

# Lexical Innovation

*Seminar 'Lexicology'*

Quirin Würschinger, LMU Munich

July 9, 2025

# Outline

- **Warm-up:** Innovation in our lifetime
- **Cultural-linguistic innovation:** The interplay of society and language
- **S-curve model:** Diffusion and change patterns
- **EC-Model:** Entrenchment and conventionalisation
- **Diffusion pathways:** Variation and spread
- **Operationalisation:** Frequency measures and corpus methods
- **Practice:** Studying lexical innovations in corpora
- **Summary:** Key takeaways

# Recap: From words to innovation

**Recap questions** (2 minutes in pairs):

1. **Word formation processes:** Which processes create completely new word forms?
2. **Semantic change:** How can existing words develop new meanings?
3. **Dictionary evidence:** How do dictionaries track and document new words?

**Today's focus:** How do these processes work in real time? What drives innovation and how does it spread?

# Research foundation

# Social networks of lexical innovation

**Würschinger (2021)**: *Social Networks of Lexical Innovation. Investigating the Social Dynamics of Diffusion of Neologisms on Twitter*

## *Research approach*

- **Longitudinal study**: 99 English neologisms tracked on Twitter
- **Multi-method analysis**: Frequency measures + social network analysis
- **Temporal dynamics**: Usage patterns, volatility, and diffusion over time
- **Social pathways**: How innovations spread through communities

## *Key insights*

- **Frequency alone is insufficient** - need temporal and social information
- **Social diffusion varies significantly** - even words with similar frequency show different patterns

# Today's session structure

We'll explore the theoretical framework and empirical findings:

1. **Theoretical models**: S-curve and EC-Model for understanding innovation
2. **Operationalisation**: How to measure diffusion using corpus methods
3. **Practical application**: Hands-on analysis with case study words
4. **Comparative analysis**: Your findings vs. published research

**Goal**: Understand both the theory and practice of studying lexical innovation

The interplay of cultural and  
linguistic innovation

# Society and language change

**Key principle:** Society continually changes as new practices and products emerge.

- These changes typically **first manifest themselves in language** on the level of lexis.
- Language change occurs through **neologisms** (new words or new meanings).
- **Knowledge of words is conventional:** speakers learn form-meaning pairings.

# Types of neologisms

## Formal neologisms

- *smartphone*
- *iPhone*
- *Covid*

## Semantic neologisms

- *hotspot* (from physical location → infection cluster)
- *Querdenker* (lateral thinker → conspiracy theorist)
- *social distancing*
- *spreader* (person → virus transmitter)

# Lexical renaissance

**Innovation after lexical attrition:** old words returning to use.

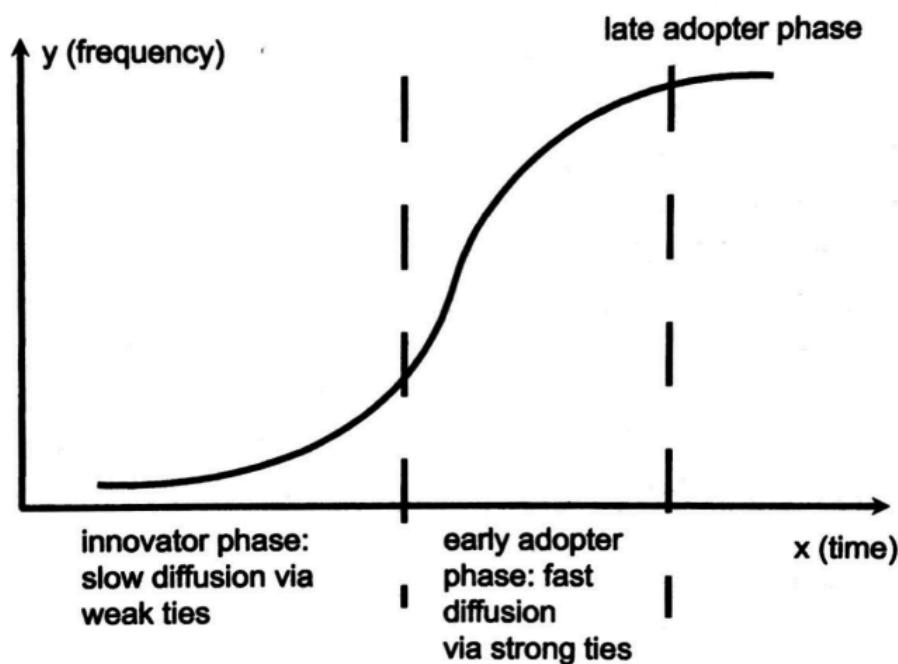
## Examples:

- *dotard* (revived during political discourse)
- *to furlough* (returned during economic crises)

Pattern: Social changes can revive forgotten vocabulary when circumstances make old words newly relevant.

The S-curve model of lexical innovation

# Integration of diffusion stages



S-curve model: Integration of Milroy's and Rogers' model of diffusion stages ([Kerremans 2015, 65](#))

**Early adoption:** Innovative users introduce new words.

**Acceleration:** Rapid spread through social networks.

**Deceleration:** Slower growth as market saturates.

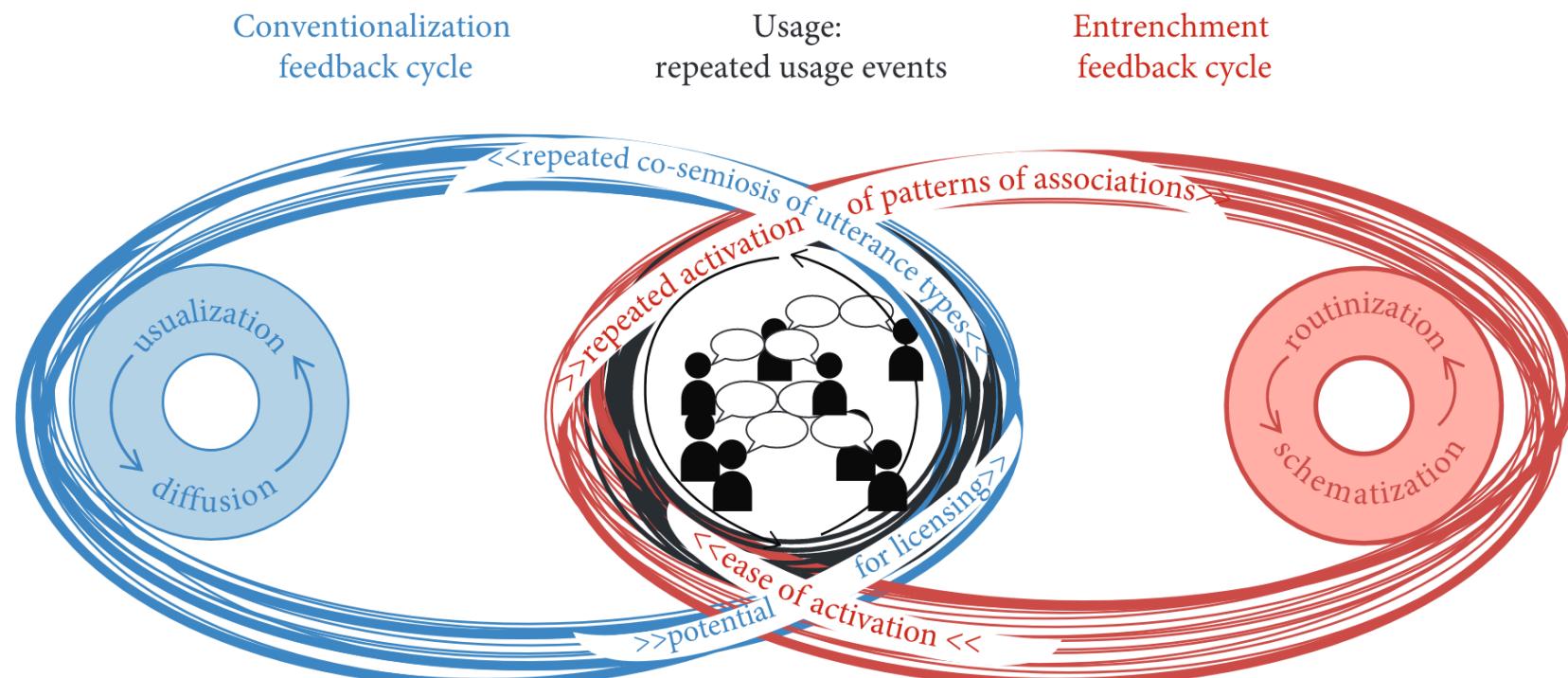
**Stabilisation:** Word becomes established in conventional usage.

# EC-Model: Entrenchment and conventionalisation

# The dynamics of the linguistic system (Schmid 2020)

**Conventionalisation:** part of the conventional language system.

**Entrenchment:** storage in speakers' mental lexicon.

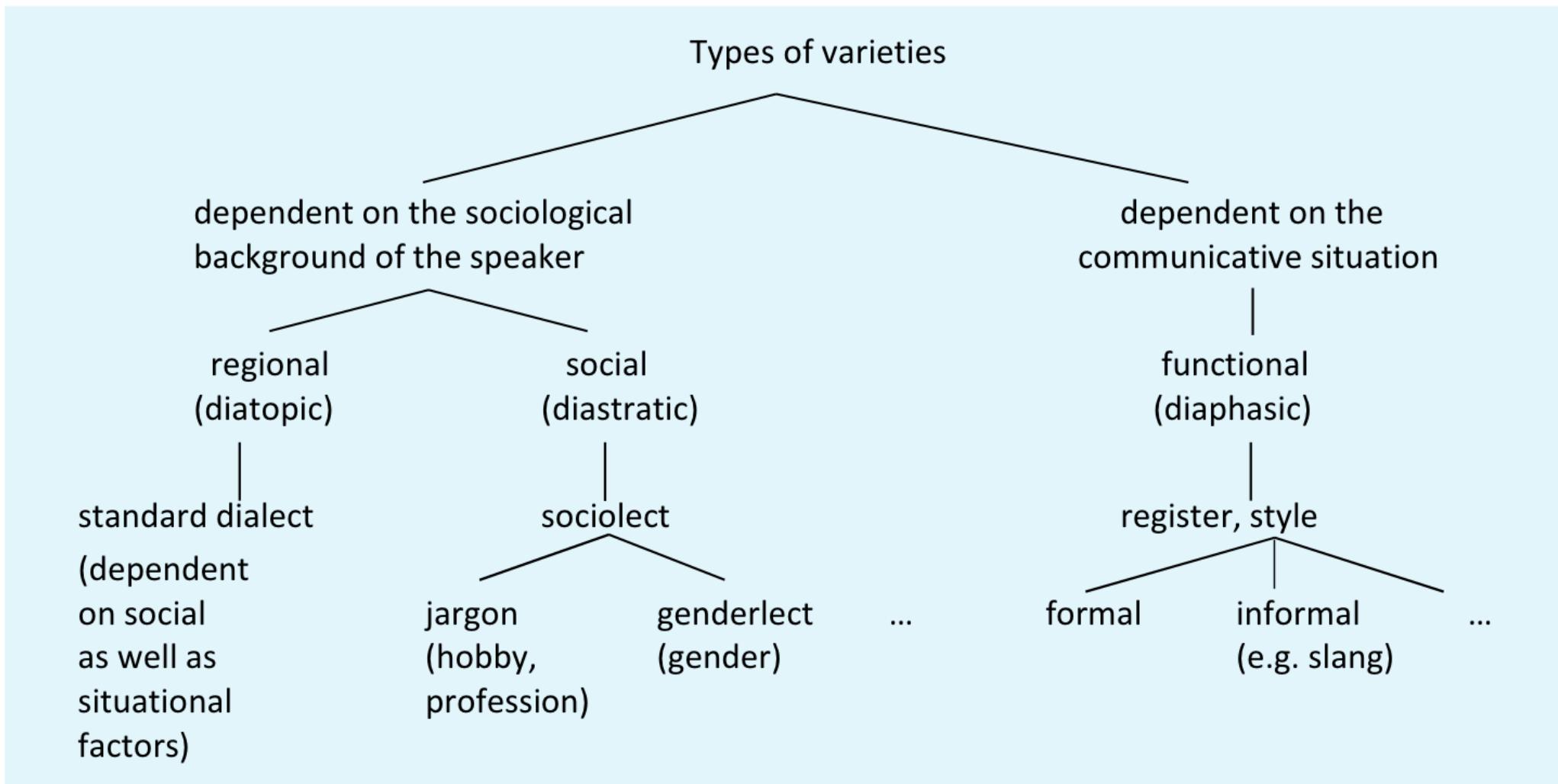


**Figure 1.2** Language as a simple Tinguely machine

The interplay between usage, entrenchment, and conventionalisation (Schmid 2020, 4).

# Pathways of diffusion

# Types of linguistic variation and diffusion



Types of linguistic variation: diatopic, diastratic, and diaphasic (Kortmann 2020, 204)

# Dimensions of diffusion

## *Across speakers and communities*

- Regional spread
- Social group adoption
- Age group patterns

## *Across text types*

- Formal vs informal contexts
- Academic vs popular writing
- Spoken vs written language

## *Examples*

usage contexts	low	high	hypostatization	electron DNA
			alt-left	DNA
	low	high	speakers	

Operationalisation: frequency  
as indicator

# Theoretical foundation

**Frequency** as indicator for entrenchment and conventionality  
[\*\*\(Stefanowitsch and Flach 2017\)\*\*](#):

- **Corpus-as-input**: language in corpora represents potential exposure to speakers.
- **Corpus-as-output**: language used by speakers represents degrees of entrenchment.

# Frequency measures

Examples from Würschinger (2021):

*Total frequency patterns*

Most frequent:

**TABLE 1** | Total usage frequency (FREQ) in the corpus. Most frequent lexemes.

Lexeme	FREQ
tweeter	7,367,174
fleek	3,412,807
bromance	2,662,767
twitterverse	1,486,873
blockchain	1,444,300
smartwatch	1,106,906

## Around the median

---

**TABLE 2 |** Total usage frequency (FREQ) in the corpus. Examples around the median.

---

Lexeme	FREQ
white fragility	26,688
monthiversary	23,607
helicopter parenting	26,393
deepfake	20,101
newsjacking	20,930
twittosphere	20,035

---

## Least frequent

---

**TABLE 3 |** Total usage frequency (FREQ) in the corpus. Least frequent lexemes.

---

Lexeme	FREQ
microflat	426
dogfishing	399
begpacker	283
halfologue	245
rapugee	182
bediquette	164

---

## Case study selection

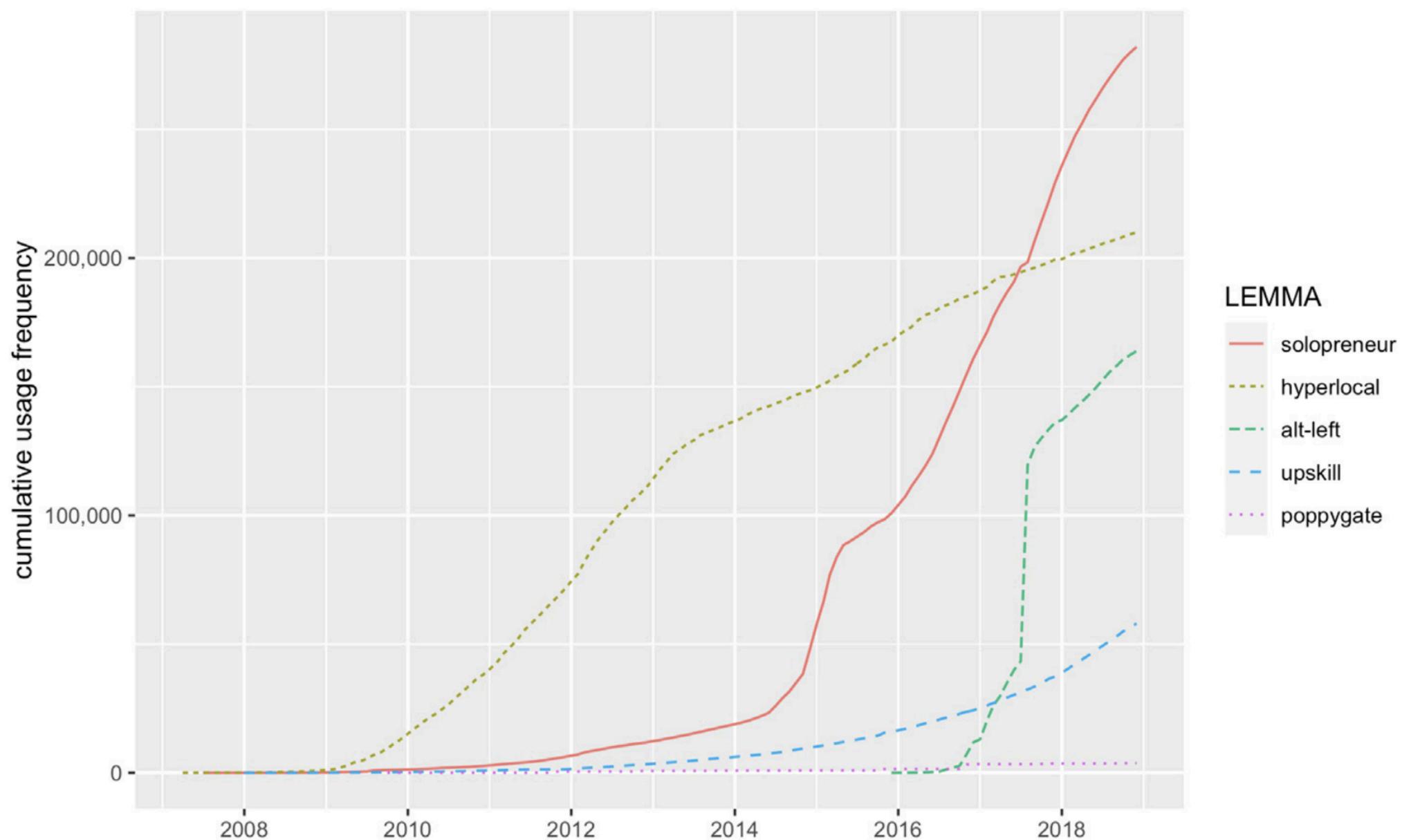
---

**TABLE 4 |** Total usage frequency (FREQ) in the corpus. Case study selection.

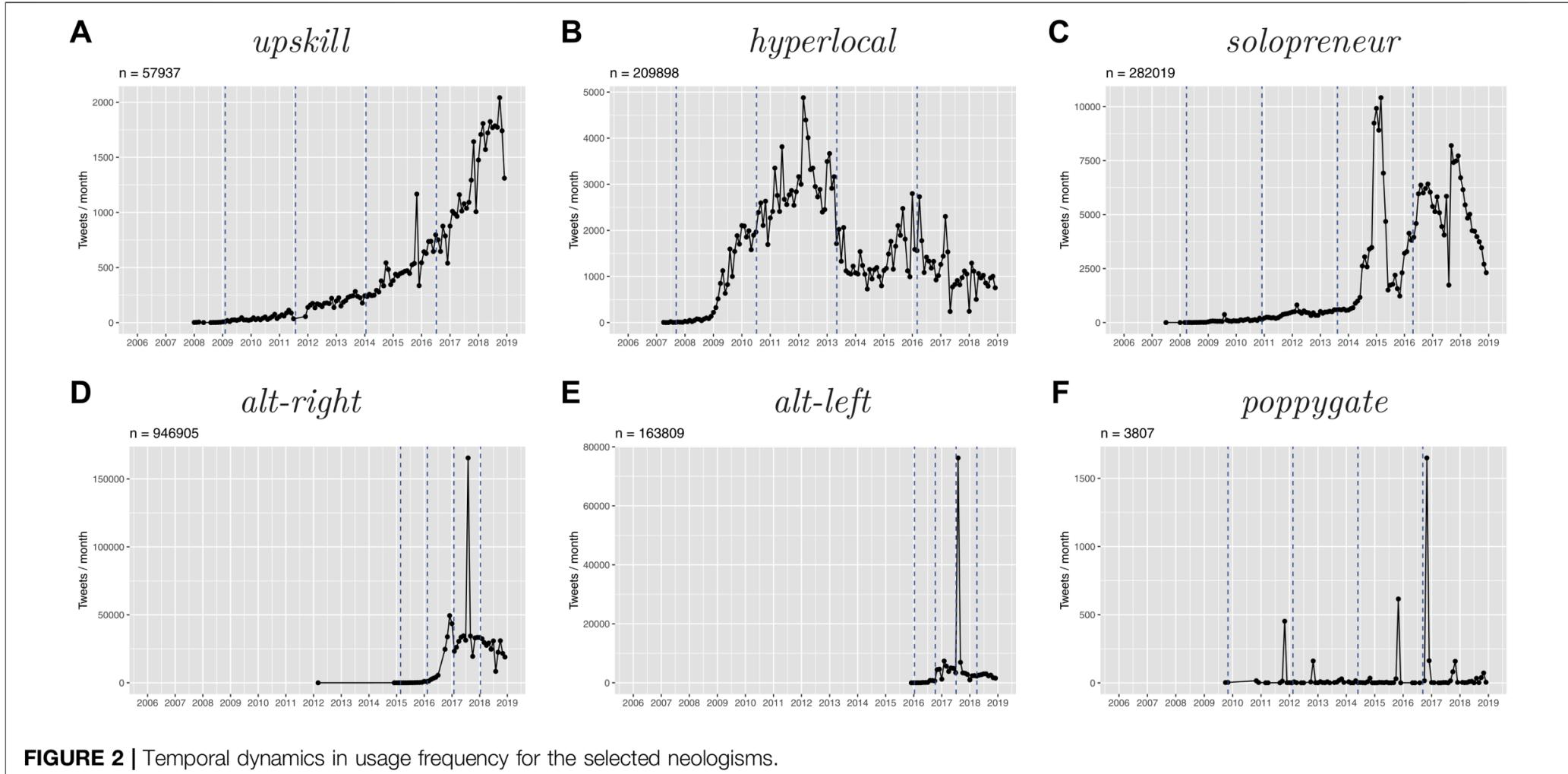
Lexeme	FREQ
alt-right	1,012,150
solopreneur	282,026
hyperlocal	209,937
alt-left	167,124
upskill	57,941
poppygate	3,807

# Frequency over time (diachronic analysis)

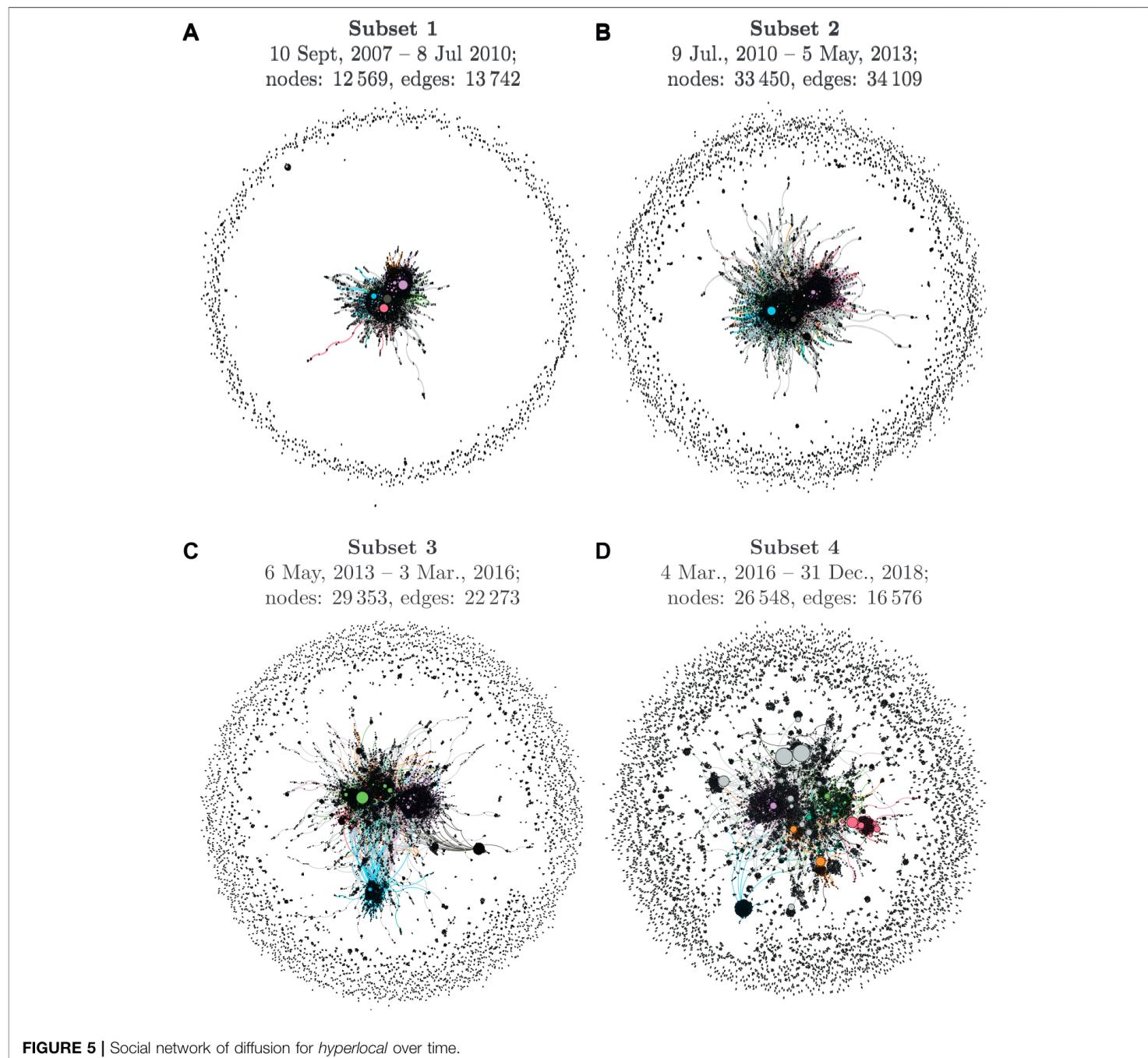
## *Cumulative frequency*



## Raw frequency

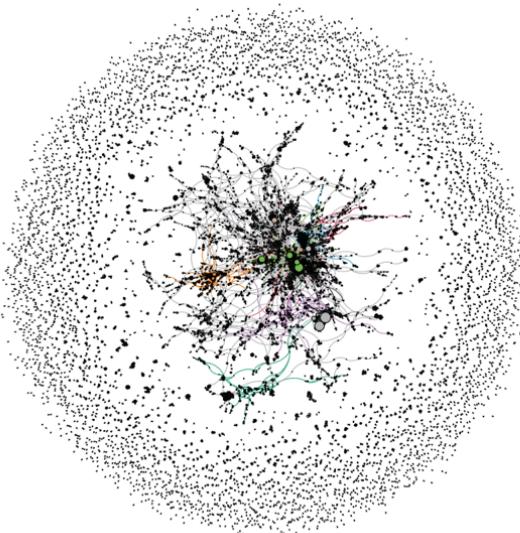


# Diffusion over time

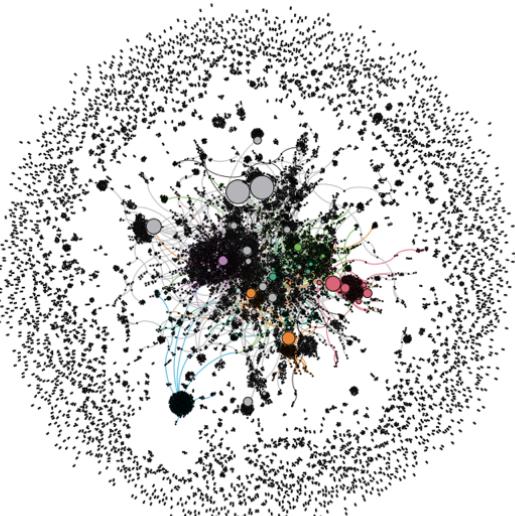


# Diffusion across communities

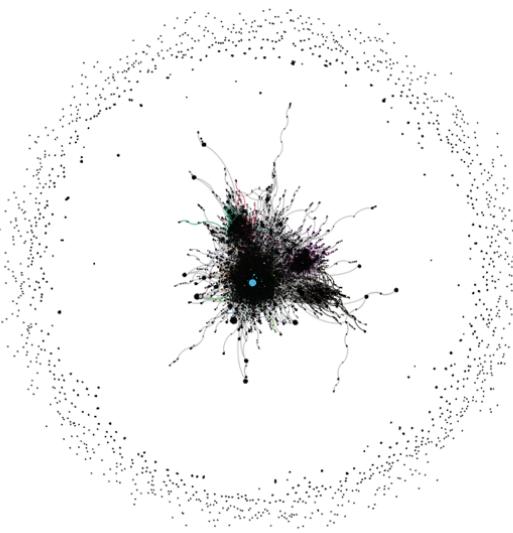
**A** *upskill*  
(37 044 nodes, 23 060 edges)



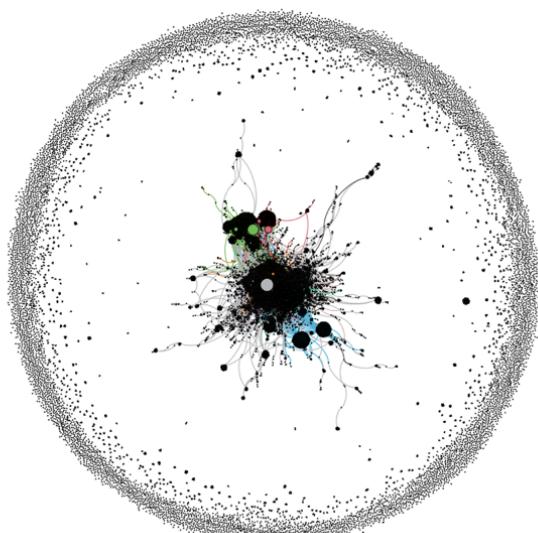
**B** *hyperlocal*  
(26 548 nodes, 16 576 edges)



**C** *alt-left*  
(26 367 nodes, 32 836 edges)

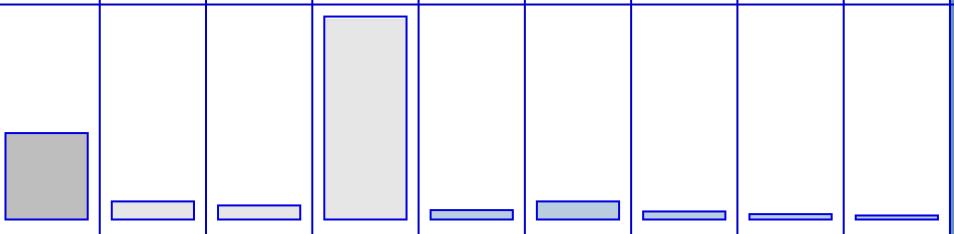


**D** *solopreneur*  
(24 585 nodes, 20 486 edges)

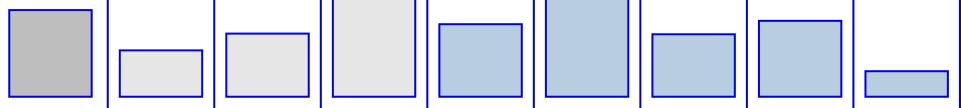


# Diffusion across text types

*bro* in COCA

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	7272	453	328	5610	204	421	160	89	7
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	7.32	3.52	2.64	43.80	1.62	3.56	1.27	0.73	0.06
SEE ALL SUB- SECTIONS AT ONCE									

*brother* in COCA

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	118471	7787	10361	35309	12205	26194	10489	12270	3856
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	119.30	60.55	83.39	275.69	96.76	221.38	83.19	100.79	32.19
SEE ALL SUB- SECTIONS AT ONCE									

Practice: identifying pattern of lexical innovation

# OED's Advanced Search

Use the OED's Advanced Search to retrieve all words that have first been used since 2000.

**SEARCH TERMS**

Headword     Quotation text  
 Definition     Quotation author  
 Etymology     Quotation work  
 Forms  
  
 Exact match

**DATE OF USE**

Use     First use

From   
To

**Add term**    **Update**

1 to 50 of 697 results    **Save search**    **1** **2** **Next >>**

Sort by **Frequency**     [Export results \(.csv\)](#)

2020-	<b>Covid-19, n.</b> An acute disease in humans caused by a...
2020-	<b>Covid, n.<sup>2</sup></b> The disease Covid-19; (also) the coronavirus...
2008-	<b>Bitcoin, n.</b> A type of digital currency introduced in 2009...
2002-	<b>selfie, n.</b> A photograph that one has taken of oneself, esp...
2001-	<b>blog post, n.</b>

Export the results as a [CSV](#) file.

# Excel analysis

Model sheet:

<https://1drv.ms/x/c/9a2ec97d593520f9/EbTMdPaevpFJqGqbQT43I3YBf2Pg0>

1. import the **csv** file into Excel
2. create a **Table** that spans the imported data (without metadata)
3. created **Pivot Tables** to analyse the data

Analysis:

1. What are the most frequent word classes among the target neologisms?
2. What is their distribution across dates of first use?
3. Which word-formation processes are most common?
4. Which subject areas are most prevalent?

# Sketch Engine analysis

Use Sketch Engine and the BNC 2014 Spoken to analyse the frequency and distribution of the target neologisms.

1. In Excel, selected a sample (e.g. **n=20**) of the target neologisms based on frequency (column **Frequency Band Number**).
2. In Sketch Engine, search for all of these neologisms in one go by concatenating them in a single CQL query. For example:

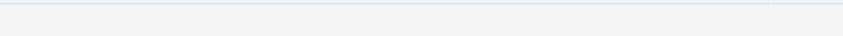
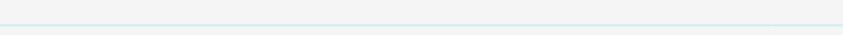
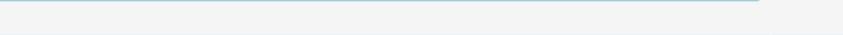
```
[word="retweet|hashtag|podcast|stan|Covid-19|photobomb|upvote|Covid|deadname|Bitcoin|stablecoin|selfie|SARS|open-world|vlog|vape"]
```

Analyse the results by using the **Frequency** tool.

The screenshot shows the CONCORDANCE interface from the British National Corpus 2014 (BNC2014). The left sidebar contains various icons for different tools. The main header says "CONCORDANCE" and "British National Corpus 2014 (BNC2014...)" with a search icon. Below the header are several buttons: "Get more space +", a refresh icon, a link icon, a help icon, a message icon, and a user profile icon. A tooltip in the center says "CQL [word='retweetlhashtaglpodcastlstanlCovid-19lphotob... • 184" and "15.55 per million tokens • 0.0016%". The toolbar below includes icons for search, download, export, eye, question mark, and other analysis options. A "KWIC" button is highlighted in red. A modal window titled "FREQUENCY" is open, showing tabs for "BASIC", "ADVANCED" (which is selected), and "LEARN". It asks "Select an attribute and its position in the concordance: ?" and has a "word" input field with a magnifying glass icon. Below it are "left context" and "right context" sections with numbered boxes (1-6) and a "KWIC" dropdown menu.

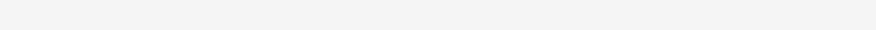
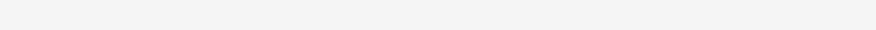
# 1. Frequency by word:

(8 items, 184 total frequency)

	Word	Frequency	Relative ?	% Of conc. ?	
1	<input type="checkbox"/> selfie	65	5.49	35.33 %	 ...
2	<input type="checkbox"/> hashtag	54	4.56	29.35 %	 ...
3	<input type="checkbox"/> podcast	39	3.30	21.20 %	 ...
4	<input type="checkbox"/> retweet	10	0.85	5.43 %	 ...
5	<input type="checkbox"/> SARS	6	0.51	3.26 %	 ...
6	<input type="checkbox"/> stan	4	0.34	2.17 %	 ...
7	<input type="checkbox"/> vlog	3	0.25	1.63 %	 ...
8	<input type="checkbox"/> vape	3	0.25	1.63 %	 ...

## 2. Frequency by Age range:

(9 items, 184 total frequency)

	Age range ↓	Frequency	Relative density ?	
1	<input type="checkbox"/> ==NONE==	1	79.74 %	 ...
2	<input type="checkbox"/> 0-10	4	168.93 %	 ...
3	<input type="checkbox"/> 11-18	52	463.15 %	 ...
4	<input type="checkbox"/> 19-29	57	84.17 %	 ...
5	<input type="checkbox"/> 30-39	20	74.92 %	 ...
6	<input type="checkbox"/> 40-49	17	65.20 %	 ...
7	<input type="checkbox"/> 50-59	23	123.50 %	 ...
8	<input type="checkbox"/> 60-69	7	41.01 %	 ...
9	<input type="checkbox"/> 70-79	3	32.44 %	 ...

# Summary

1. **Cultural and linguistic innovation are intertwined:** societal changes drive lexical innovation.
2. **S-curve model** describes diffusion patterns from innovation to stabilisation.
3. **EC-Model** explains entrenchment and conventionalisation through frequency.
4. **Diffusion occurs across multiple dimensions:** speakers, communities, and text types.
5. **Corpus methods enable empirical study** of innovation and diffusion patterns.

# References

- Kerremans, Daphné. 2015. *A Web of New Words*. Bern: Peter Lang. <https://doi.org/10.3726/978-3-653-04788-2>.
- Kortmann, Bernd. 2020. *English Linguistics: Essentials*. J.B. Metzler.
- Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System. Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.
- Stefanowitsch, Anatol, and Susanne Flach. 2017. "The Corpus-Based Perspective on Entrenchment." In *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*, edited by Hans-Jörg Schmid, 101–128. Boston, USA: American Psychology Association and de Gruyter Mouton.
- Würschinger, Quirin. 2021. "Social Networks of Lexical Innovation. Investigating the Social Dynamics of Diffusion of Neologisms on Twitter." *Frontiers in Artificial Intelligence* 4: 106. <https://doi.org/10.3389/frai.2021.648583>.