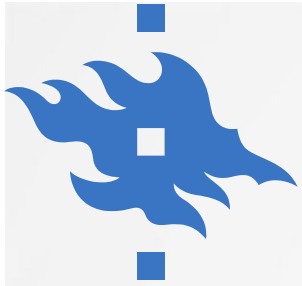


BETWEEN THE OED AND HISTORICAL TEXT COLLECTIONS:

THE CASE OF DISCOVERING NEW WORDS

Tanja Säily, Eetu Mäkelä & Terttu Nevalainen



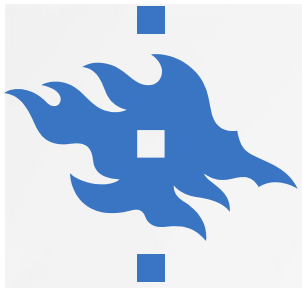


OUTLINE

VARIENG, HISTORICAL TEXT COLLECTIONS AND THE OED

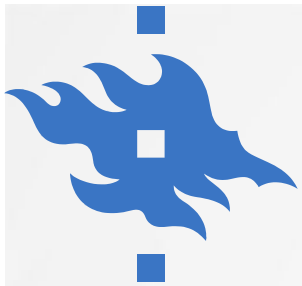
NATAS: SOCIAL ASPECTS OF **NEW WORDS** IN EARLY ENGLISH LETTERS

WORKFLOWS: BACK AND FORTH BETWEEN THE OED AND HISTORICAL TEXT COLLECTIONS



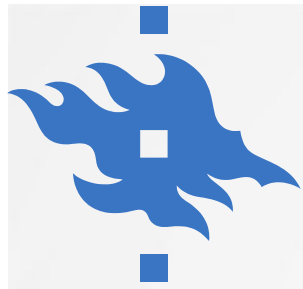
HISTORICAL TEXT COLLECTIONS COMPILED AT VARIENG

- **VARIENG** = Research Unit for the Study of Variation, Contacts and Change in English
- Community of linguists who study language variation and change
- Special focus on designing and compiling digitized historical text collections (corpora)
- <http://www.helsinki.fi/varieng/>
- *Helsinki Corpus of English Texts (HC)*
- *Helsinki Corpus of Older Scots (HCOS)*
- *Corpus of Early English Correspondence (CEEC)*
- *Corpus of Scottish Correspondence (CSC)*
- *Corpus of Early English Medical Writing (CEEM)*
- *Helsinki Corpus of British English Dialects*
- <http://www.helsinki.fi/varieng/CoRD/>

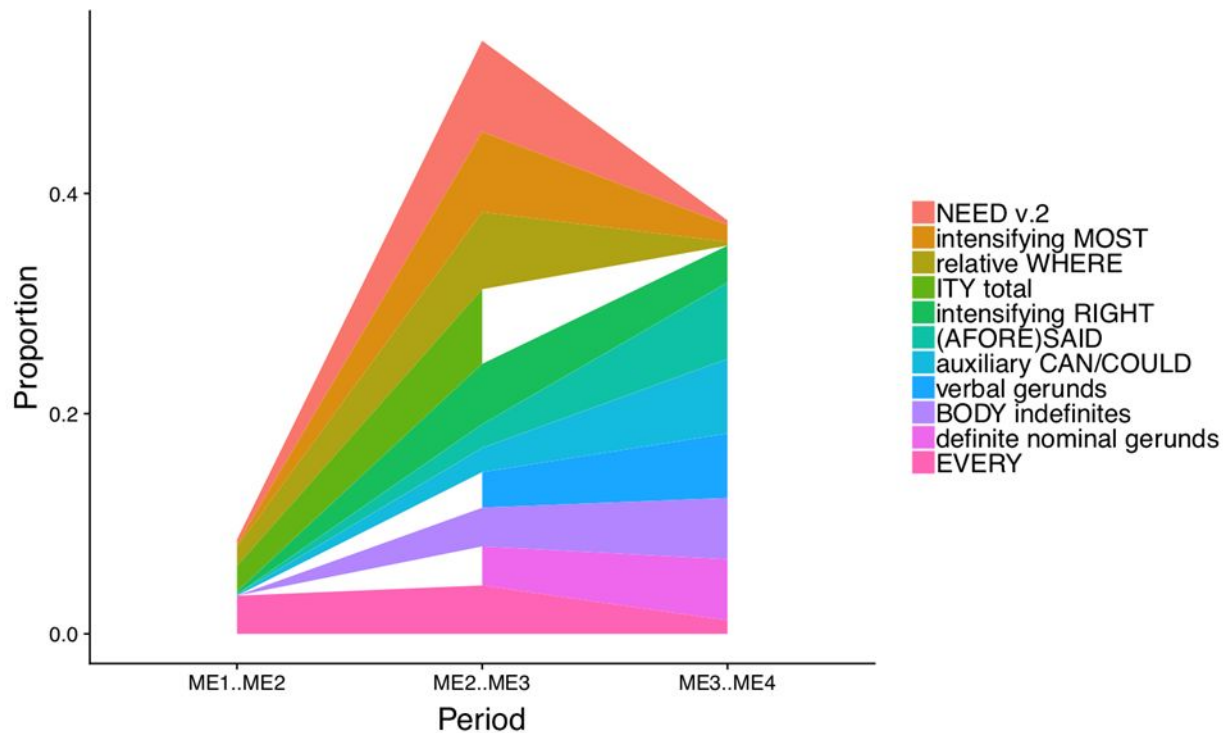


THE OED AS A TOOL AND A REPOSITORY IN CORPUS-BASED RESEARCH

- The OED is studied in its own right as a **source for vocabulary developments** over time
- The OED is a **baseline of comparison** for studies on other levels of language
- Current VARIENG Digital Humanities **projects** referring to the OED:
 - Reassessing Language Change
 - STRATAS
 - <https://www.helsinki.fi/en/researchgroups/varieng/research>
- OED antedatings from VARIENG research publications, e.g.:
- **linguist**, in the sense ‘interpreter, translator’ – 1st attestations
 - OED2 (1989) 1711
 - Kaislaniemi (2011) 1610
 - OED3 (2013) 1612
- Data from:
 - OED2 & 3: printed book (3: EEBO)
 - Kaislaniemi: Hakluyt Society edition



RATE OF CHANGE: COMPARING CORPUS FINDINGS WITH THE MIDDLE ENGLISH INTAKE OF ROMANCE LOANS



Rate of change: incoming features in the HC

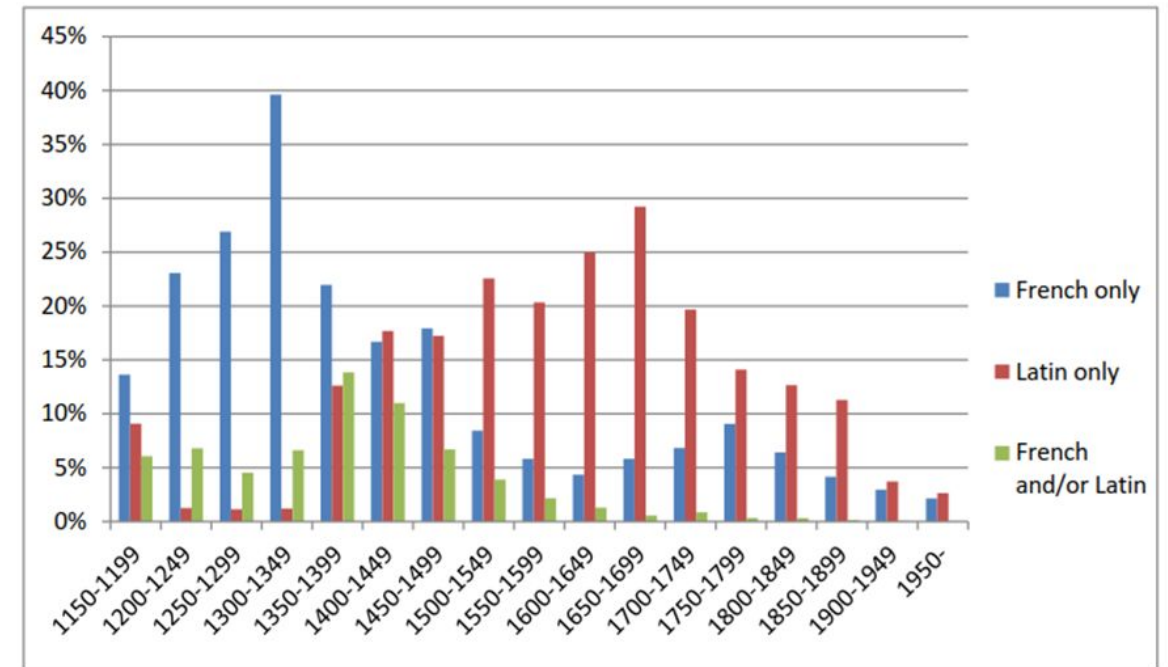


Fig. 2.6 Loanwords from French, Latin, and French and/or Latin as a proportion of all new words, as reflected by parts of *OED3* so far completed.

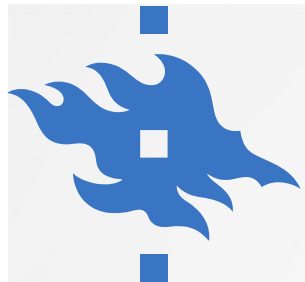


NATAS: SOCIAL ASPECTS OF NEW WORDS IN EARLY ENGLISH LETTERS



WORKFLOW

- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata. Find who there uses words not found elsewhere before. Discover differences by group, location, gender...
- How to find words not found elsewhere before?
 - Compare against **dictionaries:** *Oxford English Dictionary, Middle English Dictionary*
 - Compare against **other text collections:** *British Library Newspapers, Burney & Nichols Collections, Eighteenth Century Collections Online, Early English Books Online*



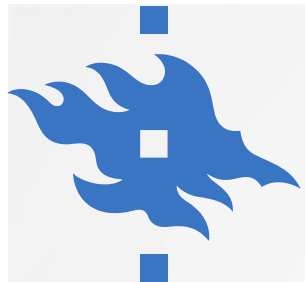
PILOT STUDY: NEW WORDS IN 18TH-CENTURY LETTERS

- *Corpora of Early English Correspondence* (CEEC), long 18th century (1680–1800)
 - 315 writers (social backgrounds known), 4,945 letters, 2.2 million words
- Criteria:
 - CEEC 1st attestation \leq OED 1st attestation
 - Can occur in max 100 contemporary published texts before CEEC 1st attestation
- Automated procedures \rightarrow only 220 candidates for human to filter in interface



81 NEW WORDS FOUND

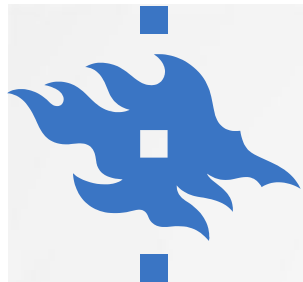
acharya, anthroponomical, anti-democrat, blueism, bonneted, canicule, cardiphonia, cast-off, catchy, chaplaincy, cleverality, curtainless, delineator, dicky-bird, double-bedded, double-cross, embodiment, envoyship, eschantillon, escritoire, freshen, fussy, godsend, grumpy, guimpe, hummingly, hydrogenate, idlish, impracticability, incomed, incontestably, inexact, inside-outness, internment, intrepid, jargonic, jumpable, keyless, kibitka, knick-knackatory, letteret, malformation, mariturient, mevrouw, missish, monotonous, moon v., moonery, nidgetty, non-papist, pacifist, paperless, pheasantry, pushery, rishi, schoolmasterishness, scratch-back, shockingly, silentious, slushy, spidery, sprawly, squeezezy, stiffish, sweet-hearted, tawdrily, trickster, truantism, unailing, unbitten, unclassed, uncommonness, undefeat, unenjoyable, uneventful, ungown, uninsured, uninteresting, unstowed, wagon-way, yester-evening



WHO ARE THE INNOVATORS?

- Surprisingly many new words compared to number of running words:
 - Lower clergy (Thomas Twining: 13)
 - **Other non-gentry** (lowest category)
 - John Jackson (2), farmer's son, upwardly mobile
 - Ignatius Sancho (2), son of a slave, upwardly mobile
 - George Culley (1), farmer
- Surprisingly few:
 - Royalty

Social rank	# of neologisms
Professional	26
Clergy, lower	22
Gentry, lower	13
Nobility	6
Gentry, upper	6
Other non-gentry	5
Clergy, upper	1
Merchant	1
Royalty	1



EXAMPLES

Mr Collings prize tup is not very capital to handle, but rather **catchy** to look at.

(George Culley to Matthew Culley, 1784;
OED 1831)

We hope he is well, and enjoys this fine weather unplagued by flies, and **unbitten** by fleas.

(Ignatius Sancho to Mrs. Cocksedge, 1779;
OED 1794)



Ignatius Sancho & Princess Elizabeth
(Wikimedia Commons)

Our dear mother is well but hurried, my sister very **fussy** & agitated, the rest of the family in full trim though *heart full* from the thoughts of so soon being seperated, with laughing faces to keep up one another's spirits.

(Elizabeth Hanover to the Prince of Wales, 1797;
OED 1831)



OED ANTEDATINGS: *CATCHY*

- OED3 **catchy**, adj., sense 2b:
‘That is designed or intended to draw attention; attractive; eye-catching.’
- Current 1st attestation:
1831 *Fraser's Mag.* July 679/2 A catchy, stage-like effect.
- Straightforward antedating:
1784 *G. CULLEY* 1 Oct. Mr Collings prize tup is not very capital to handle,
but rather **catchy** to look at.



OED ANTEDATINGS: *ANTI-DEMOCRAT*



- OED2 **anti-democrat**, n.: ‘One who is opposed to democracy or to the principles of the Democratic party in the United States.’
- Current 1st attestation:
1802 *(title)* The Republican; or, Anti-Democrat [cited in D.A.].
- Antedating:
1799 S. LENNOX 25 Mar. My husband you know is a *decided Royalist* & a supporter of *monarchy*, consequently a decided **anti-Democrat** & anti-Republican (as this term is *now* understood, for the Government of England properly speaking is a Republican form of Govt): this agrees perfectly with my own disposition.

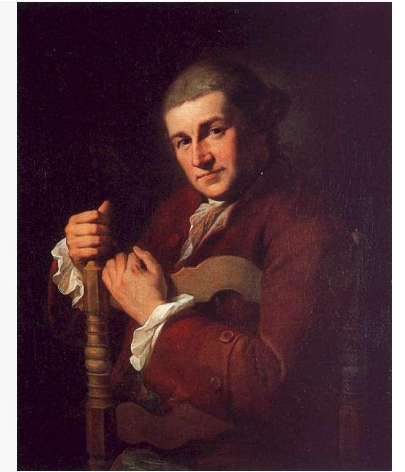


OED ANTEDATINGS: *PACIFIST*

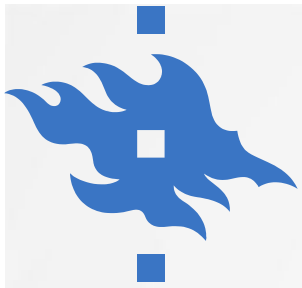
- OED3 **pacifist**, adj.: ‘Of, relating to, supportive of, or characterized by pacifism.’
- Current 1st attestation:
1906 *Times* 8 Aug 3/3 M. Paul Fribourg, waxes eloquent to-day in the *Petite République* in his reproaches of his German comrades for their opposition to the ‘Pacifist League’.
- Antedating?
1717 T. YORKE 30 May My mother..has an honest attorney, that will not hurry her into vexatious lawsuits out of any self-interest; and we take all oppertunitys to incline her to a **pacifist** disposition.



OED ANTEDATINGS: *HUMMINGLY*



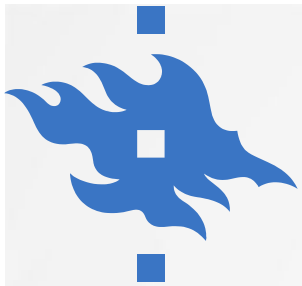
- OED2 **hummingly**, adv.: ‘With a humming sound.’
- Current 1st attestation:
1908 *Daily Chron.* 21 Jan 6/5 He..endeavours to follow the service and join hummingly in the hymns.
- Antedating?
1775 D. GARRICK 7 Nov. – the Little Gypsy* goes on **Hummingly**, & rises nightly in repute –
*Musical farce, words by Garrick
- From OED2 *humming*, adj. 2a ‘of extraordinary activity’ (1654)



OED ANTEDATINGS: *PAPERLESS*



- OED3 **paperless**, adj.: ‘Lacking official documentation; computerized’
- Current 1st attestation:
1854 T. CONNEAU
Capt. Canot 245
As soon as the cargo was safe..., I resolved to disregard the flagless and paperless craft that bore it.
- Antedating?
1799 F. BURNEY 22 July
I shall be very anxious for dry weather for them – in our curtainless & **paperless** apartment.
 - Different sense – cf. OED2 *curtainless* ‘without a curtain’ (1822)



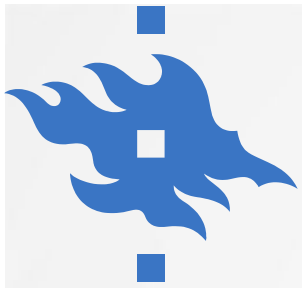
MEANINGS OF NEW WORDS IN 18TH-CENTURY LETTERS

- **People:** emotion, mental capacity, attention & judgement, behaviour, manner of action
 - *ill-natured, cleverality, nidgetty, missish, fussy*
- **Society:** communication, trade, work, faith, authority
 - *escritoire, knick-knackatory, wagon-way, chaplaincy, envoyship*
- **World:** action, space
 - *godsend, unstow*

of new words per HT category

- 6 **the world** » action or operation
- 6 **the mind** » emotion
- 5 **society** » communication
- 4 **the world** » space
- 4 **the mind** » mental capacity
- 4 **the mind** » attention and judgement
- 3 **society** » trade and finance
- 3 **society** » occupation and work
- 3 **society** » faith
- 3 **society** » authority

...



NEW WORDS: CONCLUSION

- **Big-data approach:** massive databases & automated pipeline → possible to quickly discover dozens of new words in millions of words of running text
 - Contribution to OED: not only antedatings but also new senses, etc.
- What are we **missing**?
 - Different words/senses spelled in the same way, compounds written separately, ...
 - Actual 1st attestations? Are these innovators or just early adopters?



WORKFLOWS

- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata. Find who there uses words not found elsewhere before. Discover differences by group, location, gender...
- How to find words not found elsewhere before?
 - Compare against **dictionaries:** *Oxford English Dictionary, Middle English Dictionary*
 - Compare against **other text collections:** *British Library Newspapers, Burney & Nichols Collections, Eighteenth Century Collections Online, Early English Books Online*



WORKFLOWS

- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata. Find who there uses words not found elsewhere before. Discover differences by group, location, gender...
- How to find words not found elsewhere before?
 - **Compare against dictionaries:** *Oxford English Dictionary, Middle English Dictionary*
 - Compare against **other text collections:** *British Library Newspapers, Burney & Nichols Collections, Eighteenth Century Collections Online, Early English Books Online*



WORKFLOWS

- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata
 - CEEC: Personal letters, c. 1400–1800. 1,180 writers, 11,713 letters, 5.2 million words.
- How to find words not found elsewhere before?
 - Compare against **dictionaries**:
 - Oxford English Dictionary (OED): ~280,000 words with dates, ~165,000 with earliest attestation before 1800, ~115,000 after.
 - Middle English Dictionary (MED): ~60,000 words from 1000–1500.



WORKFLOWS

- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata
 - CEEC: Personal letters, c. 1400–1800. 1,180 writers, 11,713 letters, 5.2 million words.
Contains spelling variation: ~150,000 distinct word forms. ~125,000 not tagged as foreign/proper/abbreviation
- How to find words not found elsewhere before?
 - Compare against **dictionaries:**
 - Oxford English Dictionary (OED): ~280,000 words with dates, ~165,000 with earliest attestation before 1800, ~115,000 after. **~375,000 spellings**
 - Middle English Dictionary (MED): ~60,000 words from 1000–1500. **~260,000 spellings**
- **Result:** 1/4, ~36,000 word forms can be mapped to the dictionaries, **3/4, ~88,000 cannot**



WORKFLOWS

- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata
 - CEEC: Personal letters, c. 1400–1800. 1,180 writers, 11,713 letters, 5.2 million words.
- **CEEC Contains spelling variation:** ~150,000 distinct word forms. ~125,000 not tagged as foreign/proper/abbreviation
 - “... Ships as shall be in readynesse for that service, and this matter requireing the greatest secrecy,... Your very affectionett freind” (Queen Anne, 1704)
 - “Right worshipfull and my most entierly beloude moder, in the most louly maner I recommauud me vnto youre gode moderhode,” (Elizabeth Poynings, 1459)



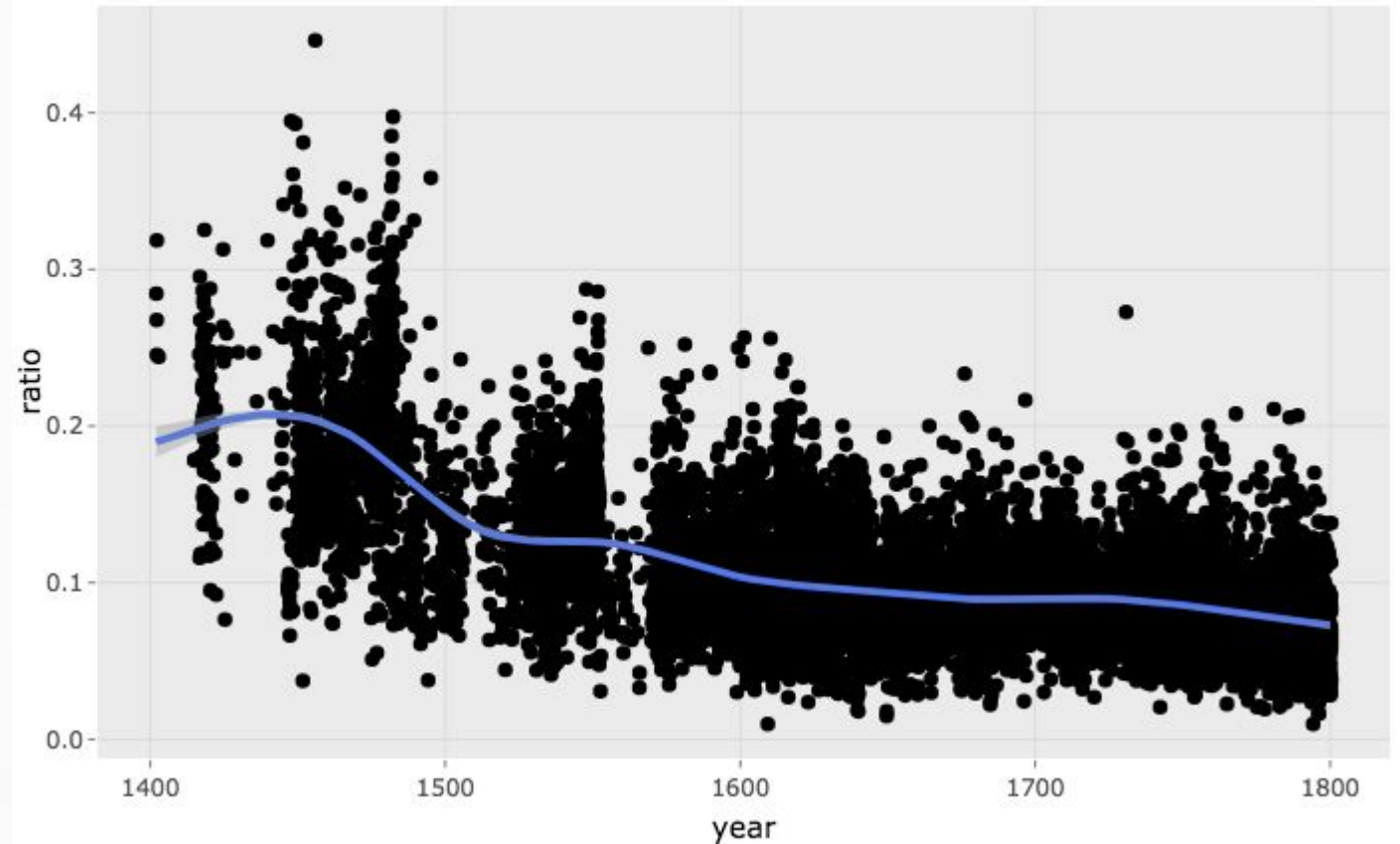
WORKFLOWS

- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata
 - CEEC: Personal letters, c. 1400–1800. 1,180 writers, 11,713 letters, 5.2 million words.
Contains spelling variation: ~150,000 distinct word forms. ~125,000 not tagged as foreign/proper/abbreviation
- How to find words not found elsewhere before?
 - Compare against **dictionaries:**
 - Oxford English Dictionary (OED): ~280,000 words with dates, ~165,000 with earliest attestation before 1800, ~115,000 after. **~375,000 spellings**
 - Middle English Dictionary (MED): ~60,000 words from 1000–1500. **~260,000 spellings**
- **Result:** 1/4, ~36,000 word forms can be mapped to the dictionaries, **3/4, ~88,000 cannot**



FILTERING STATISTICS

- Proportion not in dictionaries decreases through time (= spelling standardizes)
- ~32,000 non-foreign/non-proper words first seen in the long 18th century
- ~8,800 directly match a dictionary word, ~1,300 more are found through variants
- 487 of the matching words predate their OED attestation



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	cleveral	yes		1778	50	0	0	1
curtainless	curtaink	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-r	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

<Q A 1784 FN GCULLEY> <X GEORGE CULLEY> <P 179> [] [1..1]] [^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 **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 havesabovi eT Sen tear highly commended for making the... Comments: beingmuseer rear...s of!

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 ltfiant, 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 fignarureori rbe padke'ts~t laey ma~y be: able to

OED Oxford English Dictionary **Quick search:** Find word in dictionary **GO**
 The definitive record of the English language
 Lost for Words? | Advanced search | Help

Browse:
 Dictionary | Sources
 Categories | Historical Texts
 Timelines

My entries (13)
 My searches (15)
 Jump to:

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. **katʃi**, U.S. **ˈkætʃi**

Etymology: < CATCH v. + -Y suffix'.

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
 Editions

Entry ▼
 catchiness, n.
 catching, n.
 catching, adj.
 catchment, n.



FILTERING STATISTICS

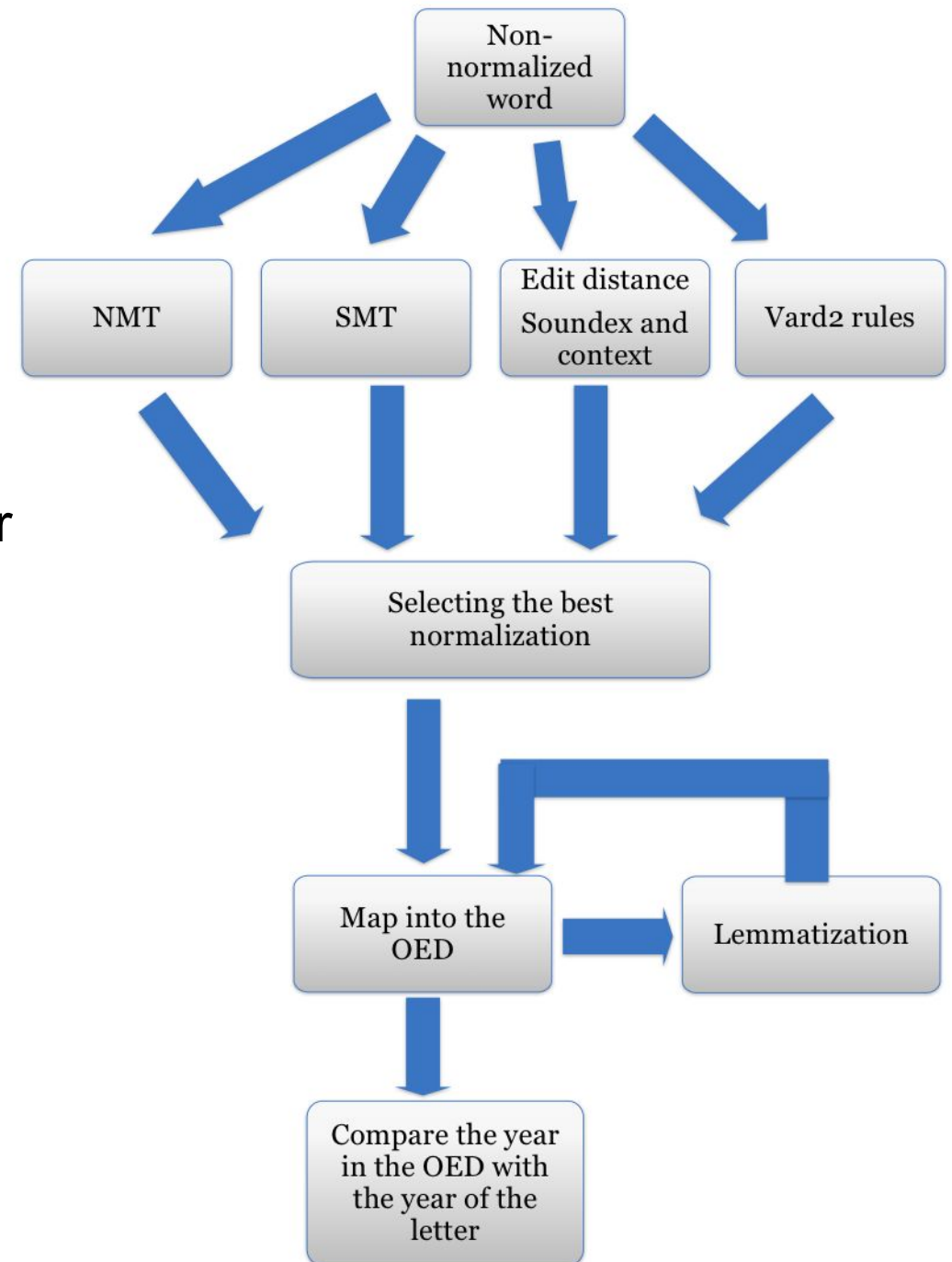
- ~32,000 non-foreign/non-proper words first seen in the long 18th century
- ~8,800 directly match a dictionary word, ~1,300 more are found through variants
- ~ $\frac{2}{3}$ of the material, ~22,000 not matched

→ cannot yet say anything reliable about social aspects of new word creation as a whole



NORMALIZATION

- **VARD2** normalization rules
 - 58 rules, e.g. “u → v anywhere”
- **Edit distance** against dictionary words + filter by context similarity
- **Character-based machine translation**
 - Neural machine translation (NMT)
 - Statistical machine translation (SMT)
 - Both use CEEC, MED and OED as input
 - Language model = British National Corpus





EXAMPLES FROM THE 18TH CENTURY

Word	NMT	SMT	Edit distance	VARD2	Correct
elyzian	elizian	elysian	elysian		Elysian
arch-type	archedip	archetype	archetype		archetype
kindelyer	kindler	kinder	kindlier		kindlier
supp'd	supped	supply	supply	cupped	supped
prosprity	prosperity	prosperity	prosperity		prosperity
affectionett	affectiont	affection	affectionate		affectionate
driction	driction	diction	direction		direction
octvo	ctuoctoo	octal	oct, oct.		octavo
mic'lemas	micelemas	miles	micklemas		Michaelmas
midetrenian	midetrenian	mid-on			Mediterranean



METHOD ACCURACIES

- NMT: 46%
 - SMT: 32%
 - Edit distance: 31%
 - VARD2 rules: 15%
 - *At least one correct: 67%*
-
- Adding social or temporal metadata to the NMT did not help

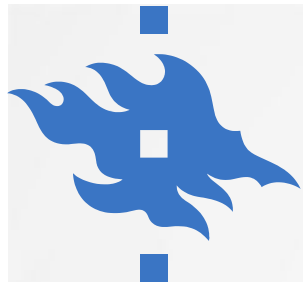


METHOD ACCURACIES

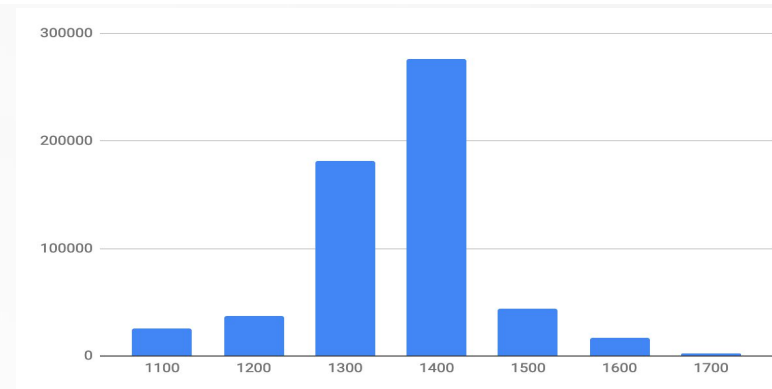
- NMT: 46%
 - SMT: 32%
 - Edit distance: 31%
 - VARD2 rules: 15%
 - *At least one correct: 67%*
-
- Adding social or temporal metadata to the NMT did not help

- Neural lemmatisation + NMT + OED filtering: 61%
- Century-specific accuracies:
 - 15th: 69%
 - 16th: 78%
 - 17th: 71%
 - 18th: 50%

→ Is 18th century variation more individual, earlier spelling variation more general?



METHOD ACCURACIES



- NMT: 46%
 - SMT: 32%
 - Edit distance: 31%
 - VARD2 rules: 15%
 - *At least one correct: 67%*
-
- Adding social or temporal metadata to the NMT did not help

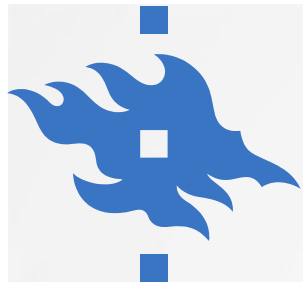
- Neural lemmatisation + NMT + OED filtering: 61%
- Century-specific accuracies:
 - 15th: 69%
 - 16th: 78%
 - 17th: 71%
 - 18th: 50%

→ Is 18th century variation more individual, earlier spelling variation more general? Or, is this an artifact due to our training data also being mostly from 14th–17th century?

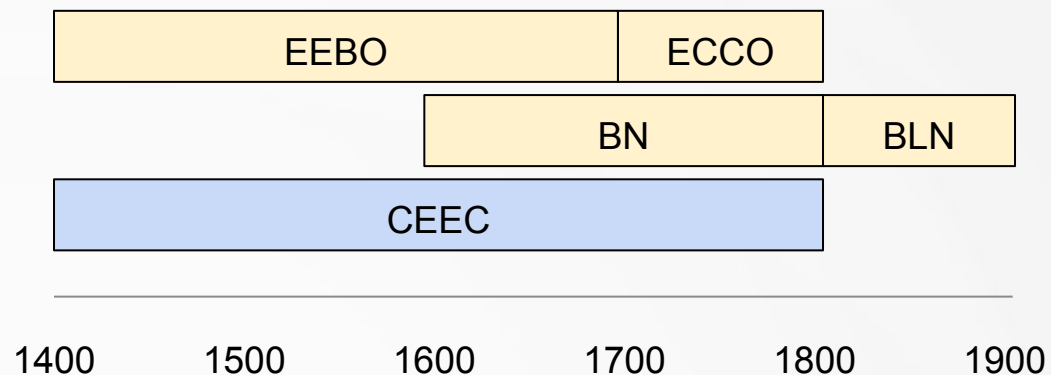


WORKFLOWS

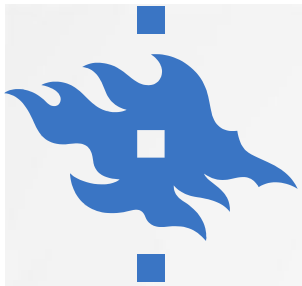
- **Research question:** discover social differences in the coining of new words
- **Approach:** take a collection of texts with social metadata. Find who there uses words not found elsewhere before. Discover differences by group, location, gender...
- How to find words not found elsewhere before?
 - Compare against **dictionaries**: *Oxford English Dictionary, Middle English Dictionary*
 - Compare against **other text collections**: *British Library Newspapers, Burney & Nichols Collections, Eighteenth Century Collections Online, Early English Books Online*



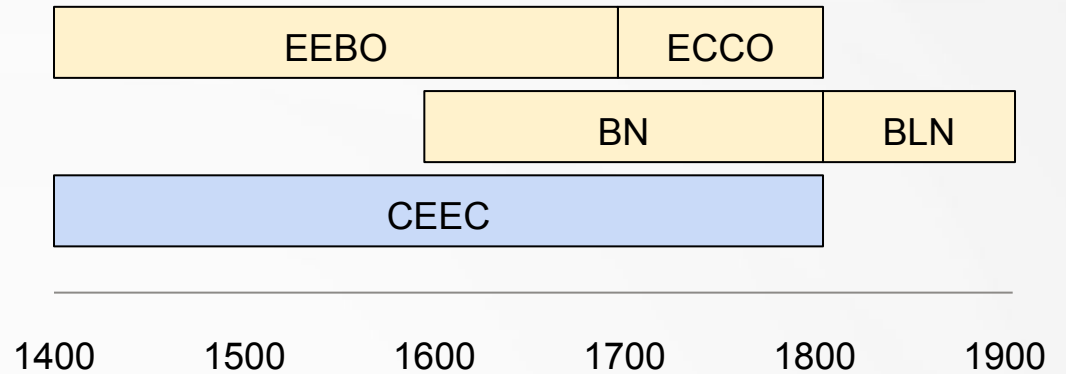
WORKFLOWS



- Research question: discover social differences in the coining of new words
- Approach: take a collection of text with social metadata
 - CEEC: Personal letters, c. 1400–1800. 1,180 writers, 11,713 letters, 5.2 million words.
- How to find words not found elsewhere before?
 - Compare against other text collections:
 - British Library Newspapers: ~655,000 newspaper issues, ~40 billion running words 1800–1900 (BLN)
 - Burney & Nichols Collections: ~230,000 pamphlets and newspapers, ~2.7 billion running words before 1600–1800 (BN)
 - Eighteenth Century Collections Online: ~210,000 books, ~10 billion running words 1700–1800 (ECCO)
 - Early English Books Online: ~60,000 books, ~1.4 billion running words 1400–1700 (EEBO)



WORKFLOWS



- Research question: discover social differences in the coining of new words
- Approach: take a collection of text with social metadata
 - CEEC: Personal letters, c. 1400–1800. 1,180 writers, 11,713 letters, 5.2 million words.
- How to find words not found elsewhere before?
 - Compare against other text collections:
 - British Library Newspapers: ~655,000 newspaper issues, ~40 billion running words 1800–1900 (BLN)
 - Burney & Nichols Collections: ~230,000 pamphlets and newspapers, ~2.7 billion running words before 1600–1800 (BN)
 - Eighteenth Century Collections Online: ~210,000 books, ~10 billion running words 1700–1800 (ECCO)
 - Early English Books Online: ~60,000 books, ~1.4 billion running words 1400–1700 (EEBO)
 - **All except EEBO contain a large number of OCR errors**



AUTOMATIC OCR

TIs not Saint *George* we Sing of here,
Nor *George*, the fatal Duke *Villier* ;
Nor *George a Green*, nor *Castriot*,
Nor *Buchanan* the learned *Scot* ;
But 'tis of *George* the Valiant *Monck*,
That made *Van-Trump* in's Blood dead-
And in the Seas his Navy funck. (drunk.
Oh! this is our brave George!

Is-w-not- Saint George we Sing of here,
Nor George, the fatal Duke Villier ;
Nor George a Green, nor Castriot,
Nor Buchanan the learned S cot q
But us of George the Valiant Monck,
That made Van-Trump in'S Blood deod
and in theseus his Navy snuck. (drunk,
Ok I this is our brave George !

THE

Firste volume of the *Chronicles of England, Scot- lande, and Irelande.*

CONTEYNING,

The description and Chronicles of England, from the
first inhabiting vnto the conquest

The description and Chronicles of Scotland, from the
first originall of the Scottes nation, till the yeare
of our Rede. 1571

The description and Chronicles of Irelande, likewise
from the firste originall of that Nation, vntill the
yeare. 1547.

Faithfully gathered and set forth, by
Raphaell Holinshed.

AT LONDON,
Imprinted for Iohn Harrison.

~ ~k ~

~ l I ~ li ~]J]O DmU ~ ov O ~ li |

~ ~ 1l ~ ~ - \ O ~ Si ~ \ r <, St ~ 5, o t %, \ ~ t, \ ~ ~ ~

~' . - bn E Is ~ l br ~; < ~ 5n ~ 1 ~

. ~ 1 1 ~ t ~ 3mo 71 ~ l < ~ 7noost I 3o ~ rsd
~ = i ~ m l m 8 7 l f i f ~ s ~

~' 3, Ilmo ~ l. 6n 3 ~ nm / 17 ~ = io \ ~ ~ 7g ~ i

.... ~ -, ~. ; l Il ~ 1 B ~] 8 ~ . ~ ~ ' ~

' . ~ ~ @ ~ ~ ~ ~ ~ p A t i l S n s t' - b ~ ~ I \ U \ ' i : ~]
~ ~ ~

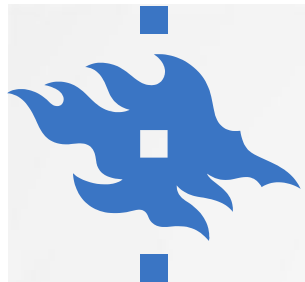
I I n o i n ~ H o d o l ~ o] b s J n i ~ q m l ' ~ 1 1 1

1 ~ . 1 ~ ~ 1 1 1

" ~ ~ | ' ? ' ~ 9 ~ 9 ~] b o O \ ~

, , - - - . ~ 1 3 ~ ~ ~

- : ~ _ 1



FILTERING STATISTICS

Out of the ~22,000 18th century words that couldn't be mapped to a dictionary word:

- only 590 appear in over 100 documents in the comparison corpora before
- ~9,000 appear in 100 or less documents at any time
- ~2,500 appear in more than 100 documents later but not before



ASKING THE RIGHT QUESTIONS

- Because we have so many sources of noise that aren't uniformly distributed, purely automated general analysis of social aspects is impossible
- Need to ask more focused questions that can be answered by combining the tools we do have with a bearable amount of manual labour
- Limit by time, filter by appearance in dictionaries and texts before/after
- “Who coins new words” →
 - “Who coins new words in the 18th century that come into general use later?”
 - “Who coins nonce words in the 18th century that are used only once and never seen again?”

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	cleveral	yes		1778	50	0	0	1
curtainless	curtaink	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-r	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

<Q A 1784 FN GCULLEY> <X GEORGE CULLEY> <P 179> [] [1..1]] [^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 **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 havesabovi eT Sen tear highly commended for making the OED. **catchy** i geon, deceafed, begs leave to inform the pullie, that he has ,le, chi
 Guy, of Duke-dtreet, York-buildings, Sur; **catchy** i geon, deceafed, begs leave to inform the pullie, that he has ,le, chi
 NTbhrfdar nert, the 12th ltfiant, 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 ldnngnm, thore **Catchy** fulfilled ;; -for if th y- lhotild have any doubts with refpefl to the fignarureori rbe padke'ts~t laey ma~y be: able to

OED Oxford English Dictionary **Quick search:** Find word in dictionary **GO**
 The definitive record of the English language
 Lost for Words? | Advanced search | Help

Browse:
 Dictionary | Sources
 Categories | Historical Texts
 Timelines

My entries (13)
 My searches (15)
 Jump to:

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

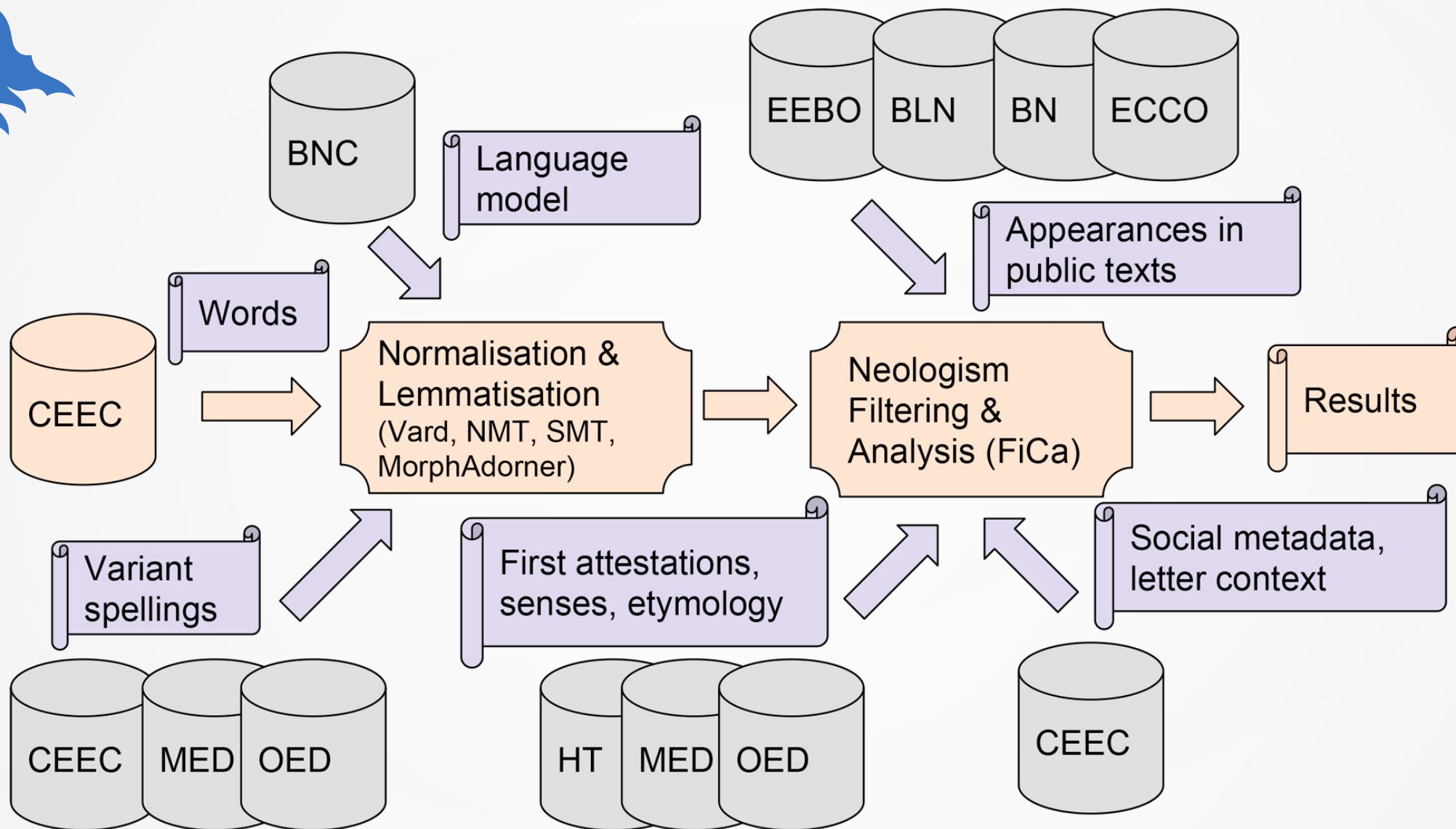
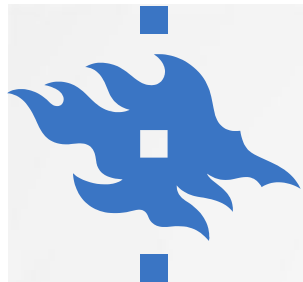
Pronunciation: Brit. **ˈ**katʃi/, U.S. **ˈ**ketʃi/
Etymology: < CATCH v. + -Y suffix'.

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

This entry has not yet been fully updated (first published 1889).
 Entry history
 Entry profile
 Previous version: OED2 (1989)
 In other editions

Thesaurus >
 Categories >

Entry ▼
 catchiness, n.
 catching, n.
 catching, adj.
 catchment, n.





FUTURE WORK

- Extend analysis to entire time period; also analyse words not in the OED, check instances in contemporary published texts
- Improve normalization, develop discovery pipeline further
- Integrate social analysis functionality to pipeline