

Scholarship in the EEBO-TCP Age

John Lavagnino

King's College London

17 September 2012

<http://www.slideshare.net/jlavagnino/scholarship-in-the-eebotcp-age>

EEBO-TCP

It's everywhere in early modern studies,
though largely hidden: overt citation and
discussion are minimal.

My topics

- 1 The necessity and uniqueness of TCP
- 2 Three kinds of TCP-based research
- 3 TCP's distinctive model for organization and funding

Other themes

- 1 How much does silence matter?
- 2 What are the unavoidable limitations of TCP?

Necessity and uniqueness: the 1520 problem

Matjaž Perc, “Evolution of the most common English words and phrases over the centuries”, *Journal of the Royal Society Interface*, forthcoming: see:

<http://goo.gl/7S0RT>

Based on Google ngram data, not TCP

A surprising claim about English

Perc, in his abstract: “We find that the most common words and phrases in any given year had a much shorter popularity lifespan in the sixteenth century than they had in the twentieth century.”

Top 3-grams, 2007 and 2008

2007			2008		
1.	one of the	0.020930	1.	one of the	0.022229
2.	as well as	0.019297	2.	as well as	0.017752
3.	the United States	0.014517	3.	part of the	0.013909
4.	part of the	0.013052	4.	the United States	0.013222
5.	in order to	0.011314	5.	out of the	0.012469
6.	the end of	0.011068	6.	the end of	0.010361
7.	out of the	0.010939	7.	in order to	0.010345
8.	the fact that	0.008849	8.	some of the	0.008103
9.	some of the	0.008248	9.	to be a	0.008015
10.	to be a	0.008080	10.	the same time	0.007895

See: <http://goo.gl/iUS3e>

Top 3-grams, early 1520s

1520			1527		
1.	of the pope	0.070325	1.	the name of	0.228102
2.	the canon law	0.056651	2.	in the name	0.228102
3.	of the German	0.041023	3.	to be baptized	0.182482
4.	of the Church	0.037116	4.	command of Christ	0.136861
5.	the German nation	0.031255	5.	name of the	0.136861
6.	the pope and	0.029302	6.	and of the	0.136861
7.	the pope has	0.029302	7.	according to the	0.136861
8.	to the pope	0.027349	8.	the forgiveness of	0.114051
9.	there is no	0.025395	9.	heaven and earth	0.114051
10.	that the pope	0.025395	10.	the Holy Ghost	0.114051

See: <http://goo.gl/r4eyh>

From 1541's top 3-grams

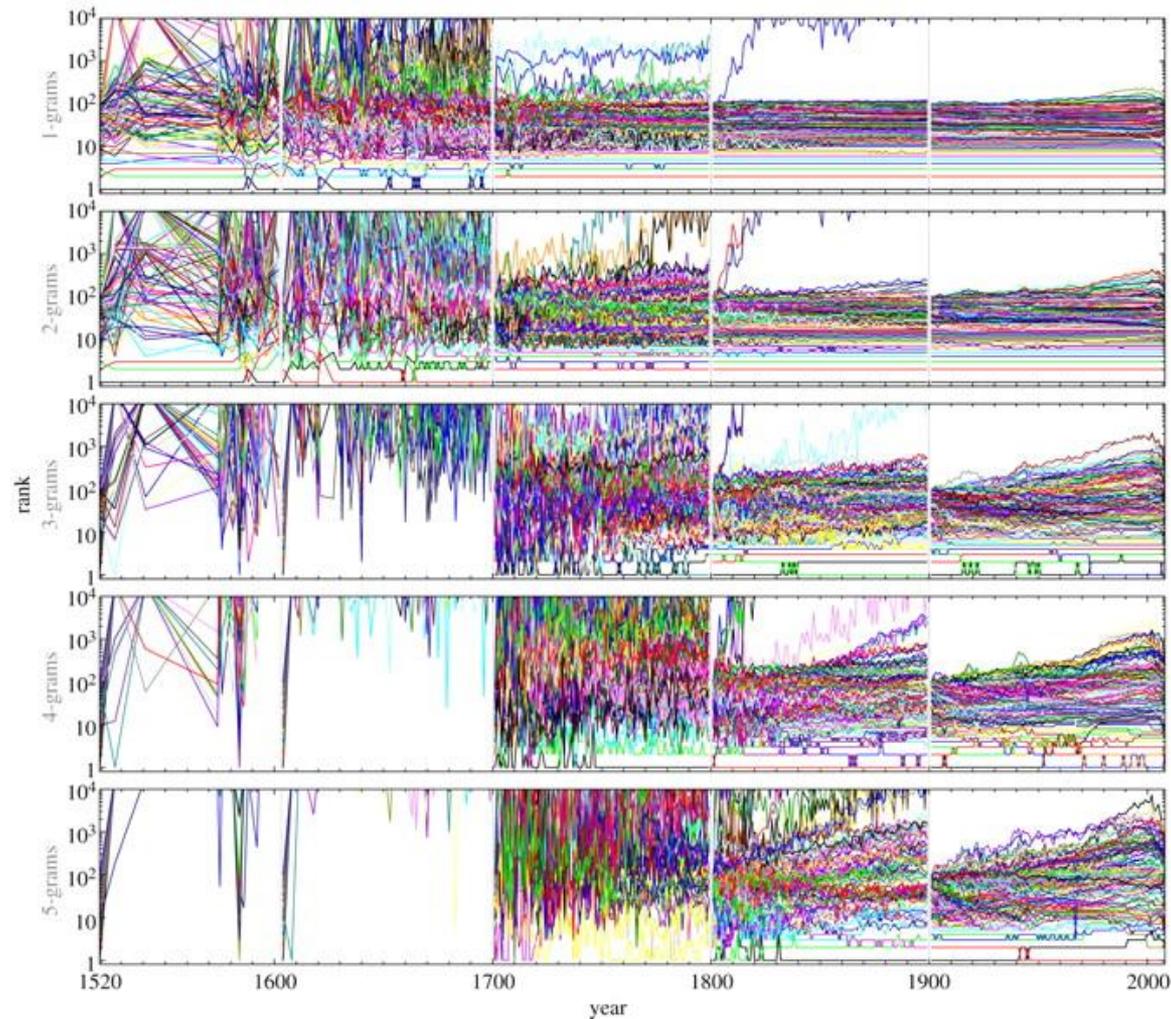
16.	editions of the	0.059335
17.	very large number	0.059335
18.	are in the	0.059335
19.	in the Breviary	0.059335
20.	an edition of	0.059335
21.	in the Bodleian	0.059335
22.	i oth January	0.039557

See: <http://goo.gl/r4eyh>

More reflections on C16 language

“Phrases that were used most frequently in 1520, for example, only intermittently succeeded in re-entering the charts in the later years.”

Evolution of popularity of the top 100 n-grams over the past five centuries.



Perc M J. R. Soc. Interface doi:10.1098/rsif.2012.0491

See: <http://goo.gl/2URVT>

Some alternative conclusions about this research

The world's best mass OCR is bad for books
before 1800

Interdisciplinary journals need to have
reviewers from many fields

Perc's publication of his data and an
interface for exploring it is praiseworthy

The necessity and uniqueness of EEBO-TCP

Despite the resources poured into it, Google Books is not an adequate representation of books prior to 1800: too few books early on, bad metadata, bad OCR.

Just how much can we know about
English writing in 1520?

How many STC titles were published in
1520? How many are planned for inclusion
in TCP?

Visualization from STC, volume 3, 1991

1518 — *cont.*

Stanbridge, *Parvula* 23167.3(?); Stanbridge, *Vulgaria* 23196.4(?); Sulpitius 23429; Thomas, a Kempis 23956(?)*; Tunstall 24320; Virgilius Maro 24828(?); Whittinton, De Heteroclitis Nominibus 25459.11(?), 25460; Whittinton, De Nominum Generibus 25479.9 (?), 25479.11(?); Whittinton, De Octo Partibus 25498.7(?); Whittinton, De Synonymis 25527.5(?), 25527.8(?); Whittinton, Syntaxis 25545, 25545.3, 25545.5.

1519

Catharine, of Siena 4815; Cock Lorrel 5456.3(?); Complaint 5609 (?), 5609.5(?); Computus 5613; England, Statutes, Abridgements/Extracts 9515.5; England, Statutes, Gen. Colls. 9268; England, Yearbooks 9598; Erasmus, D. 10450.6; Eulenspiegel 10563(c.7); Exoneratorium 10629; Fisher, John, St. 10904(?)*; Horman 13811; Hylton 14043.5; Indulgences, Cont. 14077c.93, 14077c.94; Indulgences, Eng. 14077c.48(?), 14077c.50, 14077c.53, 14077c.84A; Kur'an 15084(?); Liturgies, Latin Rite, Antiphoners 15790; Liturgies, Latin Rite, Breviaries 15816; Liturgies, Latin Rite, Hours and Primers 15922, 15923, 15924; Liturgies, Latin Rite, Hymns 16128.2; Liturgies, Latin Rite, Missals 16199, 16200, 16201, 16201.3, 16201.7; Liturgies, Latin Rite, Process. 16235; Mirk 17973.5; Narrationes 18362.5(?); Rede 20824.5(A/C); Remedy 20876; Retorna 20894.7; Rolle 21260(?); Stanbridge, Accidence 23147.6, 23154.3; Stanbridge, Gradus Comparationum 23159a(?); Stanbridge, *Parvula* 23167.5(?); Stanbridge, *Vocabula* 23181; Stanbridge, *Vulgaria* 23196.6(?), 23196.8; Thomas, a Kempis 23956(?)*; Whittinton, De Heteroclitis Nominibus 25461, 25461.5; Whittinton, De Octo Partibus 25499; Whittinton, De Syllabarium Quantitate 25512, 25514; Whittinton, De Synonymis 25528; Whittinton, Declinationes Nominum 25446; Whittinton, Libellus Epigrammatum 25540.5; Whittinton, Syntaxis 25546.

1520

Aesop 170.3(?); Almanacks (Adrian) 406.7; Almanacks (Laet, G., the elder) 470.5, 470.6; Aphthonius 699(?); Barbara 1375.5(c.); Book 3288(o.s.)*; Canutus 4593(c.); Constable, J. 5639; Croke, R. 6044a.5; Dietary 6833; Emanuel, King of Portugal 7677(?); England, Appendix 10001; England, Local Courts 7707(?); England, Proclamations, Chron. Ser. 7769.2; England, Statutes, Chron. Ser. 9362.5(c.), 9362.7(c.); England, Yearbooks 9576, 9595; Erasmus, D. 10450.2, 10450.3, 10450.7; Erasmus, St. 10435; Exoneratorium 10630(?), 10631(?); Goodwyn 12046(?); Heroum 13256(?); Hortus 13835; Indulgences, Cont. 14077c.90(?), 14077c.90A(?), 14077c.95, 14077c.96, 14077c.97, 14077c.98(c.), 14077c.99; Indulgences, Eng. 14077c.26(c.), 14077c.45(?), 14077c.59(c.), 14077c.67A, 14077c.68A (c.), 14077c.72(c.), 14077c.73(c.), 14077c.84(?); Indulgences, Images of Pity 14077c.23A(c.); Indulgences, Stations of Rome 14077c.149(c.), 14077c.150(c.); Indulgences, unassigned 14077c.154 (c.); Jacob, the Patriarch 14323.5(c.); Jesus Christ 14547.5(c.); Joseph, of Arimathea 14807; Kalenberg 14894.5(c.); Liturgies, Latin Rite, Antiphoners 15790a; Liturgies, Latin Rite, Breviaries 15791.5(c.), 15817(?); Liturgies, Latin Rite, Hours and Primers 15925, 15926(?), 15926.5(?), 15928(?), 15929(?), 16104.5; Liturgies, Latin Rite, Hymns 16128.3(c.); Liturgies, Latin Rite, Manuals 16143(?); Liturgies, Latin Rite, Missals 16202, 16202.5(c.), 16224.5 (c.); Liturgies, Latin Rite, Psalters 16259.5(c.), 16259.7(c.); Lydgate 17027(?), 17030.7(c.), 17035(?), 17038(?); Mancinus 17241(?), 17242.5(?); Mary Magdalen 17568(?); More, T. 18088; Parliament 19303.7(c.), 19305.3(A/C); Peter, of Luxemburg, Sr. 19795(?); Petrus, G. 19816(c.7)(A/C); Rastell, J., Printer 20722(?); Rome, Hospital of the Holy Ghost 21310.5(c.); Splynter 23102(?); Squire 23111.5(?); Stanbridge, Accidence 23147.8(?), 23148 (?), 23148.2; Stanbridge, Gradus Comparationum 23159a.1(?), 23159a.2(?); Stanbridge, *Parvula* 23167.7; Stanbridge, *Vocabula* 23181.2(?), 23181.3(c.), 23181.4(c.); Stanbridge, *Vulgaria* 23196a (?), 23196a.2(c.); Sulpitius 23429.5; Tenures 23879.5(c.); Terentius 23894(c.); Thomas, a Becket 23954(c.), 23954.3(c.); Treatise

1520 — *cont.*

24241(?); Valuation 24591(c.); Whittinton, De Heteroclitis Nominibus 25462; Whittinton, De Nominum Generibus 25479.14, 25479.15; Whittinton, De Synonymis 25529; Whittinton, Declinationes Nominum 25446.5(c.); Whittinton, Syntaxis 25547; Whittinton, *Vulgaria* 25569.3, 25569.5, 25569.7, 25570.

1521

Almanacks 388(bef.); Austin, St. 965; Baldwin, Abp. 1242; Barclay, A. 1384b(?), 1386; Book 3288(?); Bradshaw, H. 3506; Brendan 3600(?); Bullock 4082; Christmas 5204; Cicero 5311; Darius, S. 6279; Dietary 6834, Du Castel 7270, 7271; England, Local Courts 7709(?), 7725.9(?); England, Statutes, Abridgements/Extracts 9516, 9518.7*; England, Statutes, Chron. Ser. 9358.7(?); England, Yearbooks 9883(?); Erasmus, D. 10496; Exoneratorium 10631.5(?); Fisher, John, St. 10894(?), 10898; Galen 11536; Henry VIII, King of England 13078, 13083; Horman 13807; Indulgences, Eng. 14077c.32, 14077c.65, 14077c.66, 14077c.67; Indulgences, Images of Pity 14077c.11(C)?; Jesus Christ 14558; Justices of Peace 14866, 14867; Labellus 15580(?), 15580.5; Lily, W. 15601.5(?), 15606; Liturgies, Latin Rite, Hours and Primers 15930, 15931, 15932(?), 15932.5(aft.); Liturgies, Latin Rite, Missals 16203, 16204, 16204.5; Lucian 16896; Murellius 18292.7(c.); N. 18324.5; Ricardus 20972; Skelton 22611.5(?); Stanbridge, Gradus Comparationum 23159a.3(?); Stanbridge, *Parvula* 23168; Stanbridge, *Vocabula* 23181.5; Ten Commandments 23877; Treatise 24242(?), 24242.3 (?); Voragine 24879.5; Whittinton, Antilycon 25441.2; Whittinton, De Heteroclitis Nominibus 25464; Whittinton, De Nominum Generibus 25480.3, 25480.7, 25481, 25482; Whittinton, De Octo Partibus 25500; Whittinton, De Syllabarium Quantitate 25515, 25515a; Whittinton, De Synonymis 25530; Whittinton, Declinationes Nominum 25448; Whittinton, Syntaxis 25547.3, 25547.5, 25548; Whittinton, Verborum Praeterea 25558, 25558.5, 25559; Whittinton, *Vulgaria* 25572.

1522

Almanacks 389; Copland, R. 5732.5(c.); Dionysius, Carthusianus 6895, 6896; Elias 7571.5(c.); England, Proclamations, Chron. Ser. 7769.4(c.); England, Statutes, Chron. Ser. 9361.3(?); Fisher, John, St. 10894.5(?); Galen 11531.5(?), 11532, 11534(?); Geminus, P. 11719; Henry VIII, King of England 13079; Indulgences, Eng. 14077c.33, 14077c.59G(?)(A/C); Indulgences, unassigned 14077c.151(?), 14077c.152(c.); Jacob, the Patriarch 14324(+); Jesus Christ 14552; Lactantius 15118(?); Langton, R. 15206; Lily, W. 15606.5(A/C), 15606.7; Littleton, T., Tenures, French 15725.5; Liturgies, Latin Rite, Hours and Primers 15933; Liturgies, Latin Rite, Manuals 16144; Liturgies, Latin Rite, Psalters 16260, 16260.5; Parliament 19305.5(?); Proclus 20398.3(?); Sallustius Crispus 21626(?)(A/C); Stanbridge, Accidence 23148.3; Stanbridge, Gradus Comparationum 23159a.4(?); Tavern 23707; Tunstall 24319; Virgilius Maro 24814.5; Whittinton, De Heteroclitis Nominibus 25464.5; Whittinton, De Nominum Generibus 25479.17, 25483.5, 25484, 25484.5; Whittinton, De Octo Partibus 25501, 25501.5(A/C), 25502; Whittinton, De Syllabarium Quantitate 25517*; Whittinton, De Synonymis 25531; Whittinton, Verborum Praeterea 25560, 25560.3, 25560.7, 25561; Whittinton, *Vulgaria* 25572.5; World 25982.

1523

Almanacks 389.3, 389.7; Barclay, A. 1383.5(?); Blakman 3123(?); Dietary 6835(?); Drunkards 7260; England, Local Courts 7709.5(?), 7726(?); England, Statutes, Chron. Ser. 9362.9(?); England, Yearbooks 9806.7(?), 9820(?); Fitzherbert, A. 10949*; Fitzherbert, J. 10994(?), 11005; Froissart 11296; Galen 11533; Honorius 13686(?); Indulgences, Eng. 14077c.39, 14077c.40, 14077c.74(c.); Jesus Christ 14550(c.); Libellus 15581(?); Littleton, T., Tenures, Engl. 15759.5(+); Liturgies, Latin Rite, Hours and Primers 15934,

A third of the 1520 entries

Aesop 170.3(?); Almanacks (Adrian) 406.7; Almanacks (Laet, G., the elder) 470.5, 470.6; Aphthonius 699(?); Barbara 1375.5(c.); Book 3288(o.s.?)*; Canutus 4593(c.); Constable, J. 5639; Croke, R. 6044a.5; Dietary 6833; Emanuel, King of Portugal 7677(?); England, Appendix 10001; England, Local Courts 7707(?); England, Proclamations, Chron. Ser. 7769.2; England, Statutes, Chron. Ser. 9362.5(c.), 9362.7(c.); England, Yearbooks 9576, 9595; Erasmus, D. 10450.2, 10450.3, 10450.7; Erasmus, St. 10435; Exoneratorium 10630(?), 10631(?); Goodwyn 12046(?); Hetoum 13256(?); Hortus 13835; Indulgences, Cont. 14077c.90(?), 14077c.90A(?), 14077c.95, 14077c.96, 14077c.97, 14077c.98(c.), 14077c.99; Indulgences, Eng. 14077c.26(c.), 14077c.45(?), 14077c.59(c.), 14077c.67A, 14077c.68A(c.), 14077c.72(c.), 14077c.73(c.), 14077c.84(?); Indulgences, Images of Pity 14077c.23A(c.); Indulgences, Stations of Rome 14077c.149(c.), 14077c.150(c.); Indulgences, unassigned 14077c.154(c.); Jacob, the Patriarch 14323.5(c.); Jesus Christ 14547.5(c.); Joseph, of Arimathea 14807; ...

Some very rough numbers for 1520

STC titles: 114

In English: 47

Currently in TCP transcriptions: 14

(Figures for both 1519 and 1521 are considerably smaller, because 1520 includes many items dated c.1520.)

The ideal data set

The kind of naïve statistical study Perc performed assumes an entirely reliable and consistent data set. The Google ngram data isn't like that, but while it can be done far better, a data set for early-sixteenth-century English of that kind is not possible.

Three key TCP uses

- 1 Simple quotation-finding
- 2 Larger-scale trawl for materials
- 3 Computational analyses

A (modern) quotation to find

John Carey, “The Missing Piece of the Jigsaw”:
Mollie Evans’s only written remark following
her breakup with William Golding:

There are two things which, tho' they cannot be heard by the physical ear a mile away, cry from end to end of the earth. The one is the crash of a tree that has been felled while it is still bearing fruit; the other is the sigh of a woman whom her husband sends away while she still loves him.

Quotation finding

Often requires a very broad search, rather than one limited by period

Can be conducted using error-ridden resources, as noted by Anthony Shipps, *The Quote Sleuth* (1990)

Something huge and Googleish can be best

Does it matter to know what resource was used, or do we just want the answer?

The large-scale trawl

You, too, can be Keith Thomas.

Michael Clanchy (1999, reviewing Alexander Murray on suicide in the Middle Ages):
“The traditional subjects are simpler to handle, because the information in the sources is already parcelled out that way.”

Did this study have something to do with TCP?

Eric Langley, *Narcissism and Suicide in Shakespeare and his Contemporaries* (2010).

Arnold Hunt, exaggerating somewhat:
“research has been transformed from a
labour-intensive handicraft into a
mechanized industry”.

The location of the labour

Instead of ingenuity in choosing books to scan, ingenuity in choosing what to search for.

Should we publish the details of our queries?

The problem of data laundering

Facts are facts, however you find them...

but a negative result depends a lot on
knowing what search method failed on
what resource

And the selection of what you discuss and
what you ignore is also now a more
pressing issue

Keywords

A line of research well suited to TCP, and with a background of methodological reflection: Raymond Williams, Quentin Skinner

An example: Peter Marshall, “The Naming of Protestant England”, *Past and Present*, February 2012

The problem of context

All keyword-study theory stresses context in some form; it has not developed ideas about working with large collections

An example: Phil Withington, *Society in Early Modern England: The Vernacular Origins of Some Powerful Ideas* (2010), and Tim Hitchcock's criticism (in *Economic History Review*)

An example from Withington

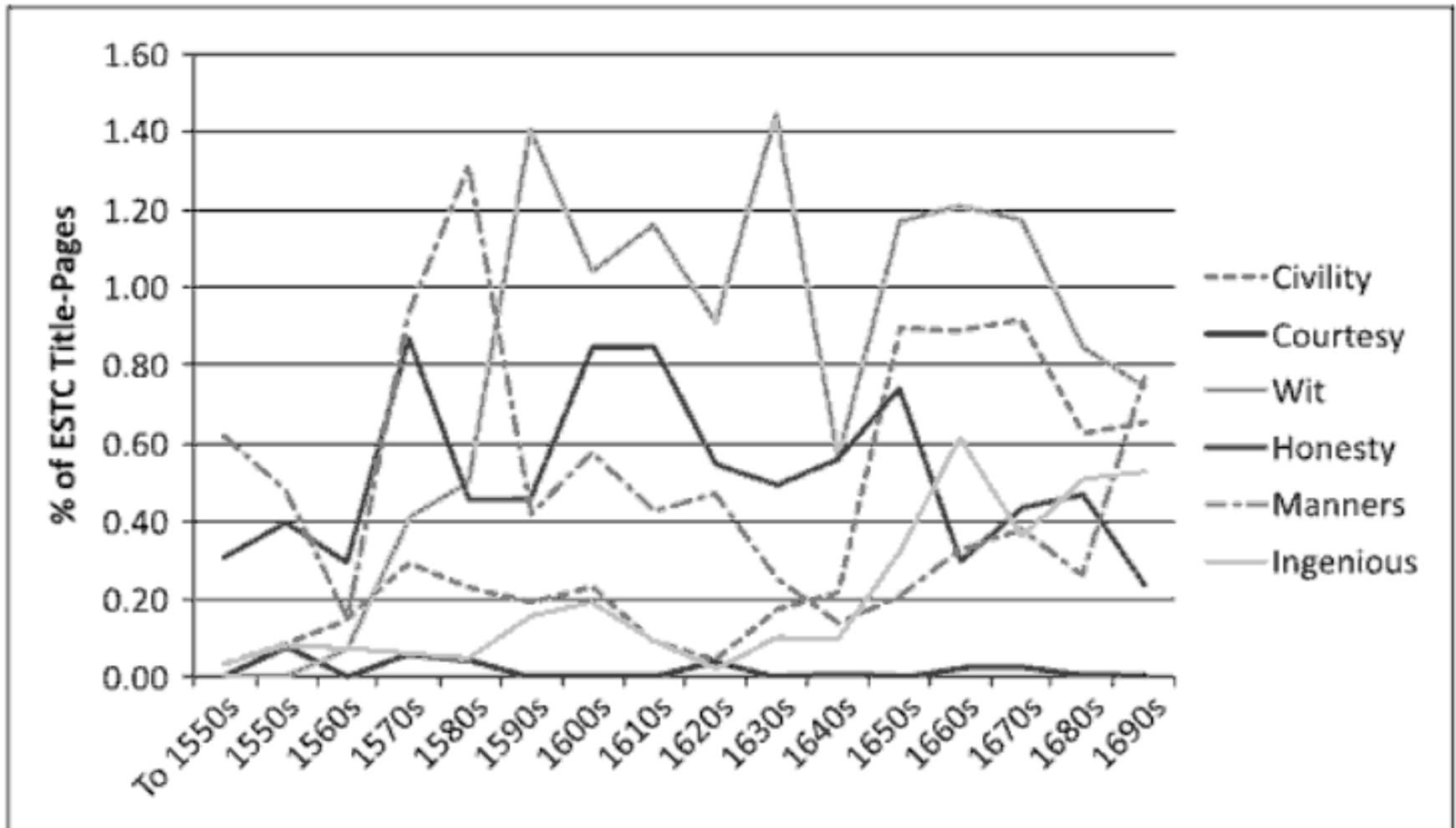


Figure 6.2 Vocabularies of Sociability (%), 1500–1700.

Open questions

We are comfortable with “unsystematic” discussion of examples gleaned through searching.

But can a large-scale study of “patterns and developments” find acceptance in early modern studies, or do we think context must always come first?

Is the data appropriate for the large-scale study?

Computational analyses

One form: finding ways to extend human understanding automatically (Moretti, Hope, Witmore)

Another form: mostly or entirely automatic systems (Jockers)

Early modern questions

Can the data really support it?

Do we need it for a small body of surviving texts?

Can we expect to get answers that resonate with traditional concerns?

Organization and funding

A superb invention: TCP's distinctive mixture of public and private funding, its discovery of an intermediate place between complete openness and effectively perpetual copyright, its avoidance of secrecy, its dissemination of work and knowledge while working on a large shared resource...

Thank you!

- John Lavagnino: please contact me at John.Lavagnino@kcl.ac.uk

