Social networks of lexical innovation

Investigating the diffusion of neologisms on Twitter

Quirin Würschinger
LMU Munich

27th March 2020

Contents

List of Figures

List of Tables

Abstract

Societies constantly evolve and produce innovative products and practices. Speakers invent new words to communicate efficiently. The resulting lexical innovations diffuse in social networks, some of them fail to catch on while other become part of the lexicon.

1 Introduction

The coronavirus has recently spread with shocking speed and has tragically affected the lives of people around the world. Its fatal consequences have demonstrated the devastating power of exponential diffusion in social networks. A leading research group of biologists analysing the contagion of Covid-19 have shown that the virus spreads supra-exponentially and that it resembles the diffusion of cultural and linguistic innovations such as internet memes. (editorial staff, 2020) Does this confirm the popular perception that certain cultural and linguistic innovations 'go viral'?

state gist of the paper

Societies constantly evolve, new products and practices emerge and speakers continually invent and adopt new words which diffuse through social networks of communicative interaction. Influential models in sociolinguistics like the S-curve model (J. Milroy, 1992) share fundamental features with economic models (Rogers, 1962) of diffusion and show commonalities between the spread of cultural and linguistic innovations. These models assume that the diffusion of innovations across social networks follows universal trajectories and that rates of spread depend on sociolinguistic dynamics such as network density and the presence or absence of weak ties (Granovetter, 1977). Unlike research on biological and cultural diffusion processes, sociolinguistic research has only fairly recently been provided with data sources that are equally suitable for large-scale, data-based approaches that use social network analyses to study these phenomena empirically.

Social media platforms like Twitter have changed the way we communicate and interact and how information spreads and they offer big amounts of data for empirical research. Sociological research has recently been concerned with pressing issues regarding the impact of online social networks for the spread of hate speech, fake news and the power of 'influencers', bots and institutions on public opinions and elections, which increasingly strain the social fabric. For (socio-)linguists, social media data provide huge amounts of authentic text along with information about speakers and their interactions which opens up new possibilities for the empirical study of language variation and change as well as the diffusion of linguistic innovations.

advantages of going beyond frequency

The size of these data as well as their informal nature allow corpus studies on the use and spread of new words which are often propagated via social media channels. In contrast to modern web corpora, which share these features, the availability of metadata about speakers and their interactions has been particularly important for the advancement of sociolinguistic research. Recent sociolinguistic work on lexical innovation has, for example, used data from social media users to gain insights about general trajectories (Nini, Corradini, Guo & Grieve, 2017) and geographical patterns (Eisenstein, O'Connor, Smith & Xing, 2014; Grieve, 2017, 2018) of diffusion and about factors that influence whether new words spread successfully (Grieve, forthcoming).

Social network analysis analyses interactional patterns between language users and allows to zoom in on the sociolinguistic dynamics of diffusion to see, for examples, whether the adoption of new words remains limited to closely connected sub-communities, whether they reach bigger parts of the speech community and whether certain individuals or groups influence this process. While the study of social networks has had a long research tradition in sociolinguistics and has shaped current models of diffusion (e.g. J. Milroy and Milroy, 1985), new data sources have only fairly made it feasible to conduct large-scale empirical analyses using this approach. Advances in data and methodology of this kind put computational sociolinguistics in an excellent position to tackle old research questions in new ways, to test long-standing theoretical models empirically and to explore new questions and insights about language and society.

Social network analyses of social media data have been successfully employed in diverse fields, for example to study the spread of diseases (Lu et al., 2018), opinions (West, Paskov, Leskovec & Potts, 2014) and political attitudes (Pew Research Center, 2019). So far the application of similar approaches to the study of lexical innovation remains scarce, although recent work shows promising results in testing mathematical models for the diffusion of new words in social networks. (Goel et al., 2016)

This paper aims to investigate the spread of new words on the social media platform Twitter, using social network analysis to zoom in on the sociolinguistic dynamics of diffusion.

ask some concrete questions: e.g. 'Who uses the word *ghosting*?'

2 Modeling the diffusion of lexical innovations

How do new words diffuse and become conventional lexical items in a language system?

2.1 Research perspectives

A substantial body of linguistic research has tackled this question from different **perspectives**. (Schmid, 2016, p. 16)

From a **structural** perspective, main areas of interest include which word-formation processes are involved in forming new words, whether they are formally and semantically transparent, whether they show variation and change in the process of lexicalization and which status the resulting neologisms have in the language system (institutionalization). (e.g. Bauer, 1983; Lipka, 2005

Cognitive perspectives focus on how individuals process and store lexical innovations. Speakers generally use new words when they experience a communicative need to talk about entities or practices that cannot be expressed by their language's inventory of conventional words yet. In order for neologisms to successfully diffuse, speakers need to successfully negotiate their meaning (co-semiosis) in discourse, others need to adopt the behaviour of using these words (co-adaption). Continued exposure and use of new words can then lead to the entrenchment of new words in the mental lexicon of speakers. (Schmid, 2008)

Sociolinguistic perspectives transcend the level of the individual to study the diffusion of new words across speakers. The diffusion of lexical innovations is commonly seen as successful when the majority of the speech community has accepted a new word as a conventional lexical unit that can and is being used in communicative practice.

2.2 The S-curve model

S-curve models of linguistic change (Labov, 2007; J. Milroy, 1992; Nevalainen, 2015) assume universal sociolinguistic dynamics for the diffusion of linguistic innovations.

- The **trajectory** of spread is expected to follow an S-curve shape, with low rates of diffusion in early stages, followed by a period of accelerating spread with a tipping point at the mid point in the diffusion curve after which diffusion slows down and the curve flattens towards the end of the diffusion process.
- These temporal trajectories are assumed to correspond to the **sociolinguistic dynamics** of which individuals and groups interact with each other and adopt the target innovation.
  - In the **first stage** of slow diffusion only a small number of early adopters take up the innovative words. The individuals who use the new word typically form dense networks connected by strong ties. The structure of tight-knit communities of potentially like-minded individuals of similar backgrounds facilitates the successful negotiation of meaning (co-semiosis) of new words. High rates of interactions in these communities leads to high rates of exposure for individuals, which fosters co-adaption, entrenchment and usualization of new words in these communities.
  - In cases of successful diffusion the initial stages are followed by an **acceleration in spread** when new words increasingly reach speakers outside these tight-knit communities via weak ties (Granovetter, 1977). Rates of diffusion increase substantially when speakers that are not part of the initial group of early adopters start to accomodate the new words, allowing the innovations to reach a broader spectrum of the speech community.
  - In **later stages**, rates of diffusion slow down again as the majority of the speech community has already adopted the new words while smaller pockets speakers remain reluctant to take up the new words.
- S-curve models have mainly been applied to the **linguistic domains** of phonology and syntax. Fundamental differences between lexemes and linguistic items on other levels such as phonemes and grammatical constructions might affect the validity and reliability of such models for *lexical* innovation.
  - For example, grammatical constructions such as the *going to* future used to express a speaker's future intention serve to fulfill relatively abstract communicative needs that remain stable over time. By contrast, on the lexical level,

linguistic innovations are typically tied to concrete cultural referents such as products and practices whose conceptual relevance is much more volatile over time. For example, many lexical innovations such as *millennium bug* denoting the fear of a computer crash at the beginning of the new millennium can show high rates of diffusion and become entrenched and conventional among the majority of the speech community. Without continual conceptual relevance in public discourse, however, these words fail pass on to the next generation of speakers. S-curves are commonly assumed to be found when linguistic innovations compete for **'semantic carrying capacity'** (Nini et al., 2017), however, in many if not most cases of *lexical* innovation the conceptual carrying capacity is far from stable over time which represents a critical deviation from the traditional assumptions behind S-curves in language change.

- However, the generally strong theoretical and empirical basis of the S-curve model for language innovation and change, also from studies on the diffusion of cultural innovations (Rogers, 1962), and the precise formulation of the sociolinguistic dynamics underlying different phases of diffusion still make it an attractive **blueprint** for the empirical study of the sociolinguistic diffusion of lexical innovations.

2.3 Current framework: the EC-Model (Schmid 2020)

I use the Entrenchment-and-Conventionalization-Model (Schmid, 2020) as a framework for modelling the diffusion of lexical innovations.

The EC-Model provides an approach integrating both structural, cognitive and sociolinguistic perspectives on the diffusion of lexical innovations.

The model also differentiates between the level of the individual ('entrenchment') and the community ('conventionalization').

Here I will only briefly outline the most important concept relevant for studying the sociolinguistic aspects of diffusion here.

Conventionalization:

definition: 'Conventionalization is the continual process of establishing and re-adapting regularities of communicative behaviour among the members of a speech community, which is achieved by repeated usage activities in usage events and subject to the exigencies of the entrenchment processes taking place in the minds of speakers.' (Schmid, 2020)

Usualization 'Usualization can therefore be defined as a process that establishes, sustains, and changes regularities of behaviour with regard to co-semiotic mappings between forms and meanings or functions and communicative goals and linguistic forms. It affects the semasiological, onomasiological, syntagmatic, cotextual, and contextual dimensions of conformity behind conventionality and is relative to communities.' (Schmid, 2020)

Diffusion 'Linking the three aspects of speakers, cotexts, and contexts, I define diffusion as a process that brings about a change in the number of speakers and communities who conform to a regularity of co-semiotic behaviour and a change in the types of cotexts and contexts in which they conform to it.' (Schmid, 2020)

less relevant for sociolinguistic aspects and for this study cotexts contexts
According to the EC-Model, for studying the sociolinguistic aspects of diffusion, investigating 'changes in the number of speakers and communities' is thus essential.

3 Measuring the diffusion of lexical innovations

3.1 Previous approaches

- **before frequency** Empirical approaches studying the diffusion of lexical innovations have only recently become feasible with the advent of new data sources and computational methods.
- Earlier work had to rely on **traditional linguistic corpora**. Due to the low-frequency nature of neologisms general linguistic corpora do not allow to study broad ranges of neologisms and thus pose limits to making broad and robust generalizations about the nature lexical innovation. Despite these limitations, case studies on selected neologisms (Hohenhaus, 2006) and studies on specific domains of neology (Elsen, 2004) managed to shed light on the spread of new words in more specific domains.
- The advent of **web corpora** in the last two decades has provided researchers with bigger and less formal data to study lexical innovation.
  - The sheer size of big corpora → bigger samples
  - monitoring corpora (Davies, 2013): tracking dynamics of diffusion, closer to coining
  - In particular, a range of tools enabled the creation of specialized corpora for the investigation of neologisms. (Renouf, Kehoe & Banerjee, 2006; Kerremans, Stegmayr & Schmid, 2012; Lemnitzer, n.d.; Gérard, 2017; Cartier, 2017)
  - the nature of web corpus data is particularly suitable for investigating lexical innovations as
    - language on the web is very creative,
    - more informal sources, bigger spectrum of language use
    - new words often first occur on the web
    - and use on the web significantly influences whether these new formations catch on or not.
  - Web corpora thus promise insights into diffusion across
    - contexts: e.g. whether new words such as *blockchain* are increasingly used in less formal
    - cotexts: e.g. whether new words such as are increasingly used in more divers
- social media corpora Grieve, Nini and Guo, 2016; Eisenstein et al., 2014

    - size
    - nature: similar to web corpora
        * creative, hotbed
        * authentic language use
        * driving force
    - social network information
- **social network analysis** getting at 'changes in the numbers and communities of speakers' by using social network information
    - number of users: active and passive
    - interactions between users: influencers
    - network properties: density, centralization

3.2 Going beyond frequency

The conventionality of linguistic units is commonly assessed by counting how often they are found to be used in linguistic corpora, with high frequencies of occurrence seen as indicators of high levels of conventionality. Diffusion as a process that drives increasing conventionalization is thus commonly assumed to be reflected by increases in the usage frequency of linguistic innovations. Previous research on lexical innovation has been largely limited to this approach and has evaluated the spread and the overall success of new words on the basis of the number of tokens found in linguistic corpora. This paper takes usage frequency as a baseline and uses social network analysis to go beyond frequency to discover sociolinguistic dynamics of diffusion and conventionality that have eluded previous frequency-based approaches.

Frequency measures are widely used to study linguistic phenomena on all levels, from investigating phonological preferences between communities, to studying the increasing establishment of grammatical constructions like the *going to*-future over time, to assessing the degree to which words are conventional lexical units of a language. Usage frequency is thus commonly used by a diverse set of linguistic sub-disciplines. From a structural perspective, for example, co-occurrence frequencies of multi-word units such as *handsome man* are taken as an indicator for whether these are free combinations or more or less fixed collocations in a language system. Historical linguistics investigates phenomena like language change and grammaticalization, by analysing changes in usage frequency of certain constructions like the *going to*-future over time. Cognitive and psycholinguistic research commonly relies on frequency measures to approximate the degree to which speakers are familiar with words that are presented as linguistic stimuli in experiments to control for effects on experimental results.

The reliance on usage frequency as a measure for different phenomena in these diverse research contexts has faced substantial criticism. Stefanowitsch and Flach, 2017 provide a good overview of the theoretical assumptions and problems that underlie frequency-based approaches in corpus linguistics.

(1) highly socially indexical and thus especially prone to be used only by certain subcommunities, (2) topical which makes freq. less reliable bc. it fails to capture 'dormant' passive knowledge of the words

freq. esp. insufficient for lex. inn.

When assessing the suitability of usage frequency as a measure for the diffusion and conventionality of neologisms a set of assumptions underlying the frequency-based approach need to be disentangled. While these theoretical and methodological considerations generally apply to all corpus-linguistic work, the focus will be on the current issue of lexical innovation.

I adopt Schmid's EC-Model (Schmid, 2020) as a framework for defining and delimitating the concepts of 'conventionalization' and 'diffusion'.

> [...] I define diffusion as a process that brings about a change in the number of speakers and communities who conform to a regularity of co-semiotic behaviour and a change in the types of cotexts and contexts in which they conform to it.

Aside from the sociolinguistic perspective on diffusion ('number of speakers and communities'), Schmid conceptualizes diffusion as a multi-dimensional process that also takes into account changes on the syntagmatic ('cotexts') and pragmatic ('contexts') level. This paper will focus on the sociolinguistic dimension of diffusion and will leave an integrative approach including all three perspectives for further research.

Applying this definition to the context of lexical innovation thus implies that the successful diffusion of a new word is marked by it being known and used by an increasing 'number of speakers and communities'.

By contrast, in a strict sense, usage frequency counts of a lexeme represent the total number of tokens produced by *all* speakers who have *contributed* to the *target text corpus*. The discrepancy between the theoretical definition of diffusion adopted here and the exact information contained by frequency show that this operationalization relies on a number of assumptions that let it only approximate the construct to be measured.

Firstly, usage frequency does not provide direct information as to how many *individual speakers* have used a new word. Especially in the case of neology, there are certain new words that are disproportionally used and propagated by a relatively small, but more active and dedicated users of the new term. This leads to high overall frequency counts which falsely suggest that larger parts of the speech community have adopted the term.

Secondly, usage frequency only captures active uses of a term and fails to include how many speakers have been passively exposed to neologisms. In the context of entrenchment, Stefanowitsch and Flach (2017) refer to this problem as the 'corpus-as-input' and 'corpus-as-output' hypothesis. The underlying assumption is that the output of the speakers who have contributed to a corpus can serve as an approximation for the potential linguistic input of a comparable speaker group. Frequency thus reflects the 'usage intensity' of neologisms in the speech community which indicates the degree of entrenchment in individual speakers as well as an approximation of the conventionality of the neologisms in the speech community. In the case of lexical innovation this can be problematic as questionnaires studies on the use of neologisms (Kerremans, 2015) show that many speakers report that they have come across target neologisms, but have not actively used them in discourse. Relying on frequency counts only can thus often lead to underestimating the degree of diffusion of neologisms.

Thirdly, usage frequency fails to capture where new words diffuse across 'communities of speakers', as suggested by Schmid's definition. This is, of course, a consequence of the fact that frequency counts cannot provide direct information about the number of speakers involved in the diffusion of neologisms, as was pointed out in the first two points above. New words often stem and quickly spread within tight-knit communities of practice that share common attitudes or interests. Frequency measures alone cannot detect whether neologisms only show increasing usualization within these groups or whether they diffuse and become conventional in other parts of the speech community, which represents an essential feature of the sociolinguistic dimension of diffusion.

how much the words might have diffused outside the *target text corpus*

temporal dynamics (e.g. *millennium bug*)

4 Data

For my empirical study I have selected XXX neologisms to study their diffusion on Twitter from their early attestations to the end of 2018. More information about the sample can be found in Section XXX.

4.1 Twitter data

Twitter[2] is a popular micro-blogging and social media platform that was started in 2006. one of the most popular social media platforms

representativity limited: certain user groups over-represented. good enough: generally okay (Grieve, Montgomery, Nini, Murakami & Guo, 2019)

advantages of data hotbed: influential hotbed for technical, cultural and linguistic innovations longitudinal: retrospective, starting from early attestations big data social network information

scope of the present study: I only study diffusion of new words *on Twitter* considering both frequency and social network characteristics might be more robust and allow more generalization than just frequency

data retrieval I have used the tool `twint` to download all tweets containing the target word as searchable via the Advanced Search Function on Twitter's official website

data overview: table

4.2 Methodology

anatomy of a tweet

post-processing treatment of hashtags filter uses as user handles duplicate removal

[2] http://www.twitter.com

social network analysis software `R`: `tidygraph`, `ggraph` `Gephi` constructing the networks elements: legitimate (Goel et al., 2016) users = nodes interactions (replies, mentions) = edges subsetting: time slices start of diffusion process: >= 3 edges 4 phases of equally sized time windows network metrics centralization network density visualizations

The resulting networks are interactional rather than static. (Goel et al., 2016) This makes them more similar to communities of practice than to traditional sociolinguistic networks based on static speaker characteristics such socio-economic status. In the case of lexical innovation networks that are based on whether speakers provide valuable information. In cases such as *alt-left*, for example, interactional networks show whether usage of the term remains centralized to a tight-knit community of speakers or whether it diffuses to be used by other sub-communities. Whether communities are distinct depends on whether users communicate with each other. While the reasons for theses communicative affiliations remain unknown (age, gender, socio-economic status), they are certainly real in mutually engage in communicative interaction. It would be interesting to complement this information with static information, however such data are currently not obtainable.

4.3 Neologism sample

Full sample

Previous empirical approaches to the spread of neologisms have discovered a number of prototypical pathways of diffusion. (Kerremans, 2015) no diffusion topical recurrent advanced

dimensions

overall degree of diffusion (synchronic): successful vs. unsuccessful: usage frequency, degree centralization no success limited advanced

temporal dynamics of diffusion (diachronic) stability: stable vs. topical coefficient of variation trend increasing: potential diffusion decreasing: potential centralization time window: speed and lifespan

Case studies

criteria covering clusters of neologism candidates frequency counts comparable cases (Kerremans, 2015) no diffusion: *microflat* limited topical: *poppygate* centralized: *alt-left* decreasing: *solopreneur* advanced diffusion: advanced: *upcycling* increasing: *hyperlocal*

5 Frequency

5.1 Total usage frequency

As described in Section XXX, the degree of conventionality is often approximated by a how many times a word has been used in a corpus. A common way to use this information is to rely on cumulated frequency counts which sum up the total number of uses.

The present sample of neologisms covers a broad frequency range. Table 1 presents candidates from four groups: six examples around the minimum, the median and the

Table 1: Total usage frequency in the corpus.

(a) Least frequent neologisms in the sample.

lexeme	freq
microflat	426
dogfishing	399
begpacker	283
halfalogue	245
rapugee	182
bediquette	164

(b) 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

(c) Most frequent examples in the sample.

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

(d) Case study selection.

lexeme	freq
alt-right	1012150
solopreneur	282026
hyperlocal	209937
alt-left	167124
upskill	57941
poppygate	3807

maximum total usage frequency as well as six words that will serve as case studies in the following sections.

In a strict sense, usage frequency only captures how many tokens of a word were produced by all speakers who have contributed to the corpus at hand. Investigating the degree to which new words diffuse to new speakers and speaker communities on the basis of frequency counts thus depends on several inferences that are commonly accepted as sufficiently reliable.

1. Frequency counts indicate how many speakers have used the term.

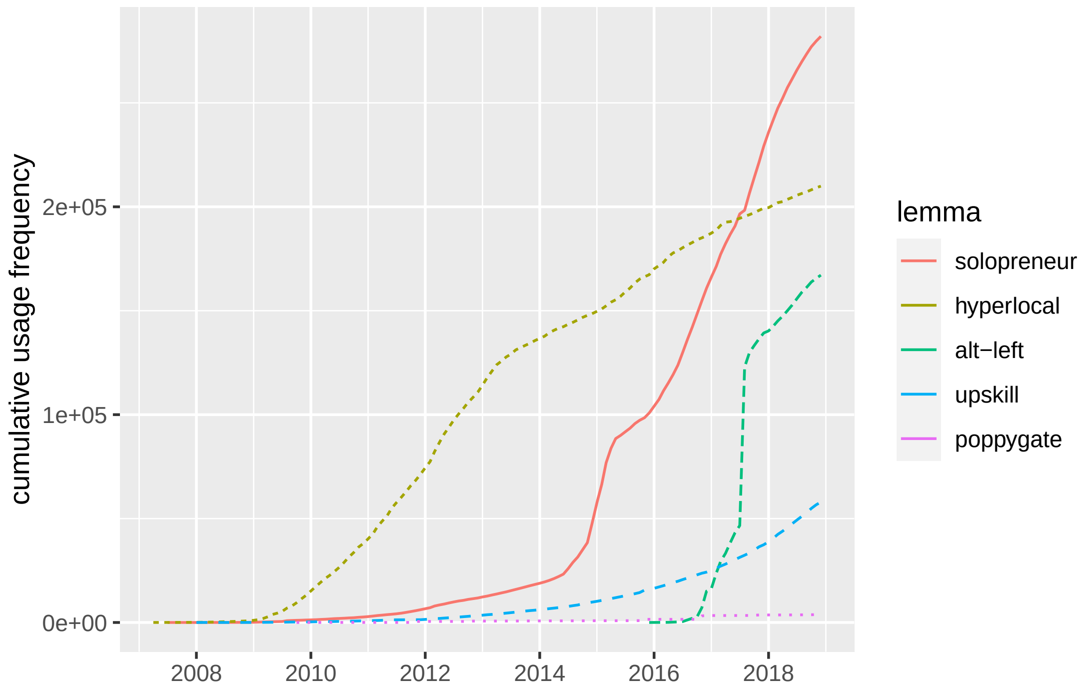
2. The number of speakers who have used the term indicates how many speakers are familiar with the term, whether they have actively used it or not.

3. The number of speakers who are familiar with the term indicate how many communities of speakers are familiar with the term.

These assumptions are to a large extent plausible and have empirically been proven to be effective for investigating both degrees of entrenchment of lexemes in individual speakers as well their conventionality in the speech community.

The frequency-based division of neologisms into groups as presented in Table 1, for example, largely seems to fit common intuition. Neologisms that show very high frequency counts such as *smartwatch* have certainly been used by (1) many speakers. It is

Figure 1: Cumulative increase in usage frequency for case studies.[3]



also unlikely that its roughly 1 million active uses stem exclusively from (3) one or few tight-knit communities of techno-enthusiasts and that the rest of the speech community (2) has never been exposed to the term.

However, even among the group of high-frequency lexemes in Table 2c, words that show similar total usage frequencies such as *twitterverse* and *blockchain*, for which semantic transcriptions even seem unnecessary, might indeed differ significantly regarding their conventionality in different parts of the speech community. For assessing and comparing the pathways of diffusion of less-established neologisms like *hyperlocal* and *solopreneur* total frequency counts alone provide a very limited picture.

5.2 Cumulative frequency

Cumulative frequency plots can supplement total frequency counts by additional information about the temporal dynamics of diffusion. Figure presents this information for all six cases.

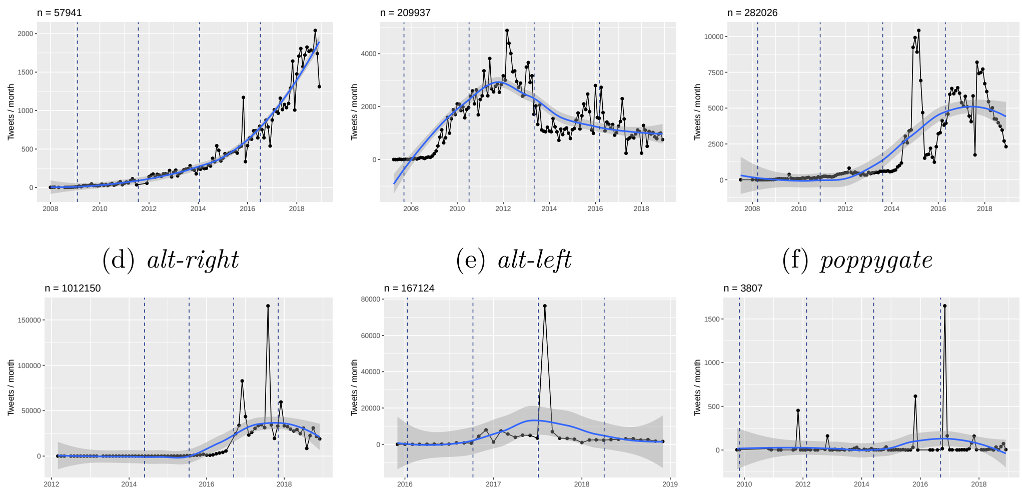
Most importantly, lifespan 1 comparison: e.g. *alt-left* vs. *alt-right*

introduce cases

Potential distortions uses != users: This can distort the picture, e.g. if some speakers have a much stronger preference to use the term than the average or the amount of

[3]For better visibility *alt-right* was omitted from this plot because of its high usage frequency.

Figure 2: Temporal dynamics in usage frequency for case studies.



(a) *upskill* (b) *hyperlocal* (c) *solopreneur*

(d) *alt-right* (e) *alt-left* (f) *poppygate*

words contributed by by each speaker is not balanced.

X is most frequent Y is oldest *poppygate*

5.3 Temporal dynamics of usage intensity

instead of cumulative counts we now look at absolute frequency counts over time (in monthly bins)

case studies

different patterns stability trend speed of diffusion

stability: shows that freq. is problematic 'dormant' spikes distort representativity of frequency for degree of conventionality underestimate: *poppygate* not forgotten in troughs overestimate: cumulating hides the fact that words like *millenium* do get lost

full sample

coefficient of variation most volatile least volatile

volatile patterns are the rule than the exception for *lexical innovation* due to nature of *lexical* innovation bound to cultural conceptual salience (variable 'semantic carrying capacity' (Nini et al., 2017)) needs to be accounted for

trend increasing: looks successful decreasing: looks unsuccessful

going beyond frequency In the following sections I will assess the value of usage frequency and compare and complement it with social network information about the diffusion of lexical innovations.

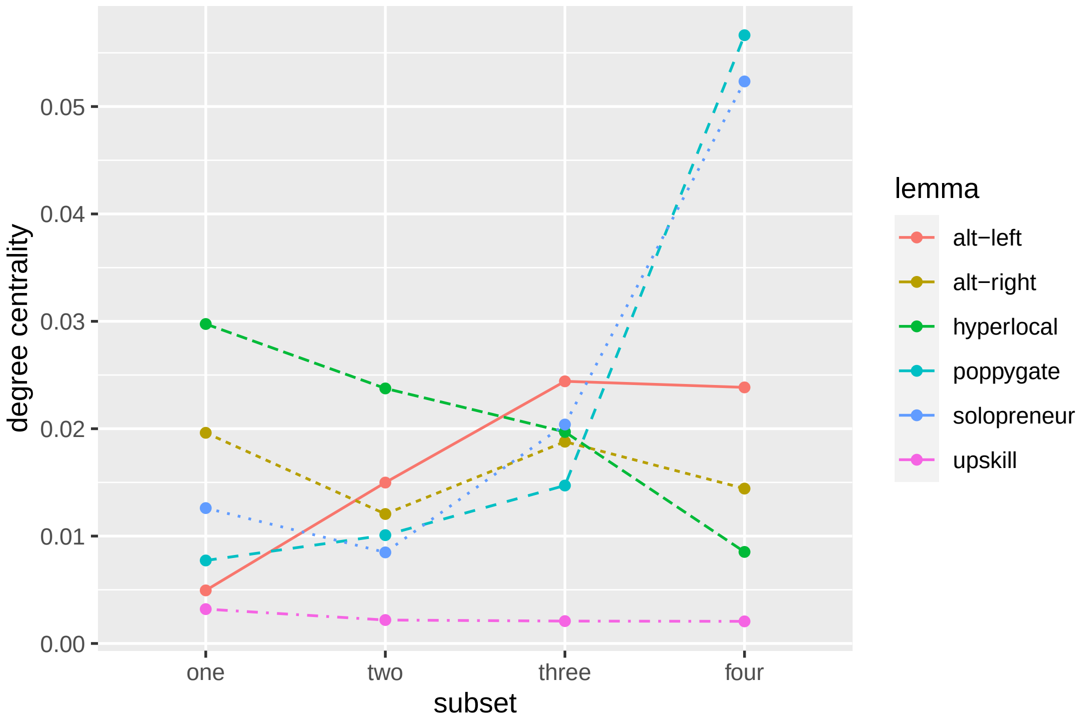
6 Social networks of diffusion

6.1 Centralization over time

going beyond frequency
def. diffusion: numbers of users communities
subsetting / time slices start of diffusion process 4 quarters
explain: degree centralization
case studies
example where freq. meets nets
example where nets add to freq.: *alt-left*

6.1.1 Overview of changes in centralization for case studies.

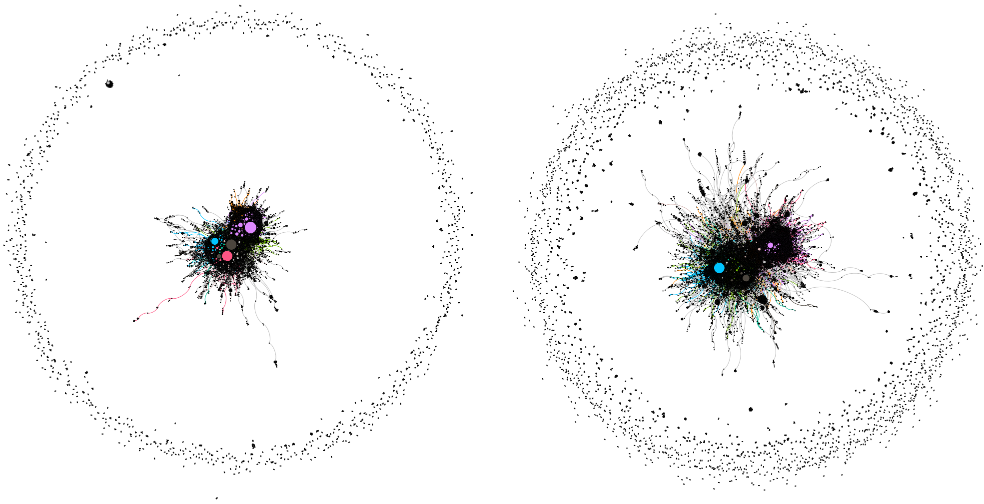
Figure 4: Degree centralization over time for case study words.



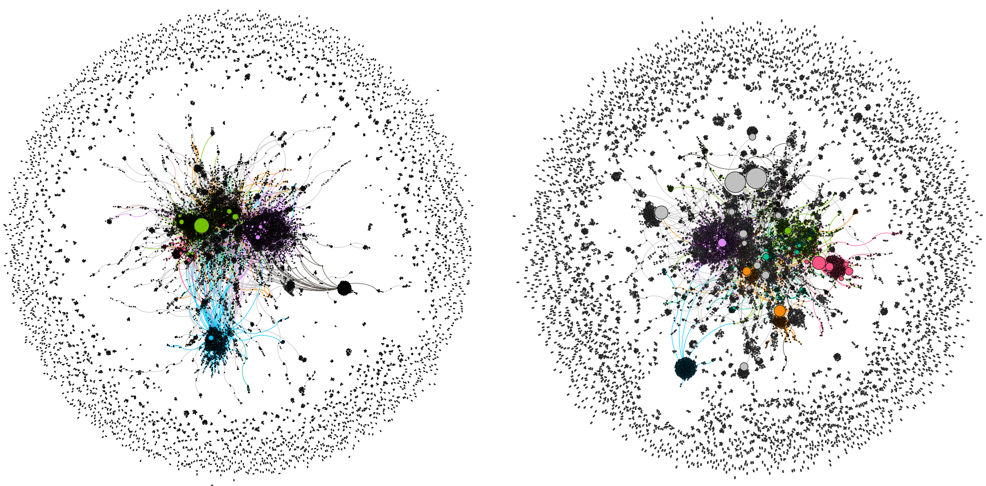
6.1.2 Advanced / increasing: *hyperlocal*

Figure 5: Social network of diffusion for *hyperlocal* over time.

(a) First stage (b) Second stage



(c) Third stage (d) Fourth stage

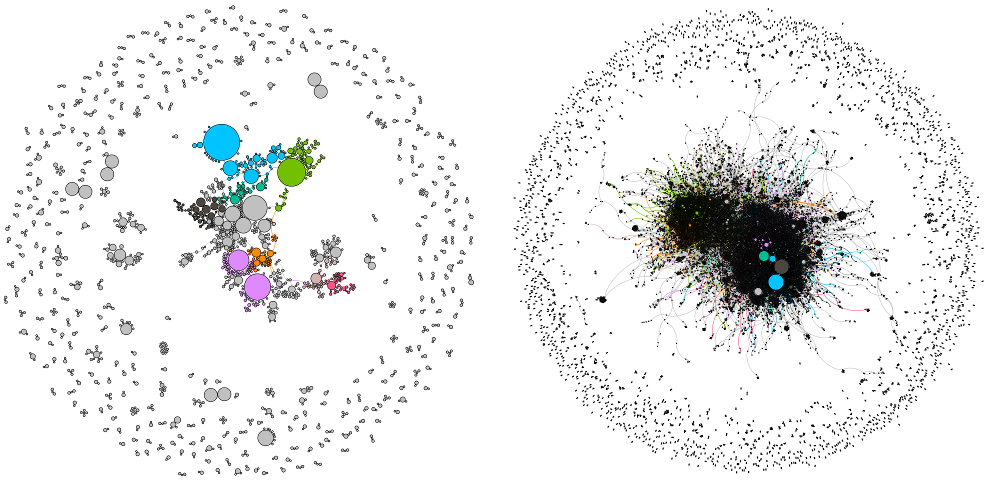


6.1.3 Limited / limited: *alt-left*

Figure 7: Social network of diffusion for *alt-left* over time.

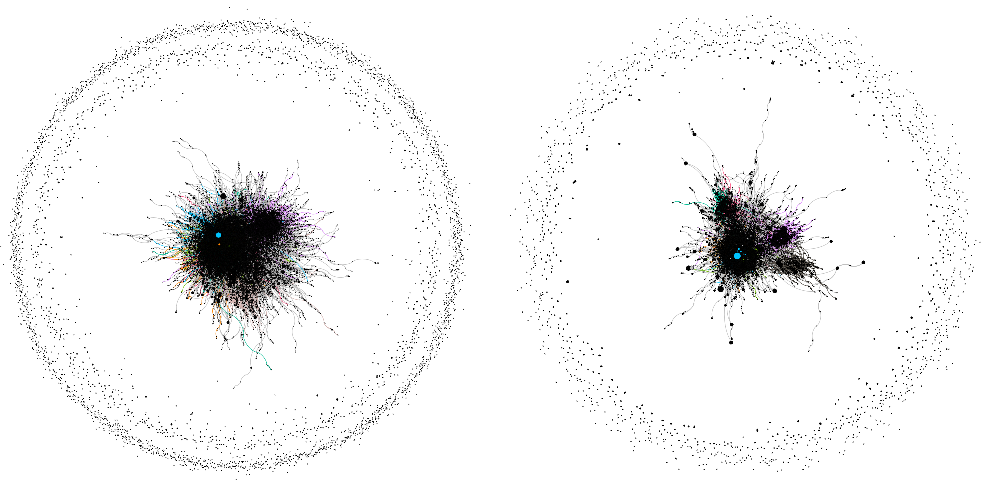
(a) First stage

(b) Second stage



(c) Third stage

(d) Fourth stage



6.1.4 Full sample

density successful unsuccessful
biggest changes

6.2 Overall centralization

most diffused least diffused

7 Networks vs. frequency

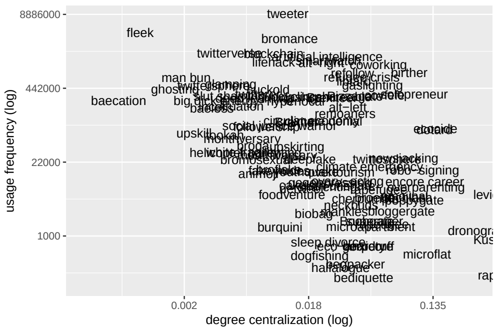
7.1 Correlation

There is a significant correlation ($p = 0.015$) between frequency and centralization

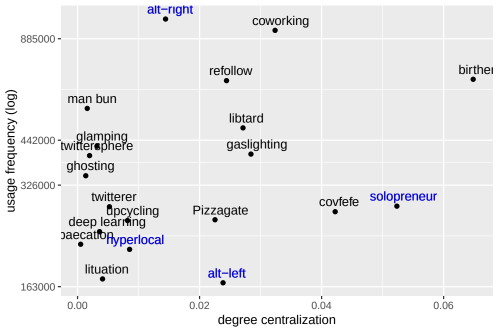
7.2 Discrepancies

However, we also see discrepancies
plots

(a) Full sample.



(b) Case studies.



cluster analysis
freq. overestimating topical propaganda: *alt-right*, *alt-left*, *covfefe*, *birther* Brexit terms: *Brexiteer*, *Brexiter*, *Brexit* nerds technical
freq. underestimating: XXX topical words
Social network metrics certainly add to freq.

8 Conclusion

freq. proves to be a pretty good indicators but temporal dynamics important social network dynamics important, esp. w.r.t. new words cross-checking other data sources (NOW corpus) shows validity social network analysis can be an important tool for sociolinguistics

References

Bauer, L. (1983). English word-formation. Cambridge: Cambridge University Press.

Cartier, E. (2017). Neoveille, a web platform for neologism tracking. In *Proceedings of the software demonstrations of the 15th conference of the European chapter of the Association for Computational Linguistics* (pp. 95–98). Valencia, Spain: Association for Computational Linguistics.

Davies, M. (2013). Corpus of news on the web (NOW): 3+ billion words from 20 countries, updated every day. Retrieved from https://www.english-corpora.org/now/

editorial staff, E. (2020). Diseases show to interact and spread like internet memes. Retrieved from https://eandt.theiet.org/content/articles/2020/02/diseases-shown-to-interact-and-spread-like-internet-memes/

Eisenstein, J., O'Connor, B., Smith, N. A. & Xing, E. P. (2014). Diffusion of lexical change in social media. *PLOS ONE*, *9*(11), 1–13.

Elsen, H. (2004). Neologismen. Tübingen: Narr.

Gérard, C. (2017). The logoscope: semi-automatic tool for detecting and documenting the contexts of french new words.

Goel, R., Soni, S., Goyal, N., Paparrizos, J., Wallach, H., Diaz, F. & Eisenstein, J. (2016). The social dynamics of language change in online networks. In E. Spiro & Y.-Y. Ahn (Eds.), *Social informatics* (pp. 41–57). Cham: Springer International Publishing.

Granovetter, M. S. (1977). The strength of weak ties. In S. Leinhardt (Ed.), *Social networks* (pp. 347–367). Academic Press.

Grieve, J. (2017). *Geographical patterns of lexical innovation.* Workshop 'Diffusion of Lexical Innovations', LMU Munich.

Grieve, J. (2018). Mapping emerging words in new york city.

Grieve, J. (forthcoming). Natural selection in the modern English lexicon. *English Language and Applied Linguistics.*

Grieve, J., Montgomery, C., Nini, A., Murakami, A. & Guo, D. (2019). Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence*, *2*, 11.

Grieve, J., Nini, A. & Guo, D. (2016). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics*, (21), 99–127.

Hohenhaus, P. (2006). Bouncebackability. a web-as-corpus-based study of a new formation, its interpretation, generalization/spread and subsequent decline. *SKASE Journal of Theoretical Linguistics*, *3*, 17–27.

Kerremans, D. (2015). A web of new words. Frankfurt a. M.: Lang.

Kerremans, D., Stegmayr, S. & Schmid, H.-J. (2012). The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. (pp. 59–96). Berlin: Mouton de Gruyter.

Labov, W. (2007). Transmission and diffusion. *Language*, *83*(2), 344–387.

Lemnitzer, L. (n.d.). Wortwarte. Retrieved January 14, 2018, from http://www.wortwarte.de/

Lipka, L. (2005). Lexicalization and institutionalization: revisited and extended. *SKASE Journal of Theoretical Linguistics*, *2*(2), 40–42.

Lu, F. S., Hou, S., Baltrusaitis, K., Shah, M., Leskovec, J., Sosic, R., ... Santillana, M. (2018). Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the boston metropolis. *JMIR Public Health and Surveillance*, *4*(1), e4.

Milroy, J. (1992). *Linguistic variation and change: On the historical sociolinguistics of english*. Oxford: Blackwell.

Milroy, J. & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, *21*(2), 339–384.

Nevalainen, T. (2015). Descriptive adequacy of the s-curve model in diachronic studies of language change. VARIENG.

Nini, A., Corradini, C., Guo, D. & Grieve, J. (2017). The application of growth curve modeling for the analysis of diachronic corpora. *Language Dynamics and Change*, *7*(1), 102–125.

Pew Research Center. (2019, October 23). National politics on twitter: small share of U.S. adults produce majority of tweets. Retrieved from https://www.people-press.org/2019/10/23/national-politics-on-twitter-small-share-of-u-s-adults-produce-majority-of-tweets/

Renouf, A., Kehoe, A. & Banerjee, J. (2006). Webcorp: an integrated system for web text search. Rodopi.

Rogers, E. M. (1962). *Diffusion of innovations*. New York: Free Press of Glencoe.

Schmid, H.-J. (2008). New words in the mind: concept-formation and entrenchment of neologisms.

Schmid, H.-J. (2016). *English morphology and word-formation: An introduction* (2nd ed.). Berlin: Erich Schmidt Verlag.

Schmid, H.-J. (2020). *The dynamics of the linguistic system. Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.

Stefanowitsch, A. & Flach, S. (2017). The corpus-based perspective on entrenchment. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: how we reorganize and adapt linguistic knowledge* (pp. 101–128). Boston, USA: American Psychology Association and de Gruyter Mouton.

West, R., Paskov, H. S., Leskovec, J. & Potts, C. (2014). Exploiting social network structure for person-to-person sentiment analysis.