

04 – Dictionaries: fundamentals, examples, use cases

Lexicology and Lexicography – [**Course Website**](#)

Dr. Quirin Würschinger, LMU Munich

November 11, 2025

Introduction

Recap

- Morphology vs word-formation: structure of words vs processes that create new lexemes
- Inflection vs derivation: grammatical endings (e.g., *walked*, *cats*) vs new lexemes/word classes (e.g., *kindness*, *happiness*)
- Morphemic vs non-morphemic word-formation: derivation/compounding vs conversion, clipping, blending, back-formation, acronyms
- Takeaway: we reasoned about words via relations and structure – not mere lists; today we link this perspective to how dictionaries model lexis

Outline

1. **The Lexicon**: How does lexicology move beyond lists to structures?
2. **Dictionaries**: What do dictionaries describe and how are they used?
3. **Empirical study with the OED**: How can we extract and analyse neologisms?

The Lexicon

Two senses of *lexicon*

Lipka (1992), p. 11

"In the following, I will use *lexicon* in two senses that are not always sharply distinguished:

- a. for a **metalinguistic level**, or a sub-component in a linguistic model; and
- b. in the **sense of vocabulary** as seen from a systematic, synchronic point of view."

Key ideas – Two senses of lexicon

- Metalinguistic component in a linguistic model (place in a theory)
- Vocabulary of a language/community (items and their relations)
- Ambiguity in usage – clarify which sense is meant in context

Objective of lexicology

- “What is most important, however, is that in lexicology the stock of words or lexical items is not simply regarded as a **list of isolated elements**.
- Lexicologists try to find out **generalisations** and **regularities** and especially consider **relations** between elements (see chapters n and IV).
- Lexicology is therefore concerned with **structures**, not with a mere agglomeration of words (cf. Jackson (1988): 222).
(Lipka 1992, 1)

Key ideas – Objective of lexicology

- Move beyond lists: explain items via patterns and relations
- Focus on structures (paradigmatic/syntagmatic) and generalisations
- Aim: connect evidence to theory and make predictions

List vs structure – quick examples

- List view: “Words in English cooking”
 - *butter, salt, pepper, pan, fry*
- Structure view (relations):
 - collocation: *strong tea* vs. ? *powerful tea*
 - substitution: *cat* → *dog* (same slot), *fry* → *sauté* (near-synonym)

Contrasts to discuss:

- strong vs. powerful (collocation)
- *cat* vs. *dog* (paradigmatic slot)
- *fry* vs. *sauté* (near-synonym proximity)

Definition (**Bauer 2022**)

- “The lexicon is what linguists call the dictionary that is assumed to be **in people’s heads**, or the linguist’s best approximation to that.
- It is fundamentally a *psychological entity*, and its contents cannot be observed directly but must be deduced from speakers’ and hearers’ **behaviour**.
- Two fundamental approaches:
 1. **Home of the lawless** (**Di Sciullo and Williams 1987**) – lexicon contains whatever cannot be predicted by general rule.
 2. Wider approach – lexicon contains anything to do with the **structure of words**, whether lawless or not.”

The mental lexicon

Evidence for a mental lexicon

- Slips and blends; priming effects; Stroop interference
 - e.g., *RED* printed in blue slows colour naming
- Reaction-time studies (lexical decision, frequency effects)
 - e.g., faster decisions for frequent *time* than rare *murrain*
- Network structure: degree, clustering, multiplex links
 - e.g., *dog* linked to *cat* (semantic) and *log* (phonological)

Cognitive-linguistic models of the lexicon

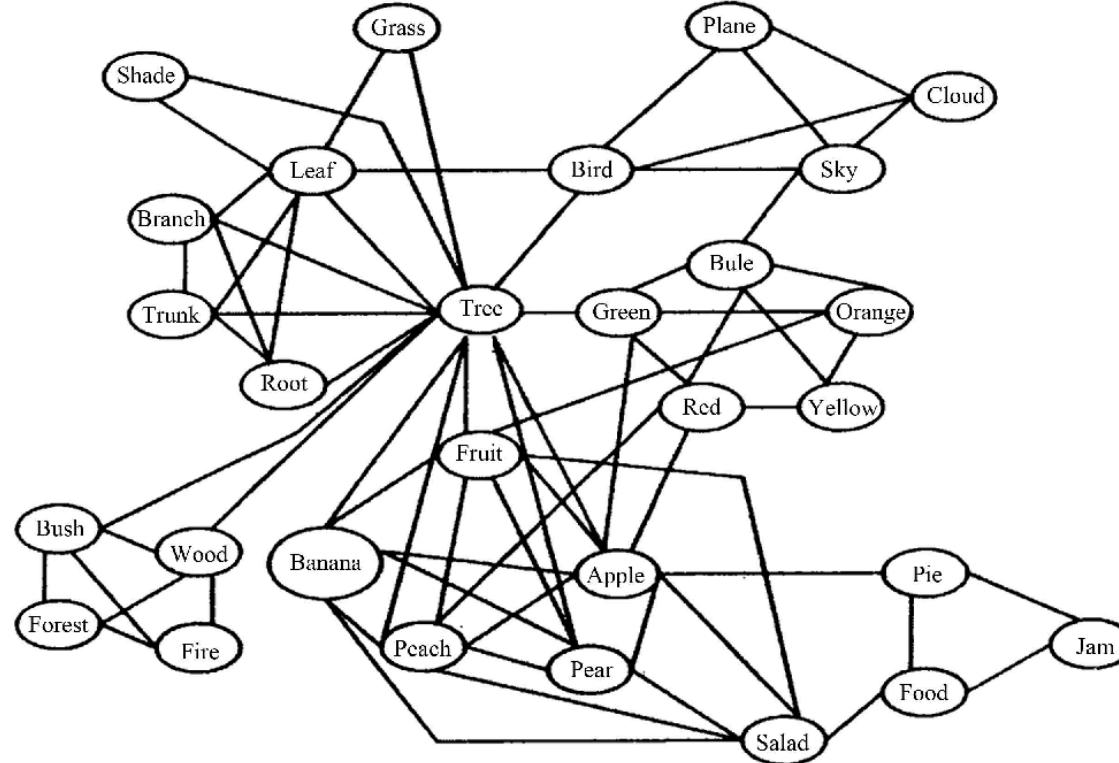
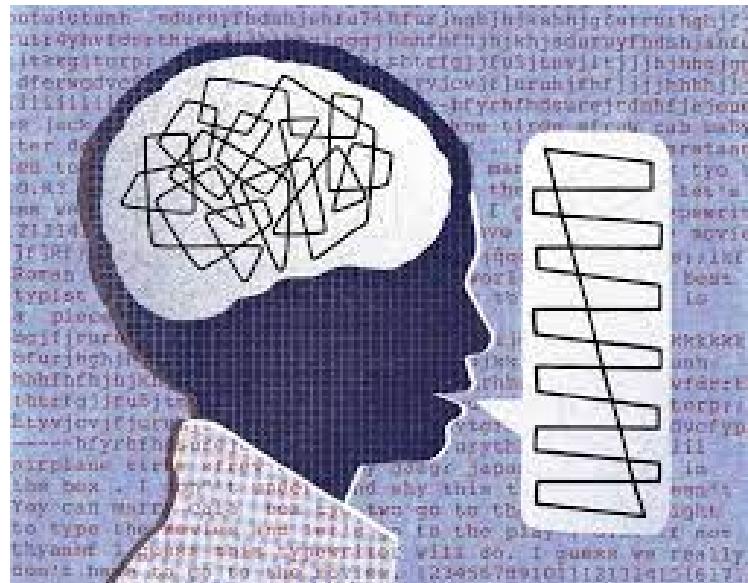
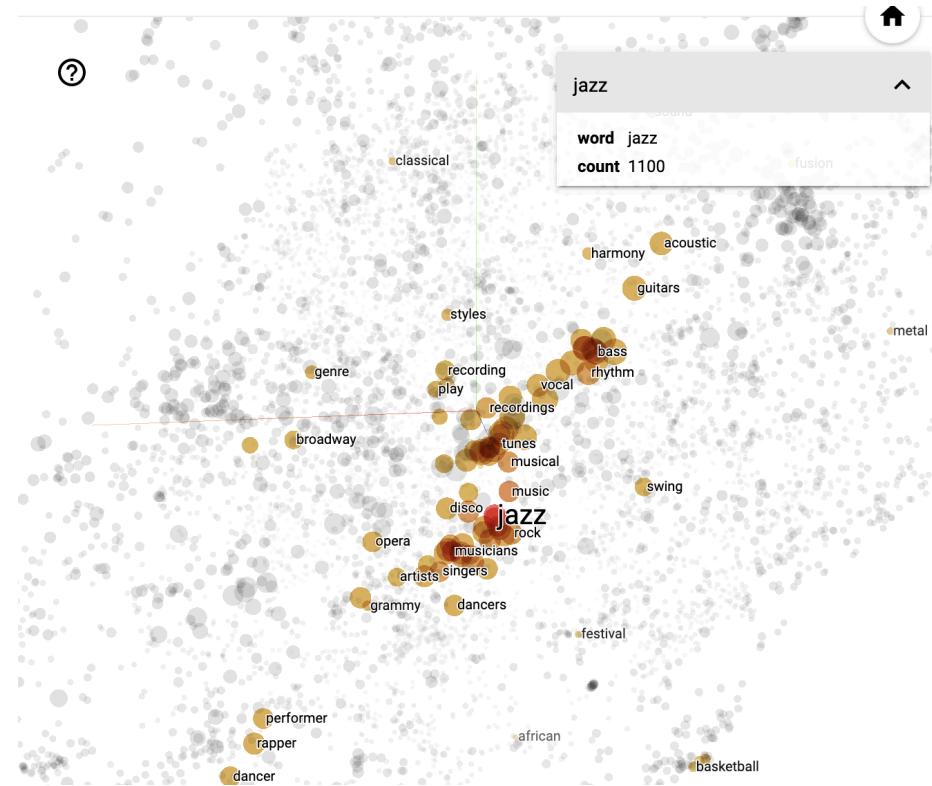
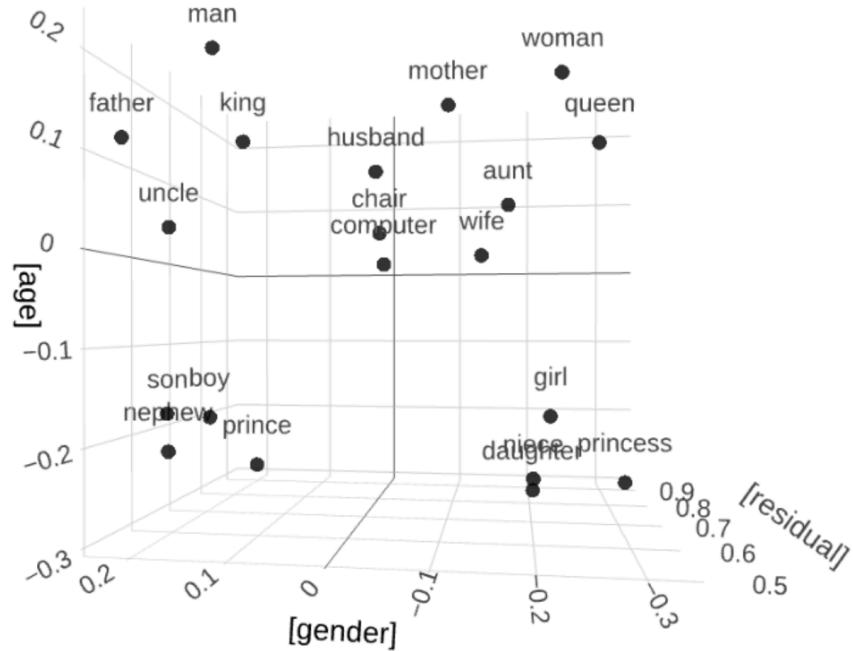


Figure 3 Collins & Loftus (1975) spreading activation model

Word embeddings (*Bandyopadhyay et al. 2022*)

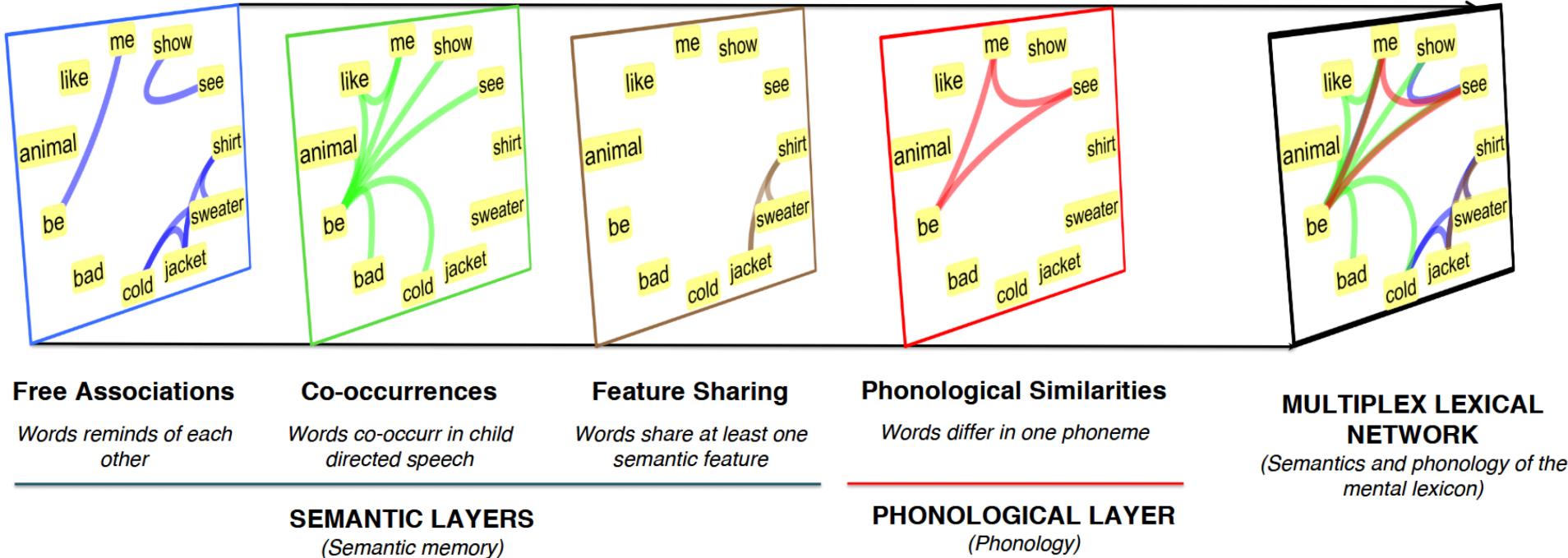
<https://projector.tensorflow.org/>



Embeddings – caveats and use

- Capture distributional similarity (usage-based); not “meaning equals vectors”
- Sensitive to training data/bias; opaque dimensions
- Best for intuition-building and retrieval; not a ground truth

Associations in the lexicon



(Stella 2019)

The Stroop Effect (Stroop 1935)

blue	red	orange	purple	green
red	blue	green	orange	purple
orange	blue	red	green	purple
purple	orange	blue	green	red
green	orange	purple	red	blue

Freudian slip (Bauer 2022)

"If you talk to someone about electricity, and then make them read the phrase *sham dock*, they might very well say *damn shock*, because of prior activation of related words; the Freudian slip depends upon a subject area being readily activated in the brain and brought out inadvertently, whether because of a word related in meaning or pronunciation."

Dictionaries (Bauer 2022)

What is a word?

- “A word is listed in the dictionary.” (**Bauer 2022, 2**) But:
 - circular reasoning – words are listed because they are words
 - dictionaries also list smaller units (e.g. prefix *un-*)
- “[...] in general we will accept the spelling conventions of English as defining words. This may not be terribly scientific, but it has the advantage of being practical.” (**Bauer 2022, 3**)

Definition of *dictionary*

- “The term *dictionary* is usually restricted to real-world dictionaries that appear in print and online.
- Dictionaries provide a list of words of whatever language they deal with – in our case, English – and then give a certain amount of information about each of them.
- They tend to have two functions, which may at times conflict:
 1. **Describe** the language as it is (descriptive)
 2. Provide an influence for establishing and maintaining the **standard** form of the language (prescriptive)”

From data to dictionary entries

- Corpus evidence → candidate headwords and senses
- Sense discrimination from usage patterns
- Definition writing and usage labels (register, region, style)
- Attested examples and cross-references (derivatives, phraseology)

How are definitions written?

- Genus–differentia vs paraphrase
- Controlled defining vocabulary in learner's dictionaries
- Clarity vs coverage: trade-offs

Check: propose a defining strategy (justify; then rewrite)

1. A bird category term (e.g., "raptor") for non-experts
2. "Platform" (as used in social media contexts)
3. "Set" (noun) – one common sense for beginners

Tasks: - Choose strategy (genus–differentia or paraphrase) and justify with intended user. - Rewrite the definition using a controlled defining vocabulary (learner's style).

Ordering senses

- Chronological (OED) vs prototypical/frequency-first (learner's)
- Impact on user tasks (history vs learning)
- Criteria: earliest attested vs most typical/current; user intent matters

Example: *tablet*

- OED (historical principle): 'slab/stone or clay for writing; pill'
→ later 'flat touchscreen computer'
- Learner's (prototype/frequency): foregrounds 'flat touchscreen computer' for modern users

Consequences

- Chronology helps trace change and citations.
- Prototype helps comprehension and production.

Usage labels

Purpose: guide choices, not police usage

- Typical labels with examples:
 - **region**: *lorry* – chiefly BrE; *soccer* – chiefly AmE
 - **register**: *ain't* – informal; *request* vs *ask* – formal vs neutral
 - **field**: *suture* – medical; *lemma* – linguistics
 - **stance**: pejorative/derogatory flags where appropriate

Principles: base on attested use; avoid over-labelling; use neutral wording in definitions.

Check: propose appropriate labels (justify with evidence)

- sick ('excellent') – labels? _____ (region/register/stance?)
- *they* (singular for a specific person) – labels? _____ (usage note?)
- *data* as a singular mass noun – labels? _____ (field/register?)
- *woke* (adjective, politicised) – labels? _____ (stance/region?)
- *soccer* in UK newspapers – labels? _____ (region/chronology?)

Phraseology in entries

- Collocations (*strong tea, heavy rain*), patterns (V + decision)
- MWEs: idioms, phrasal verbs, compounds

Which is the “better” collocate?

1. *heavy/strong rain*
2. *deeply/strongly held belief*
3. *do/make research*
4. *commit/make a crime*
5. *take/make a decision*

Why examples matter

- Citations justify senses and usage labels
- Curated vs auto-extracted: precision vs coverage
- Bias: genres, time periods, contributor communities
 - e.g., pre-2000 slang under-represented in print-heavy sources → missing senses
 - e.g., BrE vs AmE: American news sources over-represent “make a decision” vs BrE “take a decision”
 - e.g., field bias: biomedical corpora inflate plural “data are” vs general usage “data is”

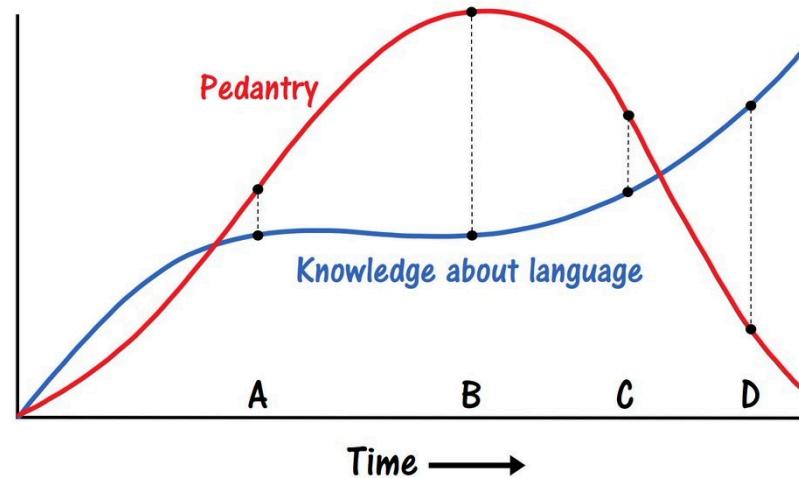
Descriptivism vs prescriptivism

Bauer (2022): 18

- “If we assume that one of ‘kilometre’ and ‘ki’lometre’ is right (and the other therefore wrong), we assume that there is a **unique solution** to this question of English usage ...”
- Language questions are more often like the jeans question (multiple acceptable answers) than like the *drive-on-the-right* question (single legal answer).

A graph and explanation of linguistic knowledge vs linguistic pedantry

@robdrummond



- A. Danger point. Increased knowledge about language has led to a surge in linguistic pedantry, resulting in an imbalance.
- B. Crisis point. Full on pedant. False confidence in good linguistic knowledge has led to pedantic half-truths outweighing facts.
- C. The road to recovery. Taking the time to understand how language actually works has led to a more tolerant outlook.
- D. A sensible place to be. The more you know about language, the more you see the nonsense of linguistic pedantry. Be aware that at point (D) you may be called upon to challenge people at point (B). This can be difficult, as people at point (B) often have the weight of public opinion behind them. Be patient, and encourage them to move towards point (C).

Proof of the existence of words

"First of all, dictionaries provide evidence of the existence of a word. The fact that a word is listed in a dictionary at all is taken to prove that there is such a word. This can be misleading in two ways:

1. Dictionaries **sometimes list erroneous words** that have no existence outside the dictionary (e.g. *banket, sardel*).
2. Dictionaries more often **fail to list perfectly good words** – no dictionary can list every word of English."

Information provided in the OED

O E D | Oxford English Dictionary   Dictionary  Advanced search

Revised 2016 (details) More entries for "bro"  Cite  Contribute Tabbed view 

bro

— NOUN —

Factsheet Meaning & use Etymology Pronunciation Forms Frequency Compounds & derived words

CONTENTS

- I. As a graphic abbreviation for brother.
 - I.1. In singular = brother, n. (in various senses).
In later use...
 - I.2. In plural (in form Bros.). In the names of companies...
- II. slang and colloquial uses.
 - II.3. colloquial (originally Caribbean and U.S. regional...)
 - II.4. A person's male sibling; = brother, n. A.I.1a.
 - II.5. slang (originally and chiefly U.S.).

I. As a graphic abbreviation for *brother*.

I.1. In singular = **brother** *n.* (in various senses). In later use chiefly in religious or ceremonial titles.

a1530 *Bro.* It is not syster, the numbre of wordes or of lynes, that maketh a tale or a boke to be called longe or shorte.
T. Lupset, *Treat. Charitie* (1533) f. 27 ...

...

2005 The lodge master, **Bro.** Cameron, gave a report.

Selected dictionaries

Urban Dictionary

<https://www.urbandictionary.com/>

The screenshot shows the Urban Dictionary homepage with the search bar containing 'Type any word...'. Below the search bar are navigation links: Browse, Categories, Store, and Blog. To the right are three blue circular icons with white symbols: a plus sign, a crossed-out symbol, and a person icon.

TOP DEFINITION

neologism

Itself an invented word used exclusively pejoratively to [dismiss](#) newly coined words. Usually used to express [distaste](#) for words inconvenient to one's [ideology](#).

Pederasts, a more [perverse](#) subtype of faggots, [declare moralphobia](#) a neologism.

by [Tokenizing Words](#) December 19, 2007

Activity

A line graph titled 'ACTIVITY' showing the volume of activity over time from 2012 to 2021. The activity is highly volatile, with several sharp peaks and troughs, indicating periods of high and low usage.

Get a **neologism mug for your dog José.**

Word

© 1999-2021 Urban Dictionary® • [advertise](#) • [terms of service](#) • [privacy](#) • [dmca](#) • [bug report](#) • [help](#) • [blog](#) • [data subject request](#)

<https://www.wiktionary.org/>

English [\[edit\]](#)

Etymology [\[edit\]](#)

From French [néologisme](#), from Ancient Greek νέος (*néos*, “new”) + λόγος (*lógos*, “word”).



English [Wikipedia](#) has an article
on:
neologism

Pronunciation [\[edit\]](#)

- IPA^(key): /ni:'ɒlədʒɪzəm/

Noun [\[edit\]](#)

neologism (*countable* and *uncountable*, plural [neologisms](#))

1. (*linguistics*, *lexicography*, *countable*) A word or phrase which has recently been coined; a new word or phrase.

Synonym: [coinage](#)

2. (*linguistics*, *uncountable*) The act or instance of coining, or uttering a new word.

3. (*psychiatry*) The newly coined, meaningless words or phrases of someone with a [psychosis](#), usually [schizophrenia](#).

4. The introduction of new [doctrine](#), for example in [theology](#).

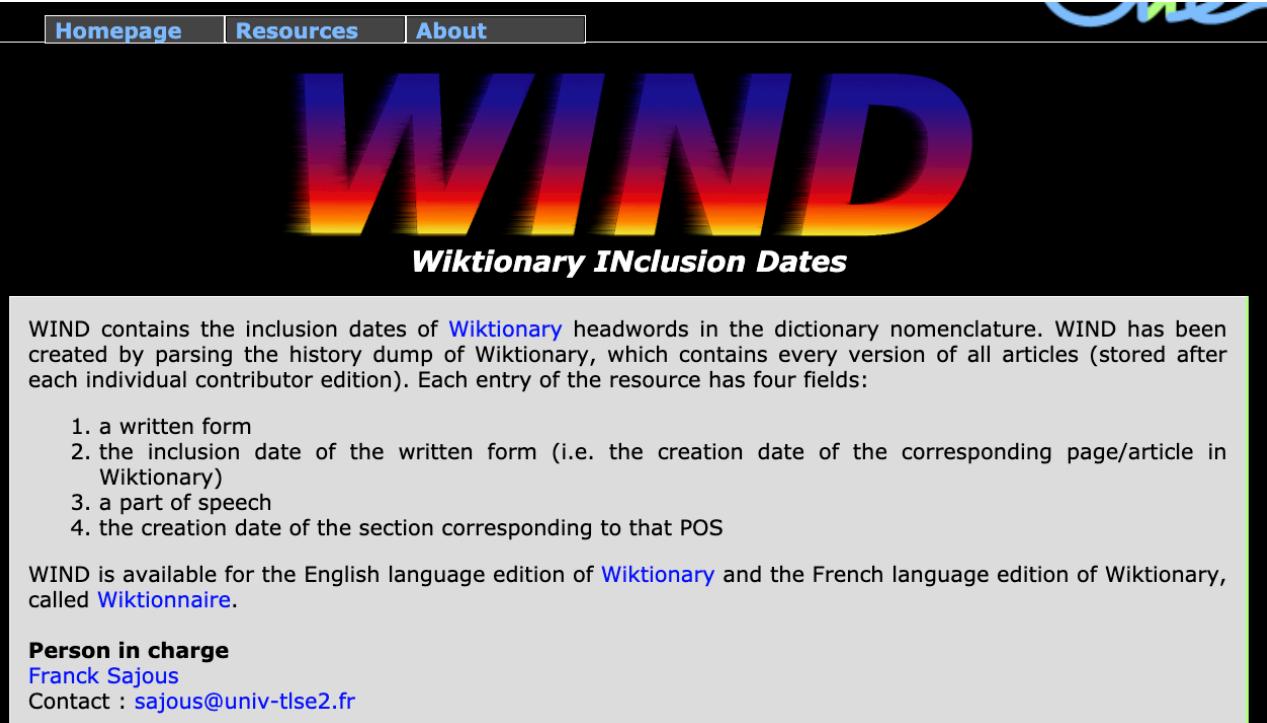
Usage notes [\[edit\]](#)

- There is no precise moment when a word stops being "new", but 15–20 years is a common cutoff (corresponding to one [generation](#) growing up potentially familiar with the word, depending on how common it is). Acceptance of a word as valid by dictionaries or by a significant portion of the population are sometimes mentioned as additional conditions. Some neologisms become widespread and standard (such as new chemical element names), others remain rare or slangy. (Distinguish from [protologisms](#), coinages which have not become common.)

WIND (Machine-readable Wiktionary)

Sajous, Calderone, and Hathout (2020)

<http://redac.univ-tlse2.fr/lexiques/wind.html>



The screenshot shows the homepage of the WIND project. At the top, there is a navigation bar with three items: "Homepage" (highlighted in blue), "Resources", and "About". Below the navigation bar is a large, stylized title "WIND" in a gradient of blue, purple, and red. Underneath the title, the subtitle "Wiktionary INclusion Dates" is displayed in a smaller, bold, white font. A main text block explains what WIND is and how it works. It states: "WIND contains the inclusion dates of [Wiktionary](#) headwords in the dictionary nomenclature. WIND has been created by parsing the history dump of Wiktionary, which contains every version of all articles (stored after each individual contributor edition). Each entry of the resource has four fields:". Below this, a numbered list details these four fields: 1. a written form, 2. the inclusion date of the written form (i.e. the creation date of the corresponding page/article in Wiktionary), 3. a part of speech, and 4. the creation date of the section corresponding to that POS. Further down, it notes that WIND is available for the English language edition of [Wiktionary](#) and the French language edition of Wiktionary, called [Wiktionnaire](#). At the bottom, there is a section titled "Person in charge" with the name "Franck Sajous" and contact information: "Contact : sajous@univ-tlse2.fr".

	A	B	C	D	E	F	G
1	Entry	PagId	Entry inclusion	#Revisions	POS	POS inclusion	#POS rev
2	enWiktHist_32.csv:"Loboda"	7339302	2019-09-01	4 NP	2019-09-01	1	
3	enWiktHist_32.csv:"moving target"	7339196	2019-09-01	3 NC	2019-09-01	1	
4	enWiktHist_32.csv:"baywing"	7339282	2019-09-01	2 NC	2019-09-01	1	
5	enWiktHist_32.csv:"bombing run"	7339197	2019-09-01	2 NC	2019-09-01	1	
6	enWiktHist_32.csv:"checkstand"	7339203	2019-09-01	2 NC	2019-09-01	1	
7	enWiktHist_32.csv:"EPAC"	7339213	2019-09-01	2 NC	2019-09-01	1	
8	enWiktHist_32.csv:"flight sequence"	7339198	2019-09-01	2 NC	2019-09-01	1	
9	enWiktHist_32.csv:"none other than"	7339189	2019-09-01	2 ADJ	2019-09-01	1	
10	enWiktHist_32.csv:"Shapoval"	7339277	2019-09-01	2 NP	2019-09-01	1	
11	enWiktHist_32.csv:"slangery"	7339168	2019-09-01	2 NC	2019-09-01	1	
12	enWiktHist_32.csv:"taraxasterol"	7339054	2019-09-01	2 NC	2019-09-01	1	
13	enWiktHist_32.csv:"thiophosphoryl"	7339222	2019-09-01	2 NC	2019-09-01	1	
14	enWiktHist_32.csv:"attb"	7339018	2019-09-01	1 NC	2019-09-01	1	
15	enWiktHist_32.csv:"cash wrap"	7339204	2019-09-01	1 NC	2019-09-01	1	
16	enWiktHist_32.csv:"döner kebab"	7339208	2019-09-01	1 NC	2019-09-01	1	
17	enWiktHist_32.csv:"e-scooter"	7339284	2019-09-01	1 NC	2019-09-01	1	
18	enWiktHist_32.csv:"go-ped"	7339220	2019-09-01	1 NC	2019-09-01	1	
19	enWiktHist_32.csv:"landful"	7339011	2019-09-01	1 NC	2019-09-01	1	
20	enWiktHist_32.csv:"lawnful"	7338994	2019-09-01	1 NC	2019-09-01	1	
21	enWiktHist_32.csv:"libraryful"	7339012	2019-09-01	1 NC	2019-09-01	1	
22	enWiktHist_32.csv:"ninety-ninety rule"	7339224	2019-09-01	1 NC	2019-09-01	1	
23	enWiktHist_32.csv:"Rossby"	7339225	2019-09-01	1 NP	2019-09-01	1	
24	enWiktHist_32.csv:"SMOP"	7339279	2019-09-01	1 NC	2019-09-01	1	

Excel Sheet

O E D | Oxford English Dictionary   Dictionary  Advanced search

Revised 2003 (details)  Cite  Contribute Tabbed view 

neologism

NOUN

Factsheet Meaning & use Etymology Pronunciation Frequency Compounds & derived words

CONTENTS

- ▶ 1. A word or phrase which is new to the language; one which is...
- 2. Theology. The holding or adoption of novel (esp...)

QUOTATIONS

Show all quotations

1.a. A word or phrase which is new to the language; one which is newly coined. 1772-

1772 Observations on this Neologism ... I thought indeed I was intirely done with this Canon when I came to the explication of the last word *Hecaterogenosem*.

J.-N. de Saussure, *Analysis of French Orthography* 163 ...

Advanced search interface

O E D | Oxford English Dictionary   Dictionary  Advanced search

Entries (518,827) Meanings (878,485) Quotations (3,840,299) Historical Thesaurus (794,871)

SEARCH TERMS

Headword Quotation text
 Definition Quotation author
 Etymology Quotation work
 Forms

Add term Update

1 to 50 of 518,827 results [Save search](#) 1 2 Next >>

Sort by Frequency Export results (.csv)

Old English- 	the, adj., pron.², & n.¹ Before a noun denoting a thing or person that is unique or considered to be unique, or of which there is only one at a time (e.g. the sun, the earth...)
Old English- 	be, v. In progressive (chiefly non-perfect) tenses. With dynamic sense: to exist as some particular thing or as having some particular quality at a specific...
Old English- 	of, prep. Preceded by a word of number or quantity.
?a1160-	a. adi.

Bridge to the empirical task

- Extract post-2000 first attestations via OED Advanced Search
- Analyse word class, formation type, subject area in Excel

Studying word-formation
based on empirical data using
the OED

Research questions

1. Which neologisms have entered the OED since 2000?
2. What are the most common word classes among these recent neologisms?
3. What are the most common word-formation processes?
4. Which semantic domains are most common?
5. Since when have these neologisms been in use?

Extracting neologisms from the OED

Using OED's Advanced Search functionality, we can extract neologisms – words that have only been used since the year 2000 – from the OED.

The screenshot shows the Oxford English Dictionary Advanced Search interface. The search terms entered are not visible. The search parameters include:

- SEARCH TERMS:** Headword (selected), Definition, Etymology, Forms, Exact match.
- DATE OF USE:** First use (selected), From: 2000, To: Year (YYYY).
- Sort by:** Frequency.
- Results:** 1 to 50 of 710 results.
- Buttons:** Save search, Export results (.csv), Add term, Update.

The results list includes:

- 2020– Covid-19, n.**
An acute disease in humans caused by a...
- 2020– Covid, n.²**
The disease Covid-19; (also) the coronavirus...
- 2008– Bitcoin, n.**
A type of digital currency introduced in 2009...
- 2002– selfie, n.**
A photograph that one has taken of oneself, esp...

Search URL

Analyzing the data in Microsoft Excel

model Excel sheet: [URL](#)

Data preparation

1. Open a blank Excel sheet.
2. Click **import** and **csv file**.
3. Select file origin encoding to **Unicode (UTF-8)**.
4. Specify that the data are **delimited by commas (csv)** as column separator.
5. Convert the region containing the data to a **Table**.
6. Insert one **Pivot Table** (and **Pivot Chart**) for each research question.

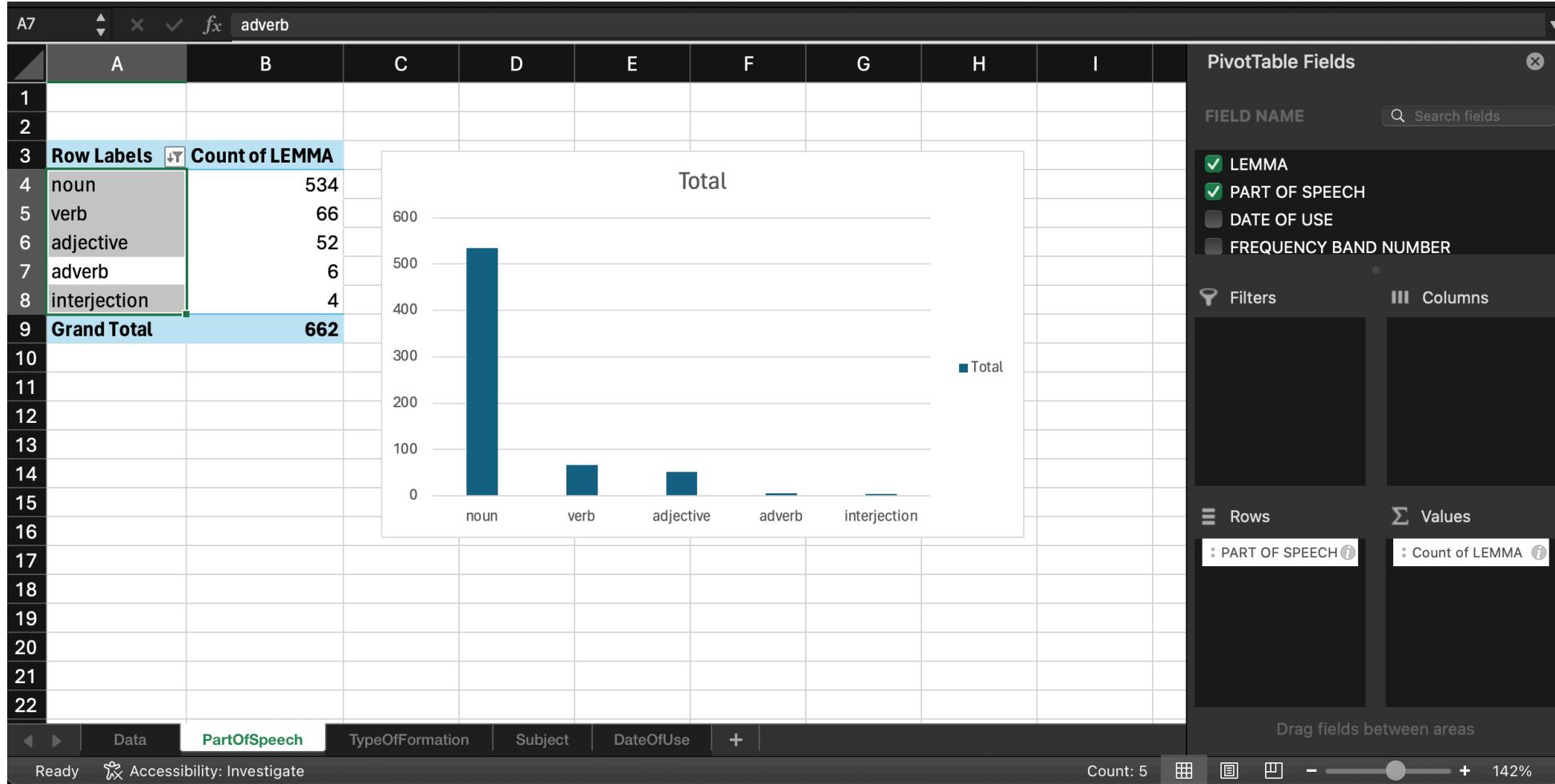
Table

first-use-from-2000 — Saved

B8 fx noun

	A	B	C	D	E	F
1	LEMMA	PART OF SPEEC	DATE OF USE	FREQUEN	DEFINITION	TYPE OF FORMATION
2	CIO	noun	-2000	5	'Cry it out', a method of sleep training in which...	initialism
3	DDoS	noun	-2000	4	= distributed denial of service, n.; usually...	initialism
4	click and collect	adjective, noun	-2000	4	Of or relating to a shopping facility whereby a...	compound
5	CubeSat	noun	-2000	4	A small, lightweight space satellite that fits...	compound
6	stan	noun	-2000	4	slang (derogatory, except when self-deprecatory)...	proper name
7	Greatest Generation	noun	-2000	3	The generation of Americans reaching adulthood...	compound
8	Sudoku	noun	-2000	3	A type of logic puzzle, the object of which is to...	borrowing
9	blue state	noun, adjective	-2000	3	A state (projected to be) won by the Democratic...	compound
10	teleserie	noun	-2000	3	A television soap opera.	borrowing
11	downvote	noun	-2000	3	In the context of social media: an instance of...	conversion
12	butterfly ballot	noun	-2000	3	A machine-readable ballot paper having the names...	compound
13	video blog	noun	-2000	3	A blog primarily composed of video postings...	compound
14	WWJD	interjection	-2000	3	'What Would Jesus Do?'; used among some...	initialism
15	bling	verb	-2000	3	intransitive. To glitter with or like bling; to...	conversion
16	EESC	noun	-2000	3	European Economic and Social Committee; = ESC, n.	initialism
17	G2C	noun	-2000	3	Government to citizen (also consumer or...)	initialism
18	geocaching	noun	-2000	3	An activity or pastime in which an item, or a...	compound
19	mobile health	noun	-2000	3	Health and medical services provided and accessed...	compound
20	nutrigenomics	noun	-2000	3	With singular agreement. A branch of genomics...	compound
21	speed dating	noun	-2000	3	A process by which people seeking romantic...	compound
22	Anthropocene	noun, adjective	-2000	3	Chiefly with the. The epoch of geological time...	compound
23	boba	noun	-2000	3	A cold drink of Taiwanese origin which typically...	borrowing
24	deep web	noun	-2000	3	The part of the World Wide Web that is not...	compound

Pivot Table



References

- Bandyopadhyay, Saptarashmi, Jason Xu, Neel Pawar, and David Touretzky. 2022. "Interactive Visualizations of Word Embeddings for K-12 Students." *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (11): 12713–20. <https://doi.org/10.1609/aaai.v36i11.21548>.
- Bauer, Laurie. 2022. *An Introduction to English Lexicology*. Edinburgh University Press.
- Di Sciullo, Anna Maria, and Edwin Williams. 1987. *On the Definition of Word*. Linguistic Inquiry Monographs 14. Cambridge, Massachusetts: MIT Press.
- Jackson, Howard. 1988. *Words and Their Meaning*. 1st ed. Learning about Language. London: Longman.
- Lipka, Leonhard. 1992. *An Outline of English Lexicology*. Forschung Und Studium Anglistik. Tübingen: Niemeyer.
- Sajous, Franck, Basilio Calderone, and Nabil Hathout. 2020. "ENGLAWI: From Human- to Machine-Readable Wiktionary." In *Proceedings of the 12th Language Resources and Evaluation Conference*, 3016–26. Marseille, France: European Language Resources Association.
<https://www.aclweb.org/anthology/2020.lrec-1.369>.
- Stella, Massimo. 2019. "Modelling Early Word Acquisition Through Multiplex Lexical Networks and Machine Learning." *Big Data and Cognitive Computing* 3 (1): 10.
<https://doi.org/10.3390/bdcc3010010>.
- Stroop, J. Ridley. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18 (6): 643–62.