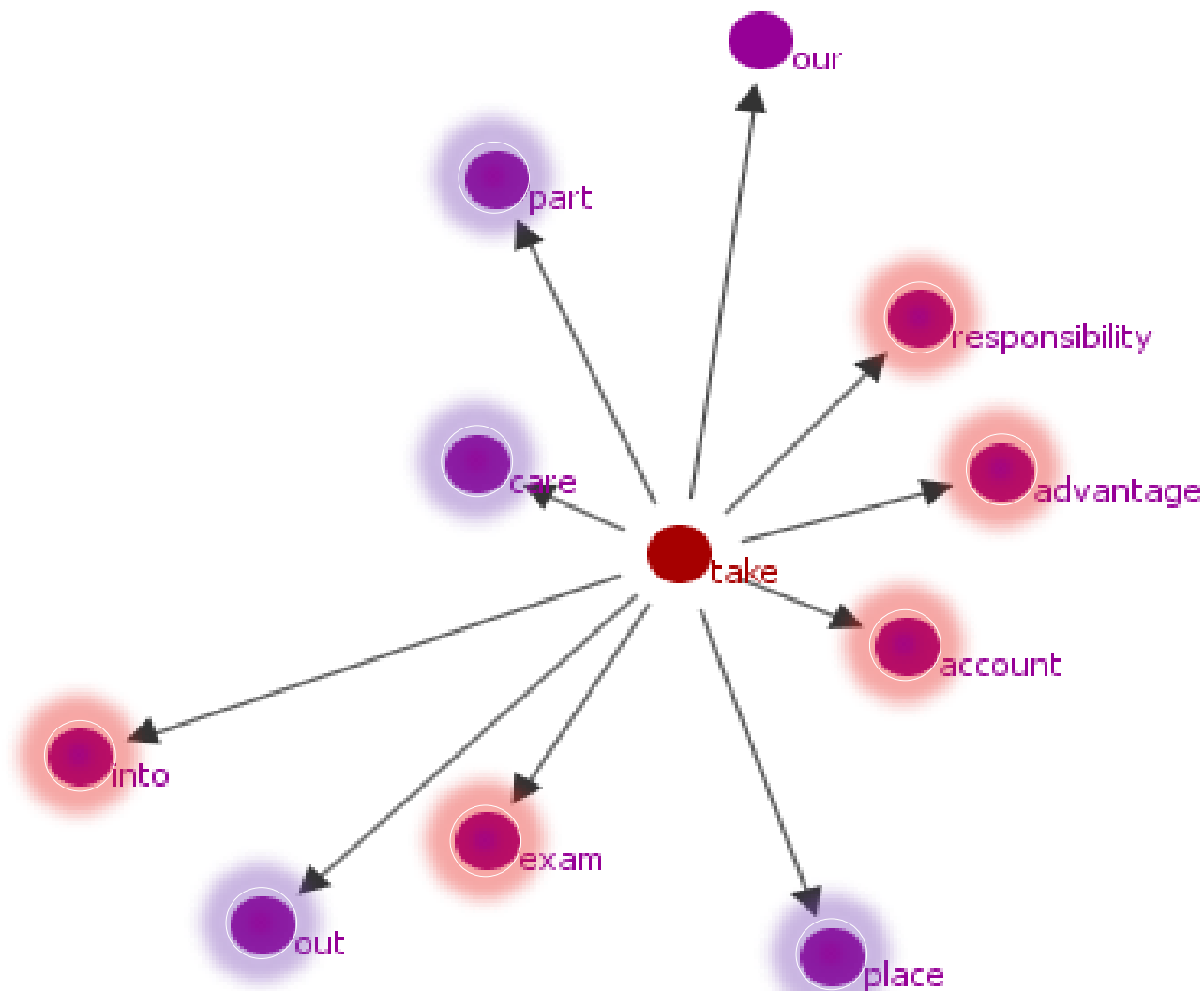
## Level: lower-intermediate (B1)



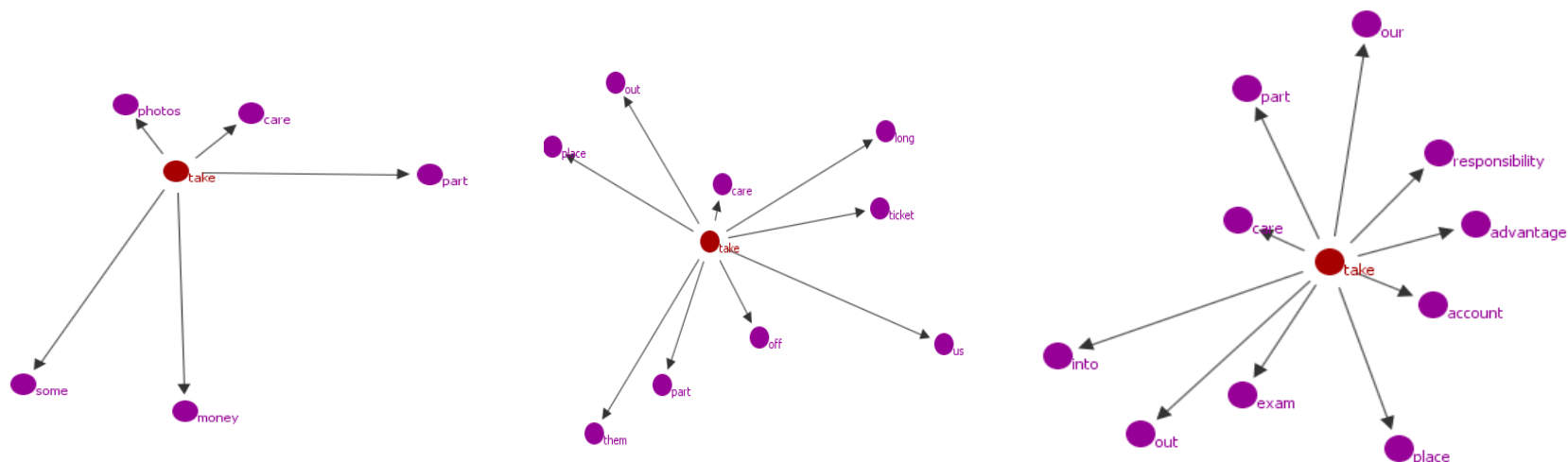
## Level: intermediate (B2)

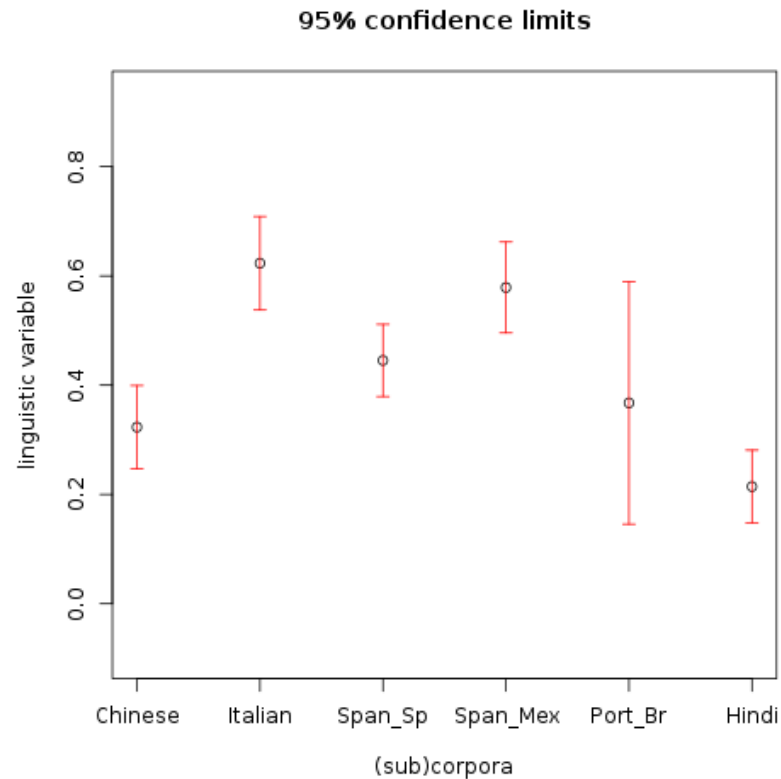


## Level: Advanced (C1-C2)



# Language development as observed in Trinity Lancaster Corpus





95% CI Error bars around the mean: Candidates' L1 background: Conversation

Spanish speakers from Mexico use more laughs than Spanish Speakers from Spain. This seems to point to cultural (pragmatic) differences.



# What is not there

- Comprehension also often typically exceeds production
- Many things may suppress production
- Sometimes close reading can be our only guide – we are offered precious insights when suppressed topics are mentioned
- Homosexuals in the 17<sup>th</sup> century – *spintrians*.
- Classical influence: *darling, mollis, obscaenus, pathicus*

Baker, H.S. and McEnery, T. (2018) 'Using corpora to explore shadows from the past: early modern male prostitution, an evasive marginal identity' in A-J. Zwierlein, J. Pertzold. K. Boehm and M. Dekker (eds) *Anglistentag 2017 Regensburg Proceedings*, Wissenschaftlicher Verlag, Trier, pp 9-20.

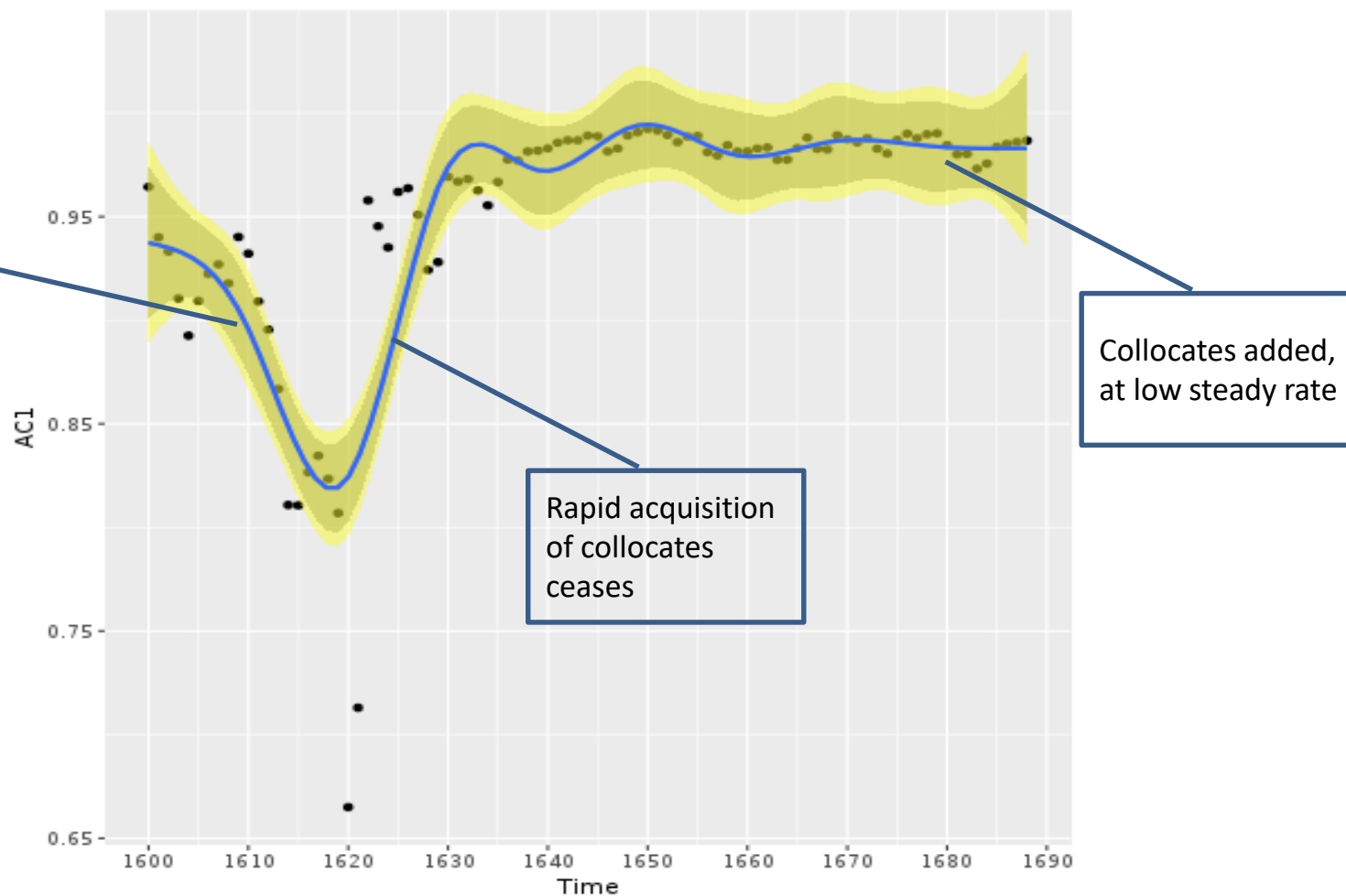
McEnery, T. and Baker, H.S. (2017) 'The public representation of homosexual men in seventeenth century England – a corpus based view', *Journal of Historical Sociolinguistics*, Vol. 3, Issue 2, pp 197-217.

# Case Study 3 – Meaningful Annotation. *its*

- Previous corpus-based studies which we can compare our results to (Culpeper and Kytő, 2010; Nevalainen and Raumolin-Brunberg, 1994 and 2003)
- Based on sections of the Corpus of English Dialogues (1.18 words, 1560-1760), the Corpus of Early English Correspondence (2.7 M words, 1410-1681) and the Helsinki Corpus (1.67 M words, 730-1710)
- Even though Culpeper and Kytő split the seventeenth century into chunks (1600-1640, 1640-1680 and 1680-1720) they gather only two examples of *its* in the Helsinki corpus in this period, for example
- it appears that *its* undergoes a shift in usage from 1640 onwards (Culpeper and Kytő, 2010: 188)
- Slowly supplants competing choices, such as *of it* for expressing what Culpeper and Kytő call the third person neuter possessive (Culpeper and Kytő, 2010:187).

Culpeper, J. and Kytő, M. (2010). *Early Modern English Dialogues: Spoken interaction as writing*, Cambridge University Press, Cambridge.

Nevalainen, T. and Raumolin-Brunberg, H. (1994). 'Its beauty and the beauty of it: the standardization of the third person neuter possessive in Early Modern English', in D. Stein and I. Tienen-Boob van Ostade (Eds.) *Towards a Standard English, 1600-1800*, Mouton de Gruyter, Berlin.



Labov, W. (1994). *Principles of linguistic change, volume 1: Internal Factors*. Oxford: John Wiley & Sons.

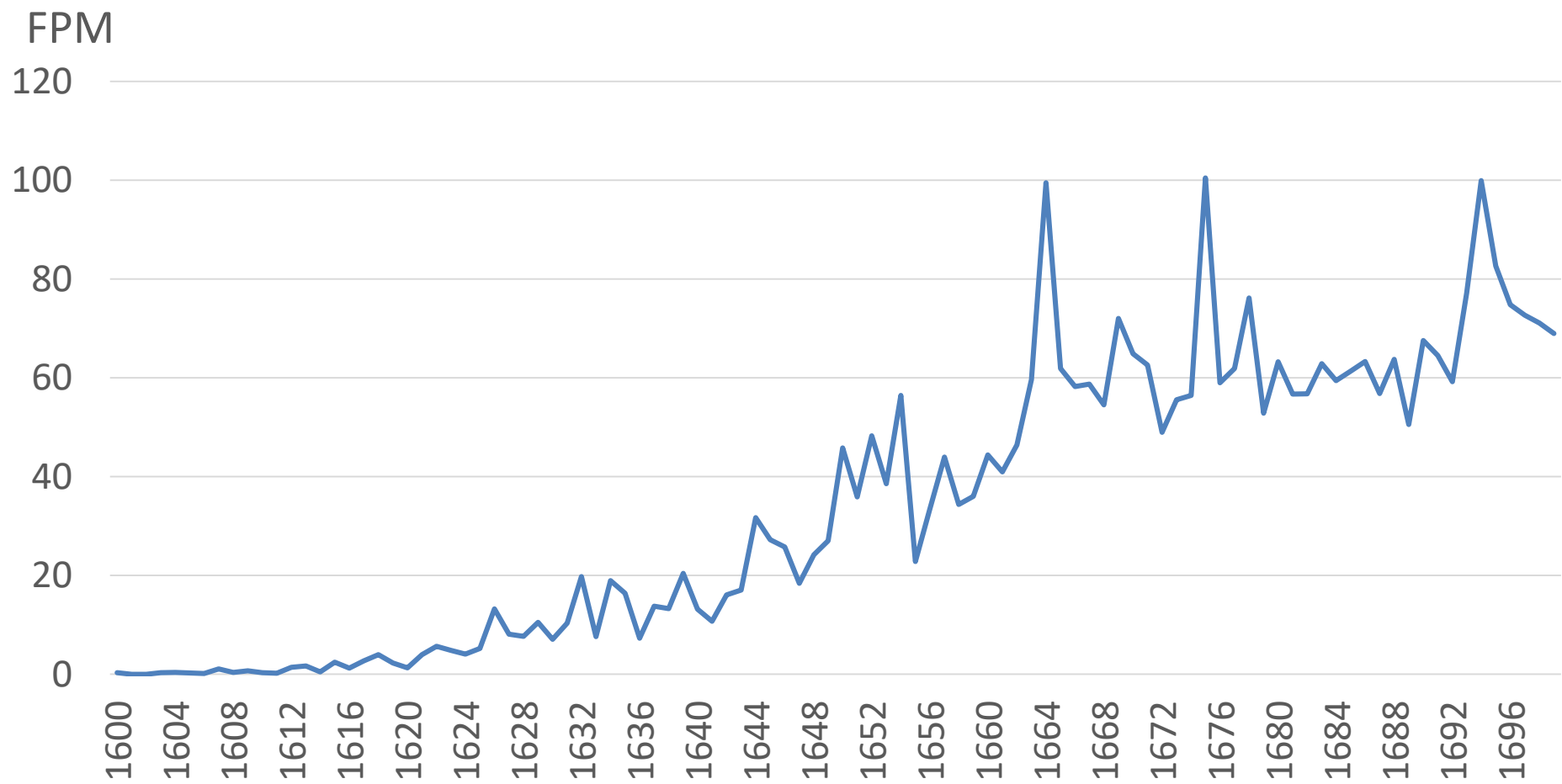
McEnery, T., Brezina, V. and Baker, H. (2019) 'Usage Fluctuation Analysis: A new way of analysing shifts in historical discourse', *International Journal of Corpus Linguistics*, 24 (4), pp 413-444.

Slot	POS	1600s	1610s	1620s	1630s	1640s	1650s	1660s	1670s	Total
-5	APPGE		1							1
-2	NN1							2		2
-2	VVN					1				1
-1	VVI							1		1
-1	VVZ	4	3			1				8
1	JJ	3	3					2		9
1	NN1	21	11			11	3	1		54
1	NN2	5	1	1		2	2			12
2	NN1	6	2							8

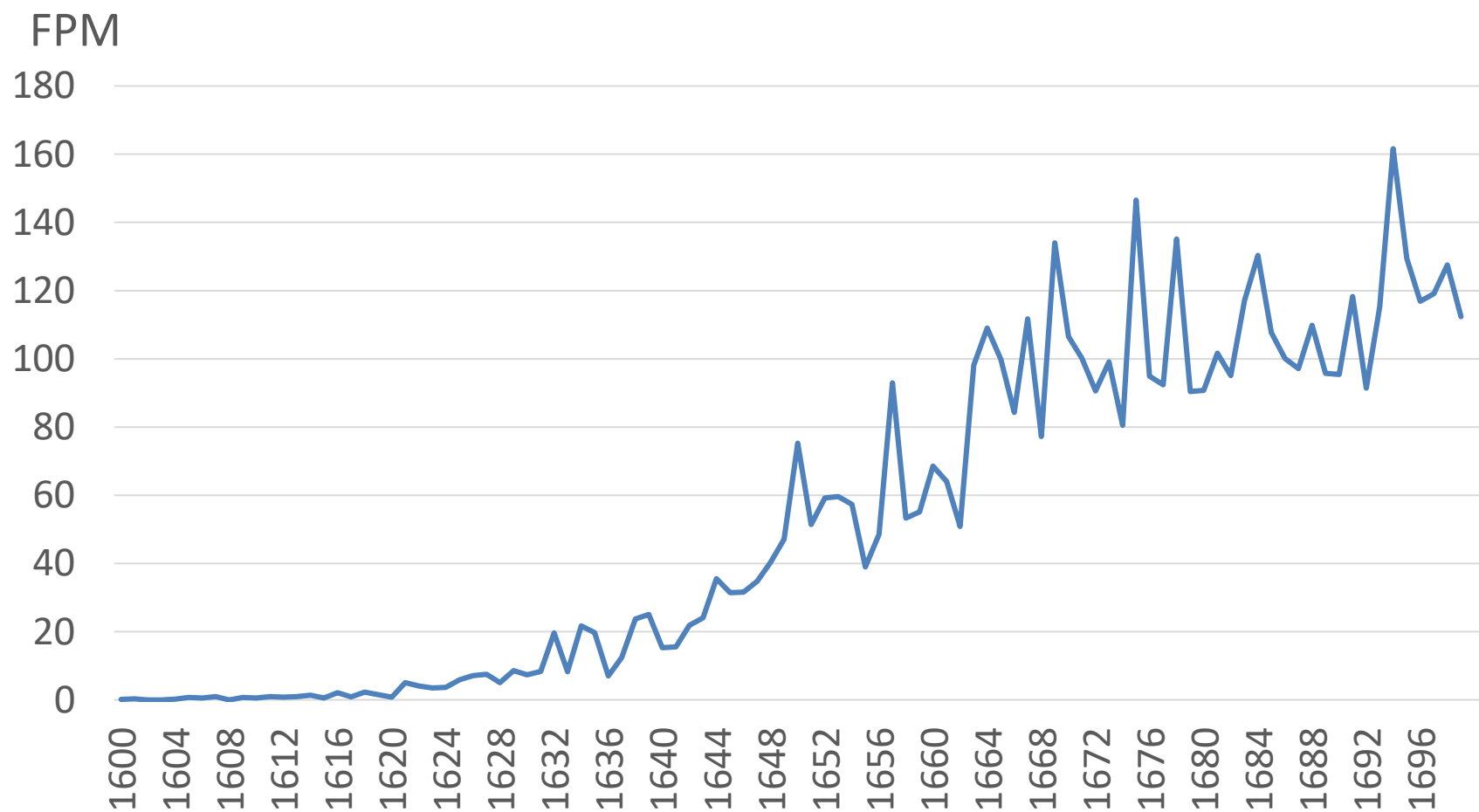
a possessive determiner used to 'specify a noun phrase by relating it to ... entities mentioned in the text or given in the speech situation' (Biber, Johansson, Leech, Conrad and Finegan, 1999:270-271) VERB its NOUN 18.2% of examples of *its* in BNC2014 Spoken

"Mephibosheth it hath caught a fall, and is lame on its feet" (Everard, 1618)

Modal slots for collocations which begin in a particular decade and then persist.



NOUN *of its* in each year of the 17<sup>th</sup> century in the EEBO corpus, FPM



*its* with a plural common noun at +1 in the 17<sup>th</sup> century in EEBO (FPM)

FPM

20

15

10

5

0

1600

1606

1612

1618

1624

1630

1636

1642

1648

1654

1660

1666

1672

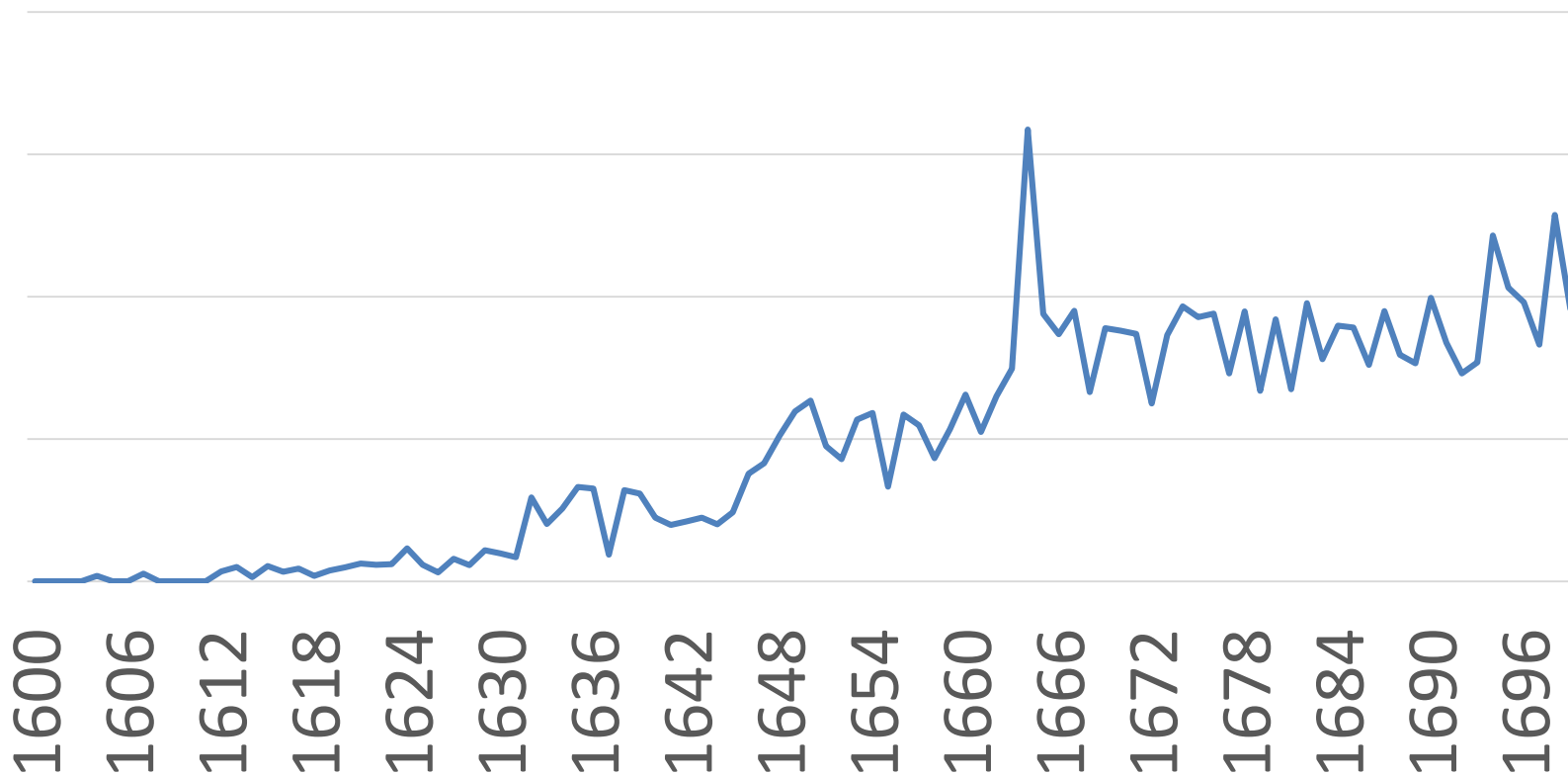
1678

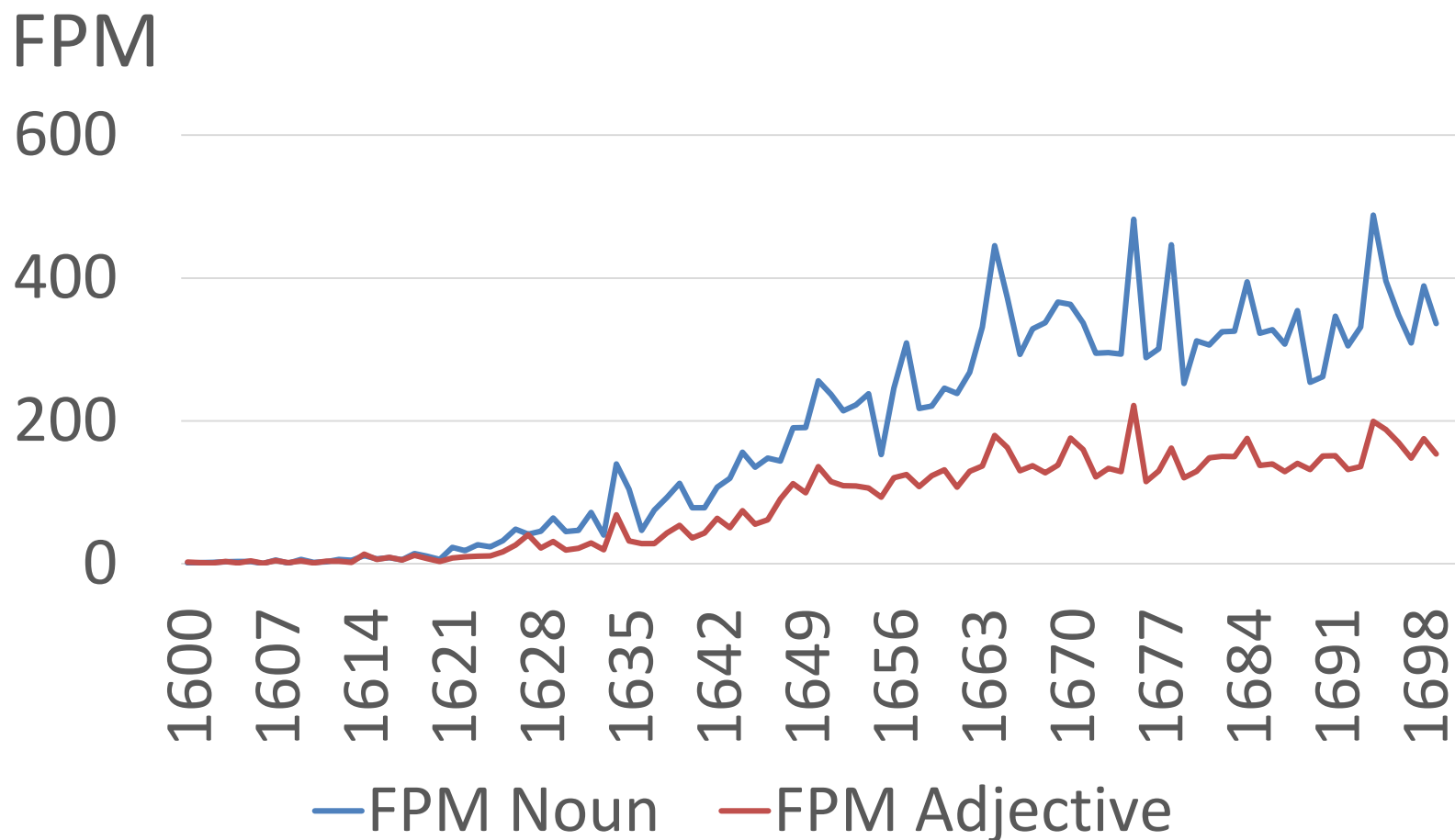
1684

1690

1696

*its* immediately preceded by a past participle in the 17<sup>th</sup> century in EEBO





FPM of nouns and adjectives immediately following *its*

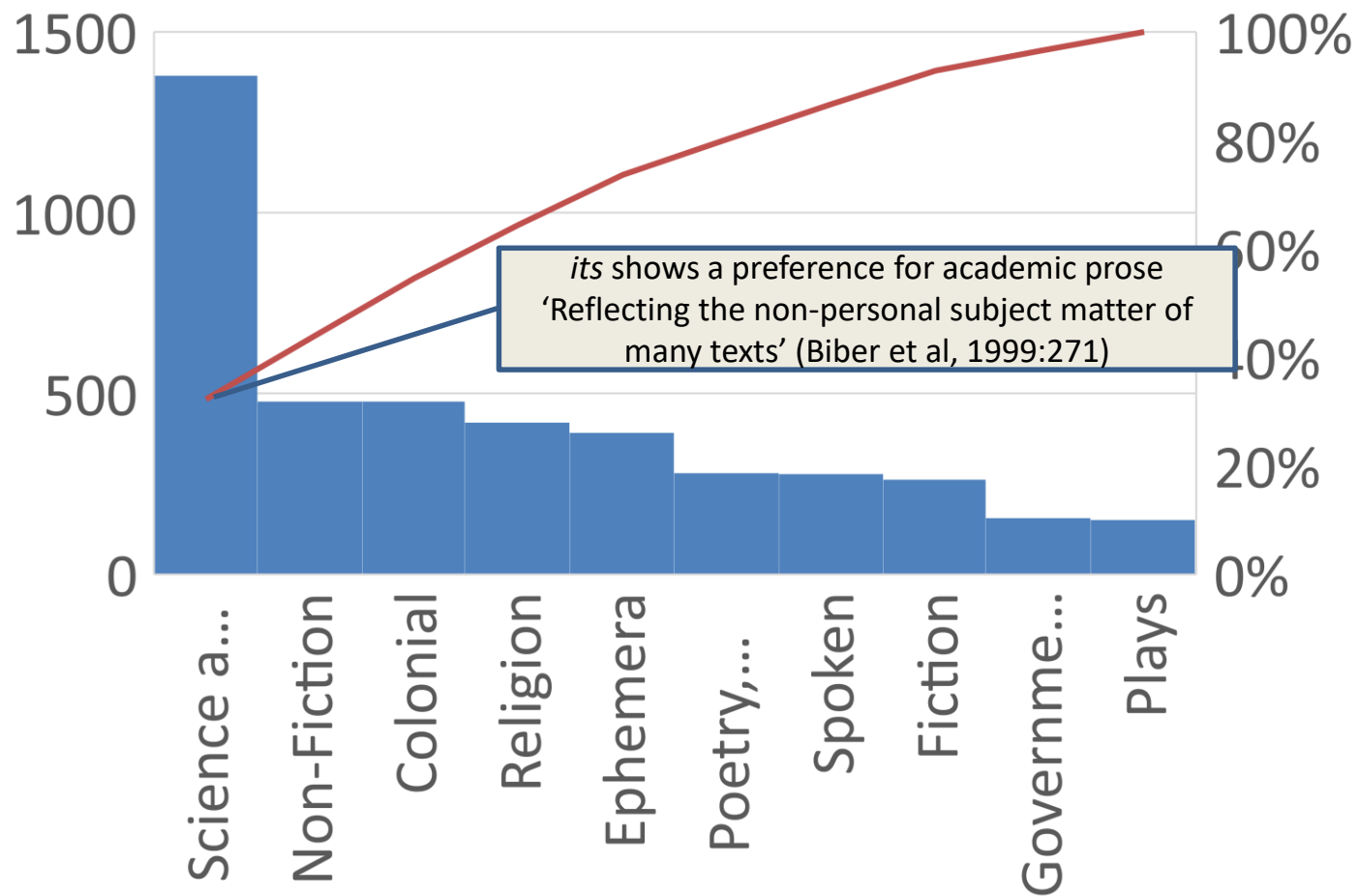


# The structure we added - Genres and Domains

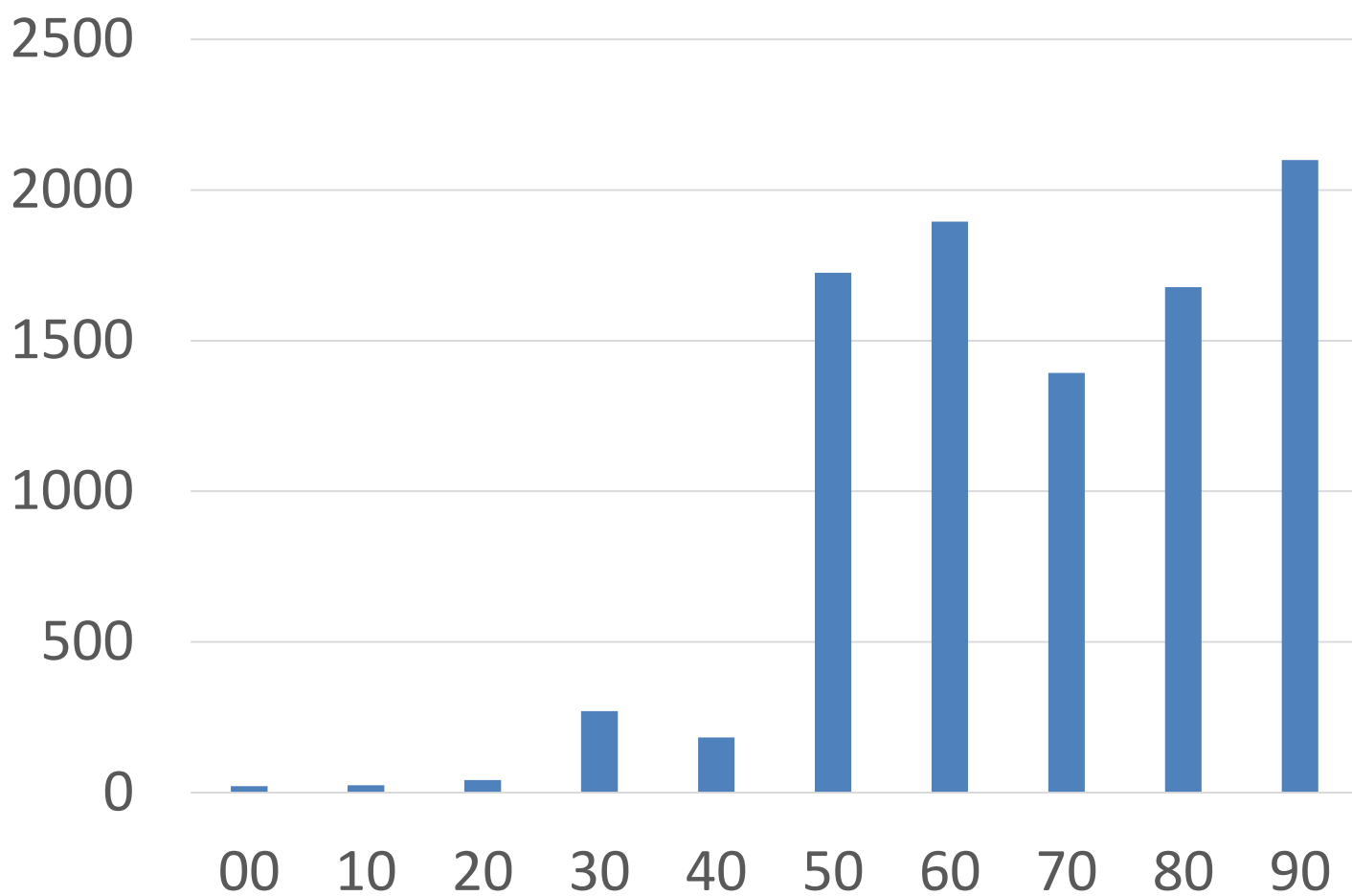
- the five domains of EEBO are **Literary** (which contains four genres, Plays; Poetry; Verse & Song; Fiction; and General);
- **Religious** (containing five genres of Bible; Catholicism; Protestantism; Doctrine, Theology & Governance; and General);
- **Administrative** (split into the four genres of Royal; Parliamentary; Legal; and General);
- **Instructional** (with five genres of Philosophical; Science; Mathematics; Medicine; and General);
- **Informational** (with six genres of Biography; Colonial; Essay; Letters; Pamphlets; and General).
- Many of these genres are further sub-divided – can be very helpful, e.g. masques.
- Since worked on by Sean Murphy and being used on the Shakespeare Encyclopaedia project (see ICAME Journal, Volume 43, Issue 1, pp 59-82).

# How did we do it?

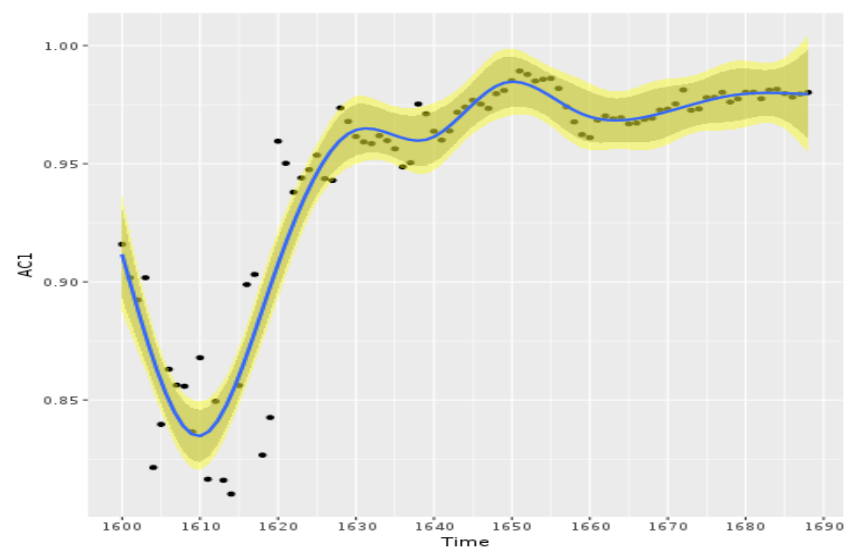
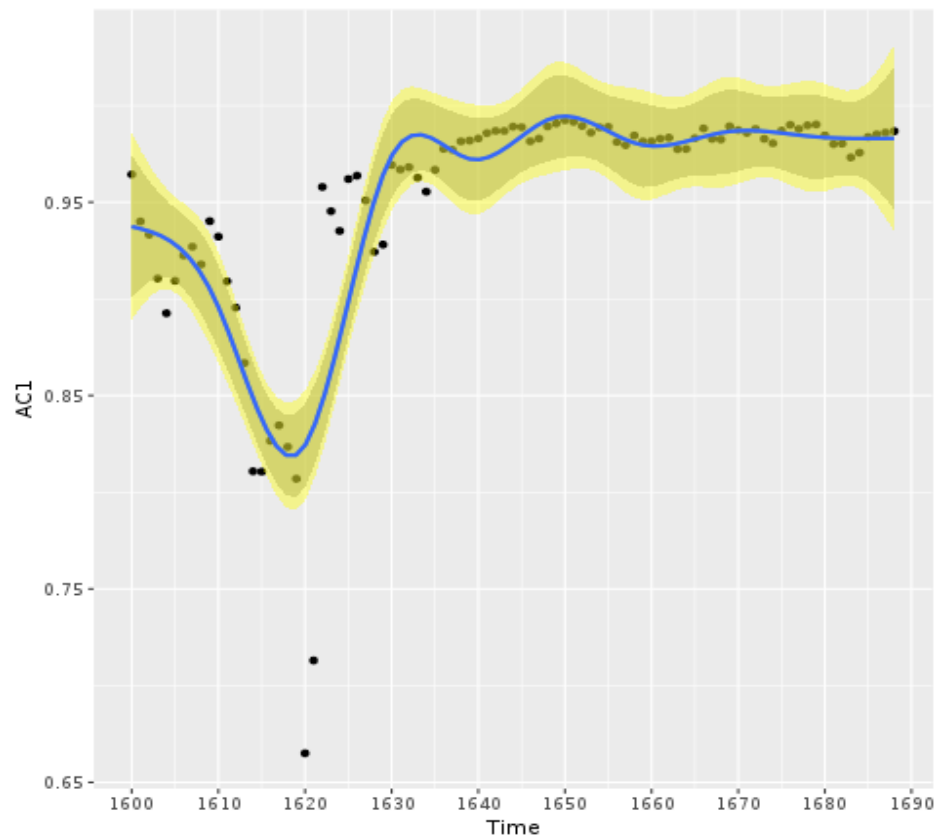
- Do not think of modern genres – think of how the book or text was advertising itself.
- Taavitsainen (2001a: 140) says that genres function as “‘horizons of expectation’ for readers to know what to expect and models of writing for authors.”
- The scheme was built from the bottom up – with the title being assumed to signal the ‘horizon of expectation’ to the reader.
- Separate work going on to look at the linguistic properties of the genres.
- Initial categorization by me, Helen Baker worked on the difficult cases.
- Since worked on by Sean Murphy and being used on the Shakespeare Encyclopaedia project.



Pareto chart showing the distribution of *its* across domains in the C 17 in EEBO



*its* in the Science and Astrology genre by decade in EEBO in the C 17



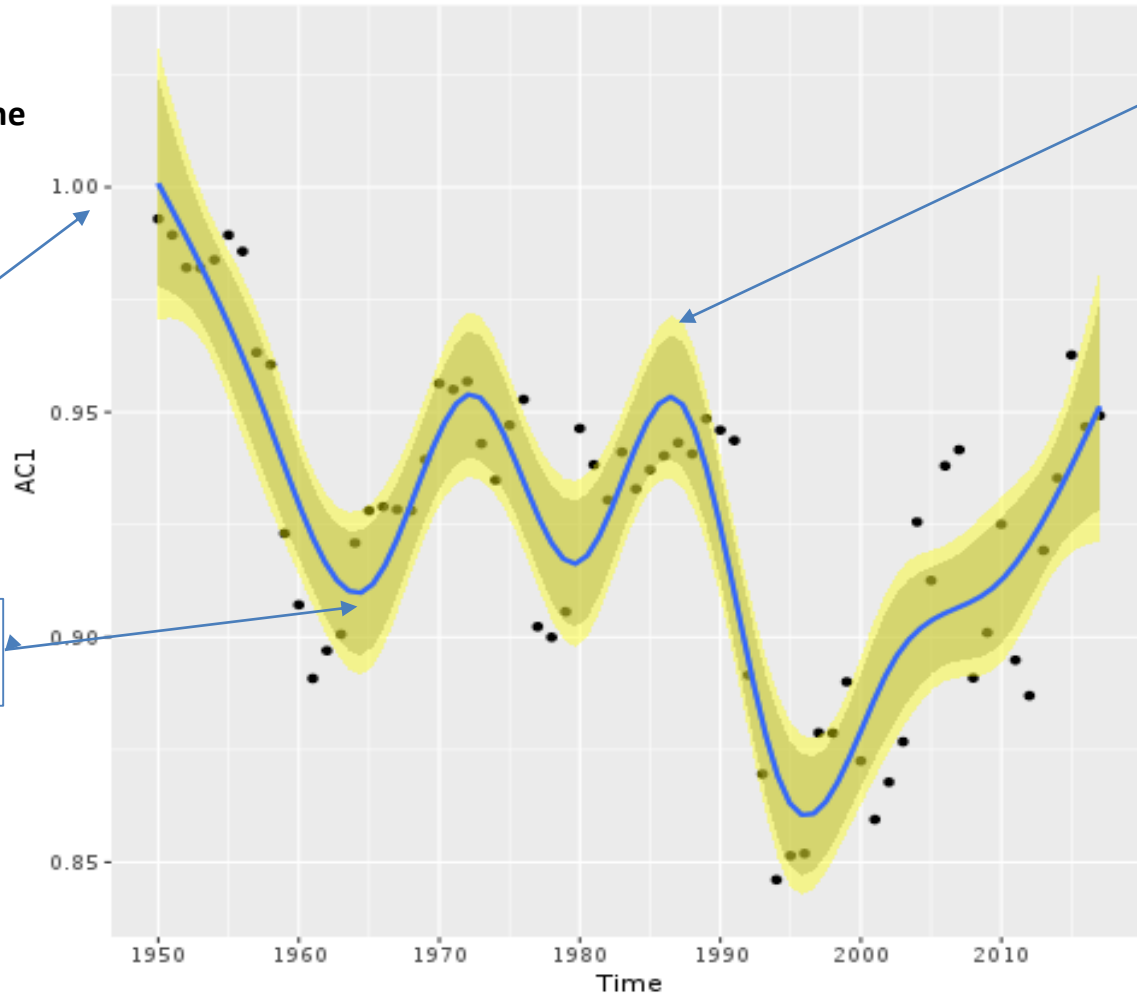
# Conclusion

- The basic message (1) – big corpora allow us to look at some things that small corpora cannot.
- The basic message (2) - looking at things we have not been able to look at before has all sorts of unpredictable consequences
- The basic message (3) – when looking at things, we may find as many questions as answers, and to answer those questions we may need new resources

**European  
Community/the  
Community in The  
Times**

Scale of  
agreement  
(0-1)

Local Minimum  
of Agreement



Local Maximum of  
Agreement

Accelerating  
convergent change

Accelerating  
divergent change

Stability