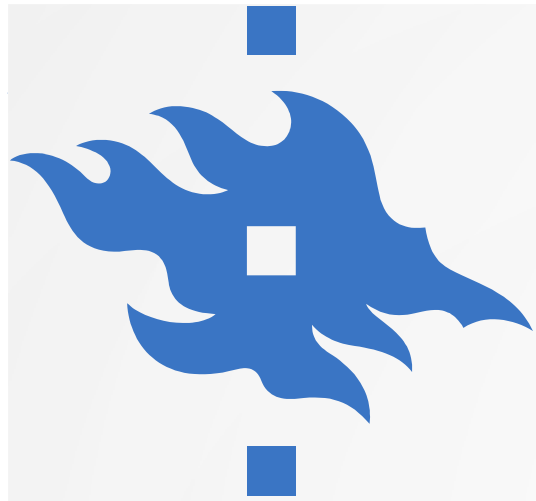


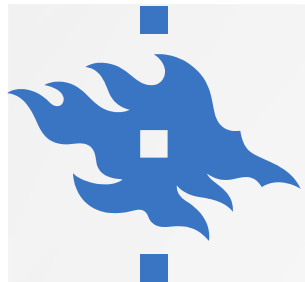


NEOLOGISMS IN EARLY ENGLISH LETTERS: HOW TO FIND THEM AND WHAT THEY CAN REVEAL

Tanja Säily, Eetu Mäkelä & Mika Hämäläinen

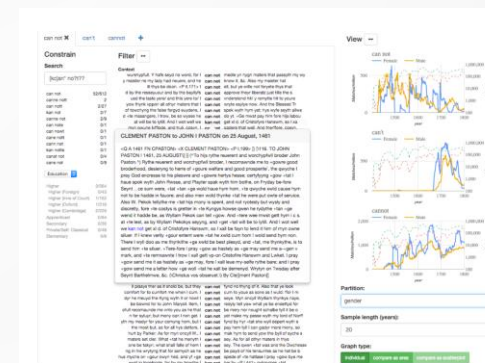


INTRODUCTION



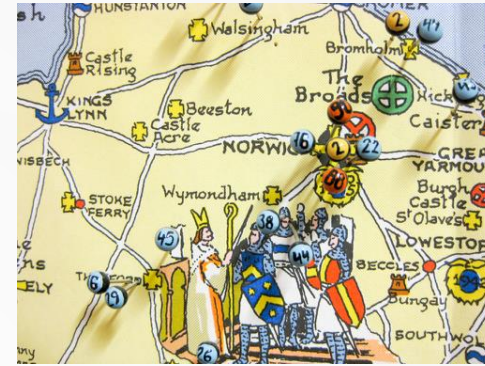
STRATAS PROJECT

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



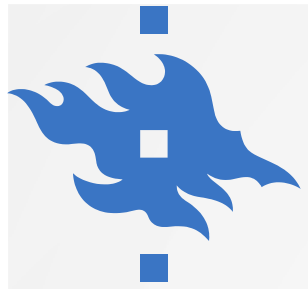


CEEC



- 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
- SCEEC = Standardized-spelling version using VARD2 (excluding 15th century)

www.helsinki.fi/varieng/CoRD/corpora/CEEC/



RESEARCH QUESTIONS

1. **Who** are the innovators/early adopters? 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?

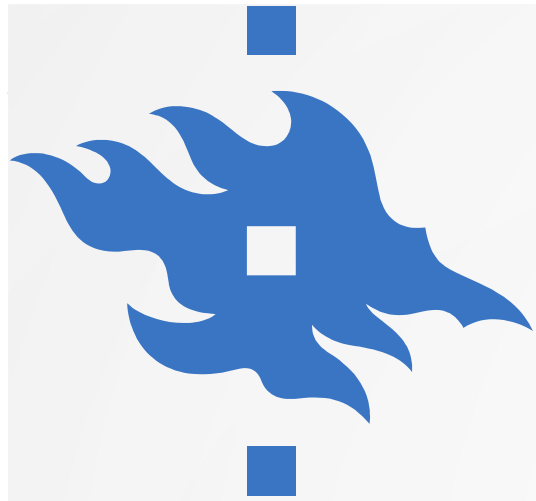




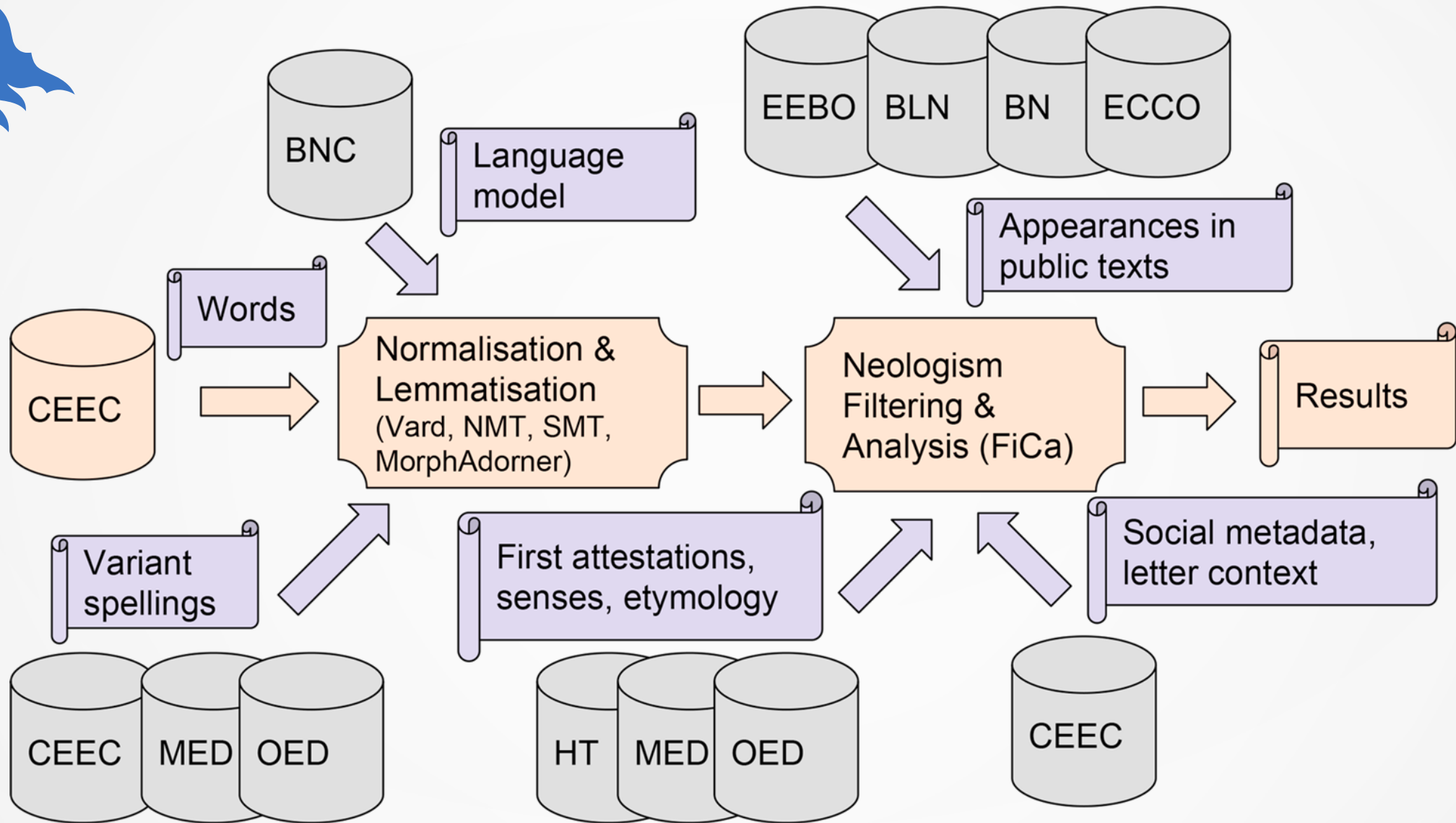
BIG-DATA APPROACH TO ANALYSING NEOLOGISMS

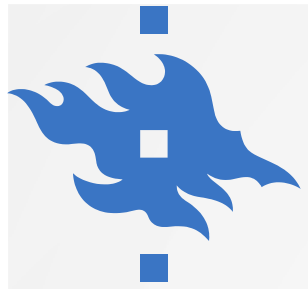


- Automatically map each word in the corpus to lexicographical data and contemporary published texts, compare first attestation dates
 - Spelling variation: SCEECS not enough, additional **normalization** required
- Automatic retrieval of related **lexicographical data**
 - *Oxford English Dictionary* (OED), *Historical Thesaurus* (HT), *Middle English Dictionary* (MED)
- Automatic retrieval of data from **databases of contemporary published texts**
 - *Early English Books Online* (EEBO), *Eighteenth Century Collections Online* (ECCO), *British Library Newspapers* (BLN), *Burney & Nichols Collections* (BN)
- **Interface** for pruning the possible neologisms found, exploring social factors



CURRENT PIPELINE





INITIAL MAPPING TO OED

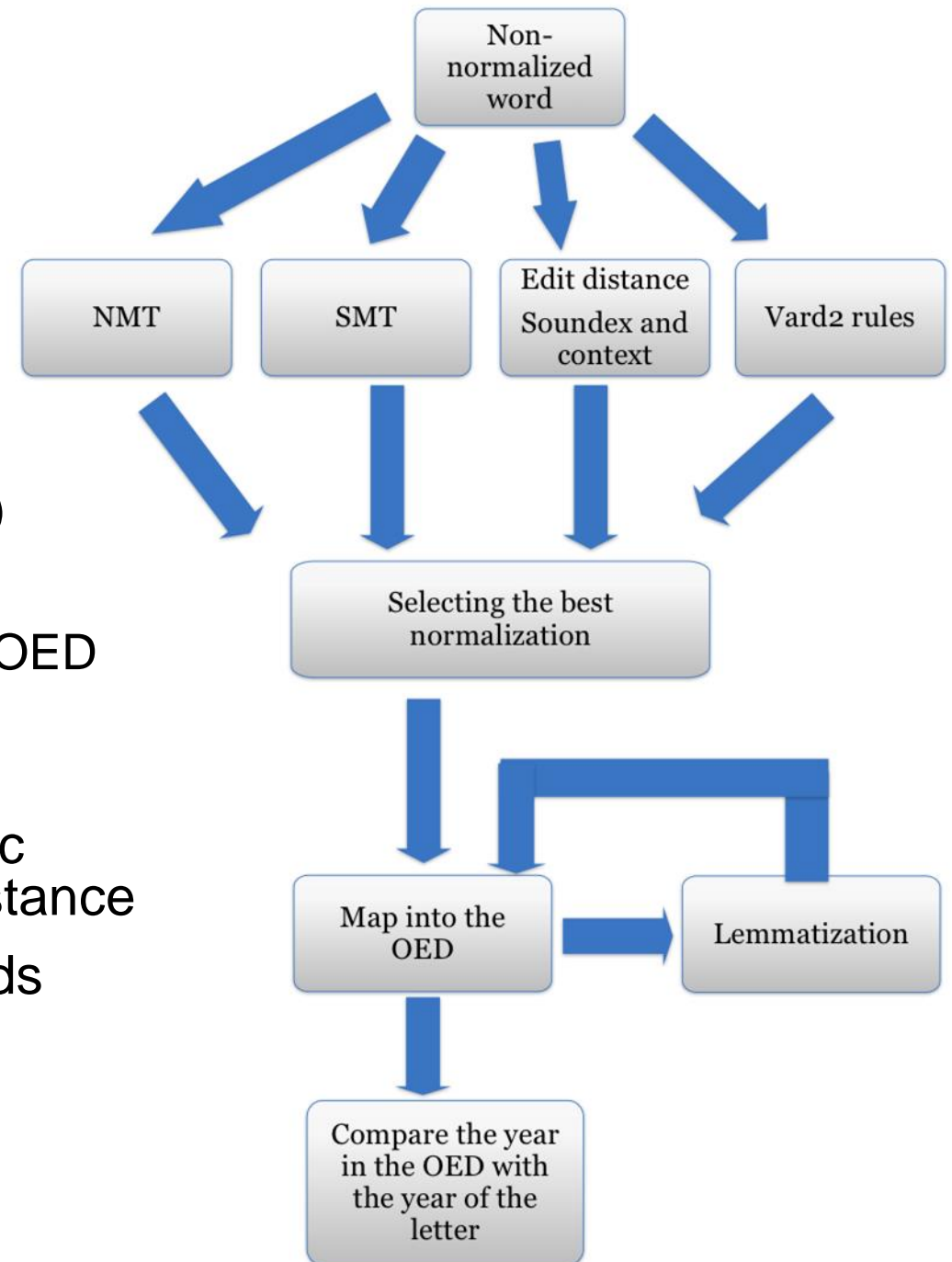
- **Prepare corpus:** convert to Unicode, remove most punctuation, tokenize
- Attempt to **lemmatize** with NLTK (based on Princeton WordNet)
- **Map** lemmas to OED (local JSON version)
- Extend already performed VARD2 normalizations to 15th century
- Use MorphAdorner to automatically **normalize further**
- **Map** again
- Successfully mapped: c. 50,000 word forms, unmapped: c. **100,000**



ADDITIONAL NORMALIZATION

(HÄMÄLÄINEN ET AL. 2018)

- Idea: use machine translation!
 - **NMT** = neural machine translation (OpenNMT)
 - **SMT** = statistical machine translation (Moses)
 - Use known VARD2 normalizations, MED and OED as input
 - Character-based; language model = BNC
- Levenshtein **edit distance** + filter by semantic similarity + Soundex pronunciation by edit distance
- Extend **VARD2** normalization rules to all words
 - 58 rules, e.g. “u → v anywhere”

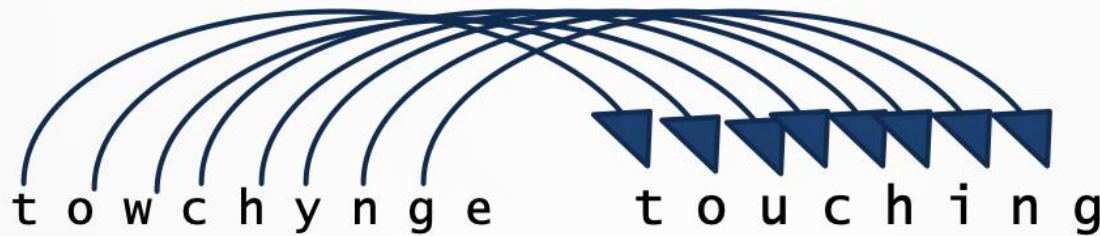




REVISITING NMT FOR NORMALIZATION

(HÄMÄLÄINEN ET AL. 2019)

- Changing the neural network to bi-directional recurrent neural network
- Looking at the top 10 predictions with OED and Spacy



	15th	16th	17th	18th	generic
OED +Lemma	69%	78%	71%	50%	61%
Lemma	61%	67%	63%	45%	53%



INTERFACE FOR FILTERING THE NEOLOGISM CANDIDATES

- FiCa (Filtering and Categorization)
- Developed by Eetu Mäkelä; see Säily et al. (2018)

OED Lemma▼	Word	Category	Notes (62)	Earliest letter	OED - CEEC	Total DF before	Total TF before	Frequency
acharya	acharya	yes		1789	3	0	0	1
anthroponomical	anthropi	yes		1734	0	0	0	1
anti-democrat	anti-den	yes	different	1799	3	0	0	1
blueism	bluism	yes		1795	0	0	0	1
bonneted	bonnete	yes	even if V	1781	43	17	17	1
canicule	canicule	yes		1701	0	5	5	1
cardiphonia	cardiphn	yes	actual 1s	1780	1	2	2	1
cast-off	cast-off	yes		1692	48	71	73	2
catchy	catchy	yes		1784	20	35	35	1
chaplaincy	chaplair	yes		1741	4	0	0	1
cleverality	cleverali	yes		1778	50	0	0	1
curtainless	curtainle	yes		1799	23	0	0	1
delineator	delineat	yes		1736	38	28	28	1
dicky-bird	dicky-bi	yes	figurative	1778	3	0	0	1
double-bedded	double-l	yes	double-t	1798	0	7	7	1
double-cross	double-ri	yes	different	1754	80	0	0	1
embodiment	embodii	yes		1777	51	2	2	1
envoyship	envoyst	yes		1706	30	0	0	1
eschantillon	eschant	yes		1717	3	5	5	1
escritoire	escritoir	yes		1694	13	3	3	1
freshen	fresheni	yes		1680	17	1	1	1
fussy	fussy	yes		1797	34	72	72	1

your own dinmonds &c must be kept. Barmton. Mr Collings prize	catchy	to look at. It looks like Mr Wastles judgement weakening Mr Collings
tup is not very capital to handle, but rather		says he has let one to George Lax &c better
<div> <div> <div> <div> <div><Q A 1784 FN GCULLEY> <X GEORGE CULLEY> <P 179> [] [1.1.]]]</div> <div>[^GEORGE CULLEY TO MATTHEW CULLEY^] Durham 1st October 1784 Mr Gill had so much to say about his tour into the west that I could not get from him last night. He has a nephew a son of Joseph, not of (^Arramathea^) but Shildon that he wants to put to us, to work and learn farming, we to find him victuals only, Mr Gill clothes, I told him I could say nothing as I had no house, and was rather cold about it, and said he wo... along. He says he does not kill half the large sheep he used to at this season and his neighbours the same. Bought the other day of the Cook &c near 400 small sheep and near 30 cattle. I think you must buy Mr John Robsons wethers, and may enquire about Mr Wilsons of Eslington. I now think dear as sheep are they will pay till Christmas, but am not fond of keeping all till spring, your own dinmonds &c must be kept. Barmton. Mr Collings prize tup is not very capital to handle, but rather</div> <div>catchy</div> <div>to look at. It looks like Mr Wastles judgement weakening Mr Collings</div> </div> </div> </div> </div>		
They havesabovi eT Sen tear		
highly commen		
Guy, of Duke-dtreet, York-buildings, Sur;	catchy	i geon, deceafed, begs leave to inform the pullie, that he has ,le, chi purchafed of Mr.
NTbhrfdar nert, the 12th Itflant, the O cleb x ted		
	CATCHY!	
.* But they are All of a piece; st yet they lye upon the	Catchy	to Trip up the Heels one of Another.* Prethee wilt thou make these Things Hang Together, now.
:rpondenta in' moll of the great towns in the ldnngdm, thore	Catchy	fupplied ;; -for if th y- lhotild have any doubts with refpefl to the fignatureori rbe padke'ts-t laey ma-y be: able to
.vwho live in the cuntry may be		

OED

Oxford English Dictionary

Quick search:

Find word in dictionary

GO

Lost for Words? | Advanced search | Help

Browse:

» Dictionary

» Sources

» Categories

» Historical Thesaurus

» Timelines

My entries (13)

My searches (15)

Jump to:

Entry ▼

catchiness, n.

catching, n.

catching, adj.

catchment, n.

View full results for 'catchy'

Help on Dictionary Entry | Print | Save | Email | Cite

catchy, adj.

Text size: A A

View as: Outline | Full entry

Quotations: Show all | Hide all

Keywords: On | Off

Pronunciation:

Brit. /ˈkætʃi/, U.S. /ˈkɛtʃi/

Etymology:

< CATCH v. + -Y suffix¹.

colloq.

1. Adapted to catch the attention or fancy; attractive, 'taking'.

Thesaurus »

Categories »

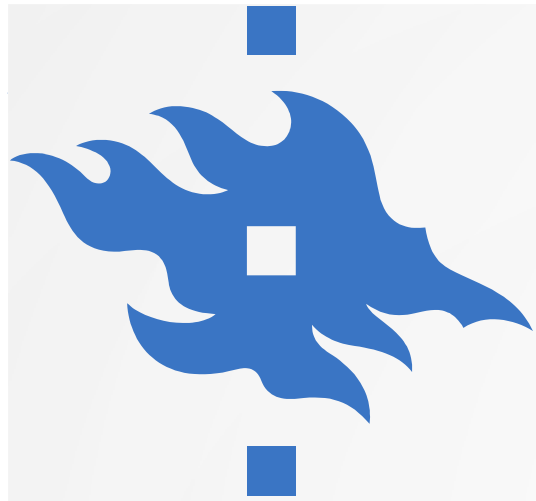
This entry has not yet been fully updated (first published 1889).

Entry history

Entry profile

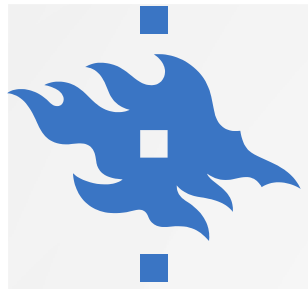
Previous version: OED2 (1989)

In other



PILOT STUDY

Social aspects of new 17th-century lexis



NEW 17TH-CENTURY LEXIS IN THE CEEC

- CEEC, socially representative sample from 1640–1660 (**Civil War** period)
 - 54 writers, 107 letters, 36,265 words
- Criteria for new vocabulary:
 - CEEC 1st attestation max 40 years after OED 1st attestation
- Aim also to test pipeline → filtering done more manually than usual
 - 820 CEEC 1st attestations for human to filter in interface
- Results: **42 new words, 12 OED antedatings, 3 “actual” neologisms** (seem to antedate both OED and contemporary published texts):
 - *packet-boat* (1641<1642), *statement* (1642<1750), *tea* (1643<1655)

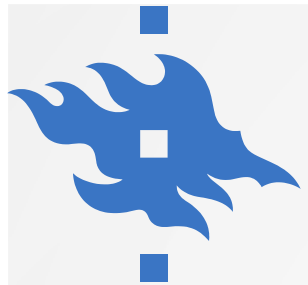


WHO PRODUCED THE 3 NEOLOGISMS?

- Two noblemen, father and son!
 - Royalists, Catholic leanings; art collectors
 - Travelled on the continent
- **Thomas Howard**, Earl of Arundel (1585–1646)
 - Courtier, politician
 - Wrote in his late 50s
- **William Howard**, Viscount Stafford (1612–1680)
 - Younger son, minor role in politics
 - Wrote in his early 30s



William Howard and Thomas Howard
(Wikimedia Commons)



EXAMPLES

just nowe Count Fabroni, & President
Cognewe are come unto me from Q:
Mother, to entreate very earnestly, that
the gentleman cominge alonge wth this
called Don Martino Dugaldi may
instantly passe to Dunkerke for her M^{ties}
especiall service w^{ch} depends soe much
upon it as upon his retorne or any
others sent before by y^e **Packette**
Boate ...

(Thomas Howard in Dover to Sir John
Pennington, 1641; OED3 1642)

I have receaved onely one letter in
which there is a **statement** that the
ssouldiers went to Mr John Penneducks
house at King berry and ransaked it
totally ...

(William Howard in Antwerp to Thomas
Howard, 1642; OED3 1750)

I have scarce bought any thinge for my
selfe but an Indian Brewhouse for **tee**,
which hath beene very good Black Lack
worke, but it is all spoyled and rased
and yett I payed exceeding deare for it.

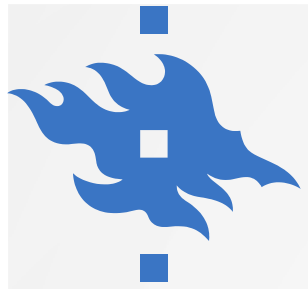
(William Howard in Amsterdam to his
mother, 1643; OED2 1655)



WHO ARE THE EARLY ADOPTERS?

- Surprisingly many new words compared to number of running words:
 - **Harrison** (4), parliamentarian army officer
 - **L'Estrange** (4), royalist politician
- Social networks:
 - Elizabeth & Charles Stuart
 - Charles Nichols & Henry Oxinden
 - **Anne Conway & Henry More**
 - William & Thomas Howard

Name	# of new words
Harrison, Thomas	4
L'Estrange, Hamon	4
Stuart, Elizabeth, Q. Bohemia	4
Conway, Anne	3
Cary, Anne	2
Fleming, Daniel	2
Howard, William	2
Nichols, Charles	2
Oxinden, Henry	2
Stuart, Charles I	2
More, Henry	1
...	1



EXAMPLES

To agree (as is alreadie) to act in dearest love expressed to him named Protector, (or Mount Sirion as the Sidonians called Hermon, and David in the spirit followed that faithfully, **believingly**, undoubtingly, **unanimously**, that He would retreat in action of undertaking (and soe witnes repentance by **condisention**) and wee would as willingly repent of o' sinfull dissentions)

...

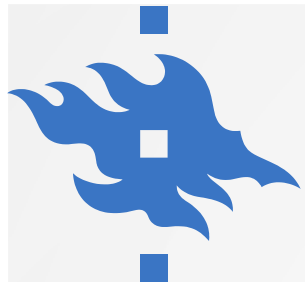
(Thomas Harrison to John Jones, 1656; OED3 1631, OED2 1642)



Thomas Harrison and Hamon L'Estrange

soone after I recd from Mr *Playford* himselfe a large and fervent letter wth profession of his skill and respects; wch letter (together wth his printed **Manifesto** or Publicacion thereinclosed) I presume to send unto you to peruse ...

(Hamon L'Estrange to Thomas Browne, 1653; OED3 1620)



EXAMPLES

but you must give me Leave to take notice of that part of yo^r Letter where you say y^t all coloures are formally in the soul and but **causally** in the obiects & organ, that green and white is not in yo^r carpett & papyr, but in yo^r mind ...

(Anne Conway to Henry More, 1651; OED2 1635)

- Philosophers, from pupil–tutor relationship to close friends



Anne Conway and
Henry More



for first the paper transmitts the colour of white w^{ch} is its own motion, and if it should transmitt the motion caused by any other obiect, then why does not everything we Looke upon yeeld the **eidolum** or representation of something else ...

(Anne Conway to Henry More, 1651; OED2 1619)



SOCIAL RANKS

- Surprisingly **few** new words compared to number of running words:
 - Upper clergy

I keep my self so close an anachoret,
and make so few **visits**, that I am not
likely suddainly to meet with Sir John
Dingly ...

(Brian Duppa to Justinian Isham, 1650;
OED 1626)

Social rank	# of new words
Clergy, lower	7
Genry, lower	7
Gentry, upper	7
Professional	7
Nobility	7
Royalty	6
Clergy, upper	1
Merchant	0
Other non-gentry	0



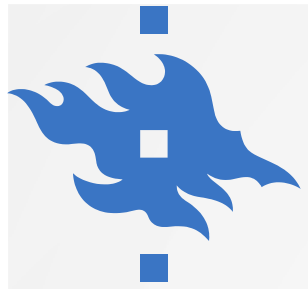
GENDER, AGE; REGISTER

- More data from men, more new words used by men
 - More advanced statistics needed
- **More** new words by **older** people (40–70)
 - Previous research on Dutch: highest lexical productivity among highly educated older men (Keune 2012)
- Surprisingly **many** new words in **letters to acquaintances and close friends**, fewer to nuclear family members
 - Consistent with “bulge theory” (Wolfson 1990); less stable relationship triggers more creative language use (cf. Säily 2018, *-ity*)



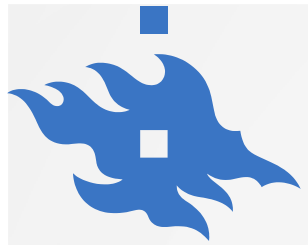
DISCUSSION

- New 17th-century vocabulary used by the **upper & middling ranks**
 - **!** 18th century: surprisingly many new words by the lower ranks
 - **Access** to education, specialized registers, new things & ideas; social networks
 - Lack of data from the lower ranks
- **Age, audience design** also factors
 - 18th-century results similar although more focused on close friends, not other acquaintances – changing functions of letter-writing?
- Novelty sometimes indicated in the text
 - Underlining, glossing, part of a list



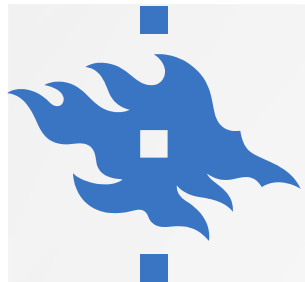
CONCLUSION

- **Big-data approach:** massive databases & automated pipeline → possible to quickly discover dozens of neologisms in millions of words of running text
 - Letters seem to be a great source for historical lexicography & lexicology
- What are we **missing**? Homonyms, zero derivation, MWUs, ...
 - Spelling variation still an issue, could disproportionately affect lower ranks
 - Actual 1st attestations? Innovators vs. early adopters
- **Future work:**
 - Extend analysis to entire time period
 - Ask more focused questions (e.g. who are the coiners of nonce-words, early adopters of vocabulary that comes into general use later)
 - Improve normalization & pipeline; integrate social analysis functionality into pipeline



REFERENCES

- Hämäläinen, Mika, Tanja Säily, Jack Rueter, Jörg Tiedemann & Eetu Mäkelä (2018). “Normalizing early English letters to Present-day English spelling”. Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Feldman, Anna Kazantseva, Nils Reiter & Stan Szpakowicz (eds.), *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (ACL Anthology W18-45), 87–96. Stroudsburg, PA: Association for Computational Linguistics.
- Hämäläinen, Mika, Tanja Säily, Jack Rueter, Jörg Tiedemann & Eetu Mäkelä (2019). “Revisiting NMT for normalization of early English letters”. Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Kazantseva, Nils Reiter & Stan Szpakowicz (eds.), *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (ACL Anthology W19-25), 71–75. Stroudsburg, PA: Association for Computational Linguistics.
- Keune, Karen (2012). *Explaining Register and Sociolinguistic Variation in the Lexicon: Corpus Studies on Dutch*. PhD dissertation, Radboud University Nijmegen.
- Säily, Tanja (2018). “Change or variation? Productivity of the suffixes *-ness* and *-ity*”. Terttu Nevalainen, Minna Palander-Collin & Tanja Säily (eds.), *Patterns of Change in 18th-century English: A Sociolinguistic Approach* (Advances in Historical Sociolinguistics 8), 197–218. Amsterdam: John Benjamins.
- Säily, Tanja, Eetu Mäkelä & Mika Hämäläinen (2018). “Explorations into the social contexts of neologism use in early English correspondence”. *Pragmatics & Cognition* 25(1). 30–49.
- Wolfson, Nessa (1990). “The bulge: a theory of speech behavior and social distance”. *Penn Working Papers in Educational Linguistics* 2(1). 55–83.



THANK YOU!

*acrimonious, believingly, candid, candour, causally, compensate, compliance, condescension, coney ground, congregational, **covenanting** (adj.), **crawling** (n.), dishearten, dragoon, **efficaciously**, eminently, **endeared**, entanglement, **helpfulness**, hint (v.), idolum, incendiary, incognito, initiatory, **joke**, **landgravine**, leveller (sense 2), malignancy, manifesto, **oversweetness**, **packet-boat**, **plenipotentiary**, remind, rickets, sequestrator, **statement**, Swede, **tea**, variously, vibrate (v.), visit (n.), voluminous*

Special thanks to Oxford University Press and the Middle English Compendium for sharing their lexicographical data (OED, HT; MED)

<https://github.com/mikahama/natas>
<https://github.com/jiemakel/fica>