

# NEW WORDS IN EARLY ENGLISH LETTERS: HOW TO FIND THEM AND WHAT THEY CAN REVEAL

Tanja Säily, Eetu Mäkelä & Mika Hämäläinen

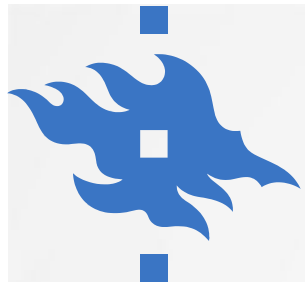


# NATAS PROJECT: SOCIAL EMBEDDING OF NEW WORDS

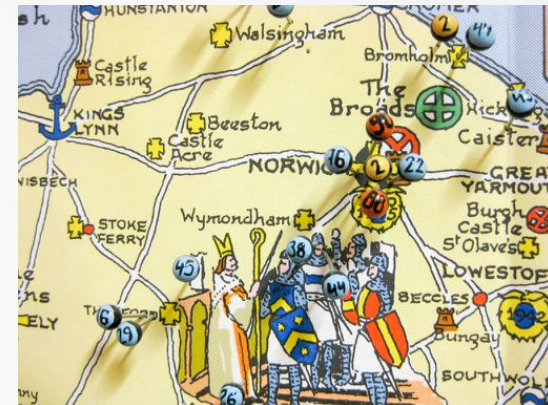


- Part of Academy of Finland funded STRATAS project, 2016–2019:  
Combining text and social information in historical sociolinguistics
  - [blogs.helsinki.fi/stratas-project/](https://blogs.helsinki.fi/stratas-project/)
- Who creates and adopts new vocabulary in the history of English?  
Differences across **social groups**?
  - Previous research focused on published writings by highly educated men
  - We take into account women, lower social ranks
  - Case studies of *-ness*, *-ity*, *-er*: sociolinguistic variation found in productivity (e.g. Säily & Suomela 2017, Säily et al. 2018)



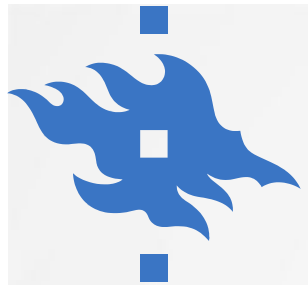


# CORPORA OF EARLY ENGLISH CORRESPONDENCE (CEECE)



- 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 15<sup>th</sup> century)

[www.helsinki.fi/varieng/CoRD/corpora/CEECE/](http://www.helsinki.fi/varieng/CoRD/corpora/CEECE/)



# BIG-DATA APPROACH TO ANALYSING NEOLOGISMS

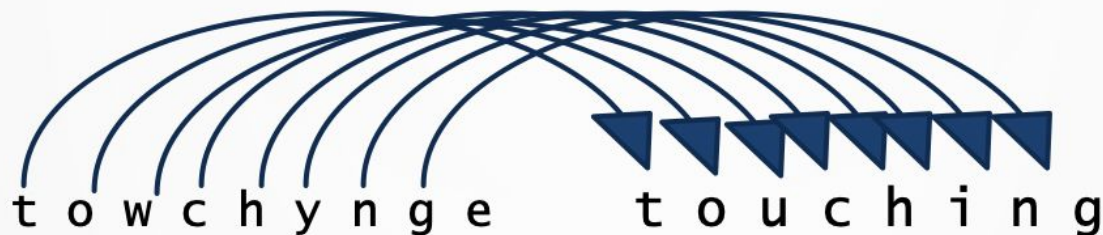


- Automatically map each word in the corpus to lexicographical data and contemporary published texts, compare first attestation dates
  - Spelling variation: SCEECS not enough, additional **normalization** required
- Automatic retrieval of related **lexicographical data**
  - *Oxford English Dictionary* (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



# REVISITING NMT FOR ADDITIONAL NORMALIZATION

- Neural Machine Translation: “translating” from historical to modern English at the level of **characters**
- **New:** change the neural network (Hämäläinen et al. 2018) to a bi-directional recurrent neural network (Hämäläinen et al. 2019)
- Focus on words found in the OED



	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	cardiph	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	figurativ	1778	3	0	0	1
double-bedded	double-b	yes	double-t	1798	0	7	7	1
double-cross	double-i	yes	different	1754	80	0	0	1
embodiment	embodii	yes		1777	51	2	2	1
envoyship	envoysh	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 diamonds &c must be kept. Barmton. Mr Collings prize  
tup is not very capital to handle, but rather **catchy** to look at. It looks like Mr Wastles judgement weakening Mr Collings  
says he has let one to George Lax &c better

<Q A 1784 FN GULLEY> <X GEORGE GULLEY> <P 179> [ ] [M.V.] [ ] ^GEORGE GULLEY TO MATTHEW GULLEY^> 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 Robbons 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 catchy to look at.** It looks like Mr Wastles judgement weakening Mr Collings says he has let one to George Lax &c better than him but not so pleasing to look at, he was gone so did not see him, let at 9 guineas also, Mr Colling will not be a marr trade. He is clear for letting at good prices or none. Mr Colling has some promising lambs, which I am glad off. Wheat a better crop here than I expected, oats a bad one, beans good and barley good, this last grain I am inclined to think has a chance to get...t to expect them untill they come. I was at Burdon. Mr Wastle poorly, Mrs Wastle very well. Mr Wastle says beef must be 2s 8d before March. Northallerton a pretty good shew of cattle, and not quite so high as have been. I send this by your friend Mr John Mason to Alnwick fair. Bob got here with the tups all well by 6 o'clock but I have not seen them. Mr Wastle says Mr Charge sent bad queys into Russia. Tell Nanny Brown I saw her brother Jack, who desired his love, and his wife is still poorly.

They have above it Sen tear  
highly commen

Guy, of Duke-street, 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 ltflant, the O cleb x ted

. \* 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.

:pondenta in' moll of the great towns in the Idngdnhm, there  
 .vwho live in the country may be

**Catchy**      fulfilled ;; -for if th y- lhotild have any doubts with refpeel to the  
 figurareori rbe padke'ts~t laey ma~y be: able to


**Oxford English Dictionary**  
*The definitive record of the English language*

**Quick search:**  **GO**

[Lost for Words?](#) | [Advanced search](#) | [Help](#)

[Browse:](#)

- Dictionary
- Categories
- Timelines

[Source](#) | [History](#) | [Thesaurus](#)

[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.  /'kætʃɪ/, U.S.  /'kætʃɪ/

**Etymology:** < CATCH *v.* + *-Y suffix*<sup>1</sup>.

*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

My entries (13)

My searches (15)

Jump to:

Entry ▾
 

- catchiness, *n.*
- catching, *n.*
- catching, *adj.*
- catchment, *n.*

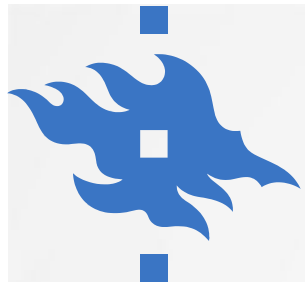


# NEW 17<sup>TH</sup>-CENTURY LEXIS IN THE CEEC

- CEEC, socially representative sample from 1640–1660 (**Civil War** period)
- 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)
- New vocabulary mainly used by the **upper & middling ranks**
  - **!** 18<sup>th</sup> 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







# EARLIEST KNOWN ATTESTATIONS OF THE ENGLISH WORDS FOR TEA

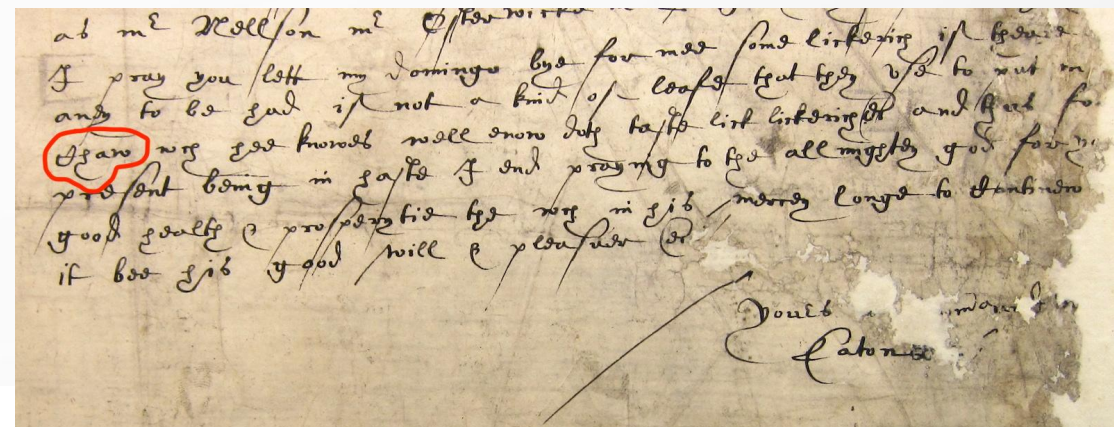


I have scarce bought any thinge for my selfe but an Indian Brew-house for **tee**, which hath beene very good Black Lack worke, but it is all spoyled and rased and yett I payed exceeding deare for it.

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

Mr Eaton, I pray you buy for me a pot of the best sort of **chaw** in Meaco, 2 farre [fair] bowes & arrowes, some half a dozen of Meaco guilt boxes, square, for to put in tobacco.

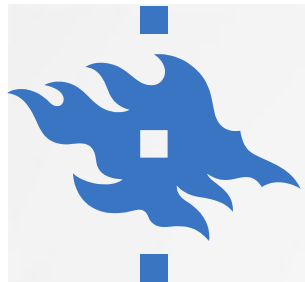
(Richard Wickham at Hirado to William Eaton at Kyoto, 1615; OED 1616)





# 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 (~60% accuracy in lemmatization. Errors evenly spread across ranks, but not according to sex or relationship)
  - Actual 1<sup>st</sup> 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



# 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>