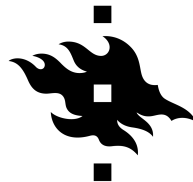


Social embedding of neologisms

in early English correspondence

Tanja Säily, Eetu Mäkelä & Jukka Suomela



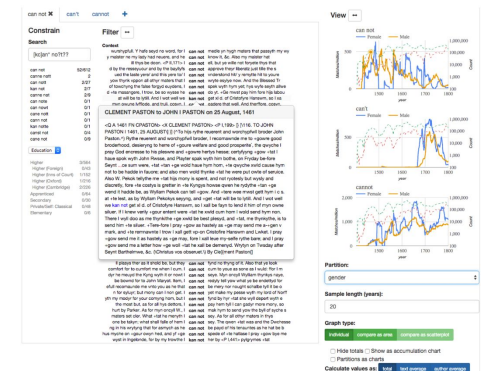
UNIVERSITY OF HELSINKI



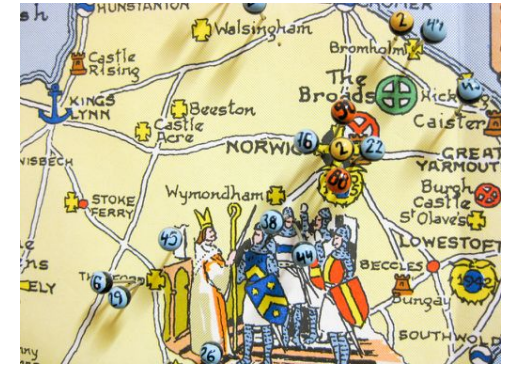
Aalto University

STRATAS project

- *Interfacing structured and unstructured data in sociolinguistic research on language change* (Academy of Finland, DIGIHUM, 2016–2019)
- Our subproject: **Social embedding of neologisms in early English correspondence**
 - Previous research: mostly lexicographical data (OED), bias towards well-known authors
 - *Corpora of Early English Correspondence* (CEEC): wide social spectrum, speech-like genre



CEEC



- **Personal letters, c. 1400–1800**
 - 1,180 writers, 11,713 letters, 5.2 million words
 - Compiled for historical sociolinguistics: **metadata** on letters, writers, recipients (e.g. gender, social rank)
- Compiled by T. Nevalainen, H. Raumolin-Brunberg et al. at the University of Helsinki
 - Based on published editions of letters
- Published versions:
 - *Sampler*, 1998 (0.5 Mw, 1418–1680)
 - *Parsed*, 2006 (2 Mw, 1410–1681)

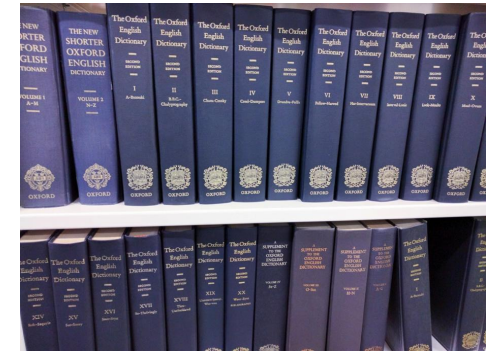
Research questions

1. **Who** are the innovators?
Which social groups do they represent?
2. **How do the new words spread** socially, geographically and diachronically?
3. **Which semantic domains** do the neologisms represent?
4. **Why** are the neologisms created and established?
Can they be linked to:
 - Specific historical events?
 - Changes in culture & society?
 - Social meanings?



Challenges

- **Identification** of neologisms
 - Spelling variation...
- Automatic retrieval of related **lexicographical** data
 - *Oxford English Dictionary*:
e.g. first attestation dates, etymologies
 - *Historical Thesaurus*: time-sensitive semantic domains
- **Interface** for pruning the possible neologisms found



Getting started: case studies

- Based on previous work: variation in the productivity of nominal suffixes *-ity*, *-er*
- **Material:** *Corpus of Early English Correspondence Extension*
 - Late C18 section, 1760–1800
- **FiCa** interface for classifying data, retrieving OED first attestation dates
 - Developed by Eetu Mäkelä

FiCa

roller (3)					RLR (3)	RLR (3)
roller	rolers	yes		2	RLR	RLR
roller	roller	yes		3	RLR	RLR
roller	rollers	yes		6	RLR	RLR
ruler					RLR	RLR
ruler	ruler	unclear	er1; 1 in ed	5	RLR	RLR
rem-ber					RMBR	RMBR
rem-ber	rem-ber	no		2	RMBR	RMBR
					RMLR (2)	RMLR (2)
rambler (2)					RMLR (2)	RMLR (2)
rambler	rambler	unclear	some NP0	7	RMLR	RMLR
rambler	ramblers	no	NP0 (perioc	1	RMLR	RMLR
remem[[b]er					RMMB	RMMBR
remem[[b]er	remem[[b]er	no		1	RMMB	RMMBR
					RMMR (8)	RMMR (7)
rememb=r=s					RMMR	RMMR
rememb=r=s	rememb=r=s	no		1	RMMR	RMMR
remember (2)					RMMR (2)	RMMR (2)
remember	rememb=r=	no		8	RMMR	RMMR
remember	remember	no		690	RMMR	RMMR
remember's					RMMR	RMMR
remember's	remember's	no		2	RMMR	RMMR
remembers					RMMR	RMMR
remembers	remembers	no		22	RMMR	RMMR
remembr					RMMR	RMMR
remembr	remembr	no		2	RMMR	RMMR
remmember					RMMR	RMMR
remmember	remmember	no		2	RMMR	RMMR
remembrancer					RMMR	RMMRNSR
remembrancer	remembranc	yes	er1	1	RMMR	RMMRNSR
					RMNT (9)	RMNTR (9)
remainder (9)					RMNT (9)	RMNTR (9)
remainder	remainder	no	MO	1	RMNT	RMNTR

of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone **rollers:** this he says is a late invention; but as it requires more workmanship than is easy to be had here I think cast iron rol

odel he has brought with him from England there is an apparatus for bruising the seed by making it pass between two iron **rollers** of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone

<Q A 1783 FN SBENTHAM> <X SAMUEL BENTHAM> <P III,209> [] [469 FROM SAMUEL BENTHAM]] [^TO JEREMY BENTHAM^] [ADDRESSED:] Jere=y= Bentham Esq=r= / Lincoln's Inn / London Petersbourg Sepr. 13th O.S. 1783. I am at length taken into the service of this country. The rank given me is that of (Conseiller de la Cour) , which is only equal to that of Lieutenant Colonel in the army. Considering that I had had no Military rank in any other country to found my pretensions on, ...^ to bruise the seed in oil mills, and what objections he sees to such a substitute. In many places I imagine the rollers of stone may be cheaper; but that would not be the case here. Capper is returned and his having seen you all has attached me to him. It is probable we may be concerned together in the erection of an oil mill which gave rise to the above question. In a model he has brought with him from England there is an apparatus for bruising the seed by making it pass between two iron rollers of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone

OED Oxford English Dictionary Quick search: Find word in dictionary GO

Lost for Words? | Advanced search | Help

Help on Search Results | Print | Email

Quick search results

Showing 1-5 of 5 results in 5 entries

Widen search? Find 'roller' in: » phrases (186) » definitions (290) » etymologies (49) » quotations (1600) » full text (751)

View as: List | Timeline Sort by: Entry | Frequency | Date

- roller, n.¹** View full entry 1295

...One of a number of (usually large) cylinders of wood or other hard material, sometimes attached to a framework, over which a heavy object can be passed....
- roller, n.²** View full entry 1678

...A jay-like bird, *Coracias garrulus* (family *Coraciidae*), having mainly greenish-blue plumage with dark blue wings and a chestnut back, noted for its characteristic tumbling display flight and found...

Your current search (entries):

- roller

Save Refine search

Refine your search

- Subject
- Language of Origin
- Region
- Usage
- Part of Speech
- Date of First Citation

My entries (2)

My searches (6)

Jump to:

Entry

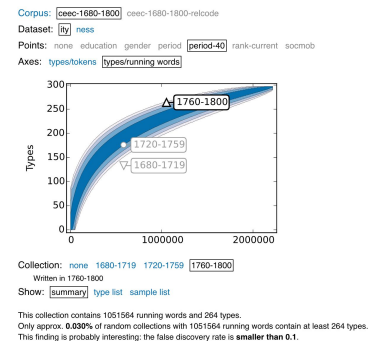
- subduement, n.
- subduer, n.
- subduing, n.
- subduing, adj.
- subduple, adj.
- subduplicate, adj.
- subdural, adj.
- subdwarf, n.
- sub-echo, n.
- subedar, n.
- subedari, n.
- sub-edit, v.
- sub-editor, n.
- sub-editorial, adj.

Case 1: *-ity*

*Corpora of Early English Correspondence,
1760–1800*

Introduction

- *-ness* & *-ity*: derive nouns from adjectives
 - e.g. *generous* → *generousness* / *generosity*
 - *-ness* native, ***-ity* borrowed from French & Latin**
 - connotations associated with those languages (e.g. polite society, learned, scientific)
- Previous research (Säily 2014): **productivity** of *-ity* increases throughout C17–18, perhaps led by male professionals
- End of C18: *-ity* highly productive – what kinds of neologisms by whom?



15 neologisms found, 1760–1800

- Working definition of neologism: corpus attestation max 100 years after OED first attestation date
 - 3 antedatings, 3 not in OED at all
- *authenticity*, *cleverality* (1778<1828), *comicality*, *conviviality* (1783<1791), *coxcombicality*, *foxity* (1788), *Germanity* (1788<1821), *impracticability*, *intrepidity*, *irritability*, *oddity*, *respectability*, *ridiculability* (1776), *scoundrellity* (1761), *versatility*

Who are the innovators?

- **Thomas Twining**, clergyman, scholar (c.1734–1804)
 - *coxcombicality*, *foxity*, *Germanity* 1788, *ridiculability* 1776
- **Charles Burney**, musician, author (1726–1814)
 - *conviviality* 1783, *versatility* 1782
 - Daughter **Fanny Burney** (author, 1752–1840) uses *conviviality* 1799
- **David Garrick**, actor, playwright (1717–1779)
 - *comicality* 1767, *scoundrellity* 1761
- **Ignatius Sancho**, author (1729?–1780)
 - *cleverality* 1778



How do they use the neologisms?

... there is not among any set of people such a comfortable scratch-back confederacy as among those old ruin-diggers [antiquarians & old-English grubbers]. Is it the consciousness of enemies & scoffers all round them – the sour'd feeling of **ridiculability** – that draws the knot closer and forms the phalanx, back to back, for mutual scrubbing & defence?

Thomas Twining to Charles Burney, 1776

It is a most infamous design, & I desir'd Churchill would Let Thornton know of it, which he will do immediatly, & prevent their **Scoundrillity** by some humourous Paragraph ...

David Garrick to George Colman, 1761

How do neologisms spread?

I have always thought, that in many particulars his equal was not to be found – his wit, learning, taste, penetration; &, when well, his **conviviality**, pleasantry, – & kindness of heart to me & mine, will ever be thought of, with the most profound & desponding regret!

Charles Burney to Fanny Burney, 1783?

The account of the Play meriting little attention indeed – I am much pleased at your independent establishment of **conviviality** at Burlington House.

Fanny Burney to Charles Burney, 1799

-ity: answers to RQs

1. **Innovators**: creative men in their 40s–50s, professionals / other non-gentry
2. Neologisms **spread** in a social network of peers writing in a similar style
3. They often describe **human attributes**
4. They also carry **social meaning**: designed to amuse / impress recipient, emphasise in-group membership

Case 2: *-er*

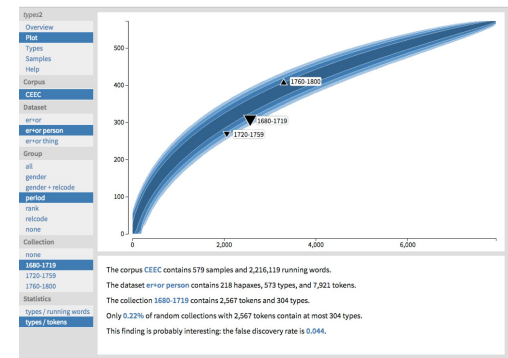
Corpora of Early English Correspondence,
1760–1800

Nominal suffixes *-er* and *-or*

- Typically derive **agentive** or instrumental nouns from verbs: *driver, governor, filler*
 - Also: person ‘concerned with N’ / ‘living in N’
 - *-or*: Latinate variant of *-er*
 - Pronounced identically (*adviser/advisor*)
 - treated as a single suffix here
- (Plag 2003: 89; Bauer 2001: 199–203)

Previous research (CEEC, C18)

- **Productivity** of *-er* increases over time
 - Stylistic change or continued semantic expansion? (Säily 2014, *-ity*; Dalton-Puffer 1994)
- Men writing to their close friends overuse *-er*
 - Less stable relationship a trigger for productivity? (cf. Wolfson 1990; Säily 2014, *-ity*)
- End of C18: *-er* highly productive – what kinds of neologisms by whom?



25 neologisms found, 1760–1800

- **Describing people:** *absconder*, *blubberer* (1782<1786), *commemorator* (1784<1856), *completer*, *complimenter*, *dangler*, *outsider* (1800=OED), *schemer*, *seceder*, *spiter* (1790<1847), *swindler*
 - **Occupations:** (*shirt-*)*airer*, *gambler*, *hairdresser*, (*China-*)*piecer*, *smuggler*
- **Connected to places:** *Chiswicker*, *Madrasser*, *Norfolker*, *Turnham-Greener* (**none in OED**)
- **Things:** *cutter* ‘boat’, *ventilator*
- **Other:** *brightener*, *plumper* ‘lie’ (1776=OED), *winterer* (1784<1795)

Who are the innovators?

- **Thomas Twining**, clergyman, classical scholar (c.1734–1804)
 - *blubberer* 1782, *complimenter* 1788, *commemorator* 1784, *plumper* 1776
- **Hester Piozzi**, writer (1741–1821)
 - *seceder*, *spiter* 1790
- **Jane Austen**, novelist (1775–1817)
 - *outsider* 1800
- **George Culley**, agriculturist (c.1735–1813)
 - *Norfolker*, *winterer* 1784



How do they use the neologisms?

As to myself, *Cecilia* has done just what she pleas'd with me: I laughed, & cried (for I am one of the **blubberers**) when she bade me.

Thomas Twining to Charles Burney, 1782

it makes me laugh when I think how the **Spiters** told us that *Siddons had lost all her Popularity*

...

Hester Piozzi to Charlotte Lewis, 1790

How do neologisms spread?

- All of the *-er* antedatings are *hapax legomena* in the corpus → no hard evidence
- However, all of them (like *-ity*) occur in letters written to **close friends or family**
 - Spread in the social networks of innovators?
 - Not all may have spread (depending on status of innovator in the network?), could have been re coined at a later date

-er: answers to RQs

1. **Innovators:** men *and women*, mostly in their 40s–50s, professional / other non-gentry
2. Neologisms **spread** in a social network of peers writing in a similar style
3. They often describe **people**
4. They also carry **social meaning**: designed to amuse recipient, emphasise in-group membership
 - Some reflect *innovations in society/(agri)culture*

Conclusion

Case studies



- Late C18 neologisms mostly produced by the emerging **middle class**, reflect **involved style** of letter-writing (affective, interpersonal)
 - *ridiculability, spiter*
- Most innovative social groups **vary by affix**
 - *-ity*: men in creative professions, *-er*: both men and women, professionals and other non-gentry
- **Purpose** of innovation may vary by affix
 - *-er* also used for societal innovations

DynLex in a sociohistorical corpus

- Following the spread of **individual words across individual people** may prove impossible
 - CEEC doesn't include complete correspondences
 - Sparse data: 5 million words over 400 years
- What to do instead?
 - Supplement with other datasets?
 - Track neologisms at the level of **semantic domains** (Alexander et al. 2015) and **social groups**?
 - Do neologisms in a particular semantic domain originate with a specific social group, spread to others?

Beyond one-off case studies

- Semi-automatically filter CEEC corpus words to neologisms
- Explore word spread and social factors interactively

roller (2)				FLR (2)	FLR (2)
roller	rollers	yes		FLR	FLR
roller	roller	yes		FLR	FLR
roller	rollers	yes		FLR	FLR
roller	roller	unclear	er1; 1 in od	FLR	FLR
rem-bar	rem-bar	no		RMNR	RMNR
RMNR (2)					
RMNR (2)					
number (2)				FLR (2)	FLR (2)
number	numbers	unclear	some NPD	FLR	FLR
number	numbers	no	NPD (peric)	FLR	FLR
number[de]	number[de]	no		RMNR	RMNR
RMNR (8)					
RMNR (7)					
RMNR (7)					
remember-ns	remember[ns]	no		RMNR	RMNR
RMNR (2)					
remember	remember	no		RMNR	RMNR
remember	remember	no	690	RMNR	RMNR
remember's	remember's	no		RMNR	RMNR
remember's	remember's	no	22	RMNR	RMNR
remember	remember	no		RMNR	RMNR
remember	remember	no		RMNR	RMNR
rememberance	rememberance	yes	er1	RMNR	RMNRNSR
RMNR (9)					
RMNR (9)					
RMNR (9)					
RMNR (9)					

of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone rollers. This he says is a late invention; but as it requires more workmanship than is easy to be had here I think cast iron rollers made of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone rollers he has brought with him from England there is an apparatus for crushing the seed by making it pass between two iron rollers.

of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone rollers. This he says is a late invention; but as it requires more workmanship than is easy to be had here I think cast iron rollers made of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone rollers he has brought with him from England there is an apparatus for crushing the seed by making it pass between two iron rollers.

of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone rollers. This he says is a late invention; but as it requires more workmanship than is easy to be had here I think cast iron rollers made of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone rollers he has brought with him from England there is an apparatus for crushing the seed by making it pass between two iron rollers.

OED Oxford English Dictionary Quick search

Quick search results

Showing 22 of 22 results by gender

Wide search: find roller in: - phrases (266) - definitions (266) - etymology (69) - quotations (566) - full text (75)

View as: List | Timeline

Sort by: Entry | Frequency | Date

1. roller, n. View full entry

One of a number of usually light cylinders of wood or other hard material, sometimes attached to a framework, over which a heavy object can be passed.

2. roller, n. View full entry

A jet-like bird, *Chordeiles germicola* (family Corvidae), having mainly greenish-blue plumage with dark blue wings and a chestnut back, named for its characteristic rolling display flight and song.

can not ✕ can't cannot +

Constrain Search [kɔːlən̩ ˈnɔʊ???

Filter

Contast

can not 52/12
can not 2
can not 2/27
can not 2/7
can not 2/9
can not 0/1
can not 0/1
can not 0/1
can not 0/1
can not 0/4
can not 0/9

Education

Higher 3/584
Higher (Foreign) 0/43
Higher (Iris of Court) 1/152
Higher (Oxford) 1/216
Higher (Cambridge) 2/226
Apprenticed 0/64
Secondary 0/50
Private/Self: Classical 0/48
Elementary 0/6

CLEMENT PASTON to JOHN I PASTON on 25 August, 1461

<O A 1461 FN CPASTON>-<X CLEMENT PASTON>-<P I 199> [0] [116. TO JOHN PASTON I 1461, 25 AUGUST[1]] (*To his rythe reuerent and worchypful broder John Paston.*) Rythe reuerent and worchypful broder, I recomaunde me to +gouore good brotherhood, desyryng to herre of +gouore welfare and good prosperie, the wyche I pray God encrease to his plesure and +gouore hertys hessas; certyffynng +gou- +tat I have spok wyth John Rouse, and Playter spok wyth him bothe, on Fryday be-fore Seynt ...ce sum were, +tat +at +ge wold haue hym hom, +te wyche wold cause hym not to be hadde in fauoure; and also men wold thynke +tat he wete pur owte of seruice. Also W. Pekok tellythe me +tat hijs mony is spent, and not rotyetly but wyssly and discretly, +te costys is gretter in +te Kyngys howse qwen he rydythe +tan +ge wend it hadde be, as Wyllam Pekok can tell +gou- And +ere were most gett hym i e s all +he best, as by Wyllam Pekokys seyng, and +ge +tat will be to hyll. And I wot well we kan not get xl d. of Cristofyre Hanssum, so I xall be fayne to lend it myn owne siluer. If I knew verily +gouur entent were +tat he wold cum hom I wold send hym non. There I wyl doo as me thynkthe +ge wold be best pleyd, and +tat, me thynkthe, is to send him +te siluer. +tere-fore I pray +gou as hastily as +ge may send me a+open v mark, and +te remonaunte I towl xall gett vp-on Cristofre Hanssum and Lwket. I pray +gou send me it as hastily as +ge may, fore I xall leue ny-selfe rythe bare; and I pray +gou send me a letter w-ge wold +tat he xall be demerit. Wryfyn on Twesday after Seynt Barthelme, &c. (Christus vs obseruet.) By Cie[ment Paston]

If plays the ar as it should be, but they comfort for to comfort me when I cum. I don't meuyd the Kyng wyth it or nowt I be bound for to John Meryk. I can't rekommende me vnto you as he that n is forky; but myn can I non gett ym my moody for your coming hom, but I the most but, as for all ny detors, I hurt by Parker. As for myn onryll W., I maters set cler. What +tat he mayen I one be takyn; what shall falle of hem I ng in his wryng that for sump as he hus myche on +gouur oynn hed, and yf +ge wyst in Ingelonde, for by my trouthe I can not fynd no thyng of it. Also that ye look cumt to youe as some as I wold: for I don't seye. Myn onryll Wyllam thynkys rays, nelytly tall you what ye be ended for mer ny rougt schabbe tyll it be o yet make my pesse wyth my lord of Norff ym my moody for your coming hom, but I fynd by hyr +tat she wyl depert wyth e pay hem tyll I can gadre more mony, so mak hym to send you the byll of syche s say. As for all other maters in thys sey. The ouer +tat was and the Duchesse be payd of his fortautes as he nat be spe of +te hattase I pray +gou by me her by <P L441> pylgrimes +tat

View

can not

can't

cannot

Partition: gender

Sample length (years): 20

Graph type: Individual compare as area compare as scatterplot

Hide totals Show as accumulation chart

Partitions as charts

Calculate values as: total text average author average

Discovering -ER in CEEC

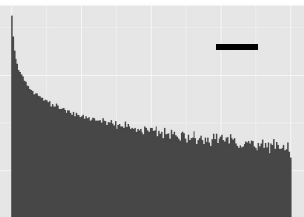
- -er(e), -ar(e), -or(e), -our(e), -owr(e), -ur(e), -r + plural, possessive...
- `\S*([rR]| [eEoO]~)(=?|=?[eE]=?| [= '~]*[eEilyY] ?['= '~]*[sSzZ]['=]*)?(?![a-zA-Z '~ = +])`
- 6800 candidate types, 300 000 appearances
- 5080 types out of 6800 irrelevant after manual study
- 153 types out of 6800 needed further study
 - 11768 individual uses

Semi-automatically filtering CEEC corpus words to neologisms

- Filter candidates by first attestations in OED, EEBO(-TCP), ECCO, BURNEY, BLN, TDA
- Problems: spelling variation, OCR errors in comparison corpora
- Solutions:
 - match based on multiple algorithms (phonetic, lemma, ...)
 - calculate match confidence (e.g. factor in total word frequency in all corpora)
 - use grouping and sorting functionalities of FiCa to efficiently verify candidates and spot errors

Semi-automatically filtering CEEC corpus words to neologisms

- If still too many candidates, create subcorpora of more manageable sizes
 - by dates
 - by social classes of interest
 - by later behaviour of the words in the combined corpora (e.g. only those that do appear later?)
- CEEC types in BLN+BURNEY+ECCO
 - ~58 000 types have less than 1000 appearances
 - ~36 000 types have less than 100 appearances
 - ~17 000 less than 10, ~2 500 appear once, ~6 700 don't appear at all



References

- Alexander, M., F. Dallachy, S. Piao, A. Baron & P. Rayson 2015. Metaphor, popular science and semantic tagging: Distant reading with the Historical Thesaurus of English. *Digital Scholarship in the Humanities* 30(suppl_1): i16–i27.
- Bauer, L. 2001. *Morphological Productivity*. Cambridge: CUP.
- CEEC = *Corpora of Early English Correspondence*. Compiled by T. Nevalainen, H. Raumolin-Brunberg et al. at the University of Helsinki.
<http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/>
- Dalton-Puffer, C. 1994. Are Shakespeare's agent nouns different from Chaucer's? – On the dynamics of a derivational sub-system. D. Kastovsky (ed.), *Studies in Early Modern English*, 45–58. Berlin: Mouton de Gruyter.
- Mäkelä, E., T. Säily & T. Nevalainen 2016. Developing an interface for historical sociolinguistics. Paper presented at DHC 2016, Sheffield, September 8–10.
- OED = *Oxford English Dictionary*. OED Online. Oxford University Press.
<http://www.oed.com>
- Plag, I. 2003. *Word-Formation in English*. Cambridge: CUP.
- Säily, T. 2014. *Sociolinguistic Variation in English Derivational Productivity: Studies and Methods in Diachronic Corpus Linguistics*. Helsinki: Société Néophilologique.
- Suomela, J. 2016. *types2*: Type and hapax accumulation curves.
<http://users.ics.aalto.fi/suomela/types2/>
- Wolfson, N. 1990. The bulge: A theory of speech behavior and social distance. *Penn Working Papers in Educational Linguistics* 2(1): 55–83.