The Semantic Drift of Quotations in Blogspace: A Case Study in Short-Term Cultural Evolution

Sébastien Lerique, Camille Roth

The semantic drift of quotations in blogspace: a case study in short-term cultural evolution

Sébastien Lerique
Centre d'Analyse et de Mathématique Sociales, UMR 8557
CNRS/EHESS
190 av. de France, F-75013 Paris
and
Centre Marc Bloch Berlin e.V., UMIFRE 14
CNRS/MAEDI, Friedrichstr. 191, D-10117 Berlin

Camille Roth
Sciences Po, médialab, Paris, France
and
Centre Marc Bloch Berlin e.V., UMIFRE 14
CNRS/MAEDI, Friedrichstr. 191, D-10117 Berlin

We present an empirical case study which connects psycholinguistics with the field of cultural evolution, in order to test for the existence of cultural attractors in the evolution of quotations. Such attractors have been proposed as a useful concept for understanding cultural evolution in relation with individual cognition, but their existence has been hard to test. We focus on the transformation of quotations when they are copied from blog to blog or media website: by coding words with a number of well-studied lexical features, we show that the way words are substituted in quotations is consistent (1) with the hypothesis of cultural attractors, and (2) with known effects of the word features. In particular, words known to be harder to recall in lists have a higher tendency to be substituted, and words easier to recall are produced instead. Our results support the hypothesis that cultural attractors can result from the combination of individual cognitive biases in the interpretation and reproduction of representations.

1 Introduction

The reciprocal influence between cognition and culture has a long history in both social science and psychology. While this question has been the subject of intense debate in the social sciences in the 20th century, today's discussion is mostly structured by proponents from cognitive science, who construe culture as an evolutionary process analogous and parallel to biological evolution. That analogy can be traced a long way back, with milestones such as Kroeber's works (1952), Dawkins' *Memetics* (1976), and later the development of *Dual Inheritance Theory* by Boyd and Richerson (1985) and Cavalli-Sforza and Feldman (1981) among others. More recently, Dan Sperber has drawn on this principle to explicitly connect anthropology and cognitive science through the theory of *Epidemiology of Representations* (Sperber, 1996), and the study of cultural evolution has been growing steadily since.

The collection of works by Aunger (2000) (in particular Bloch, 2000, and Kuper, 2000) has shown how memetics cannot account for the levels of transformation culture undergoes as it is transmitted. Mesoudi and Whiten (2008) discuss the uses of transmission chain experiments to test what dual inheritance theory can explain about cultural evolution. Morin (2013) and Miton, Claidière, and Mercier (2015), by carefully compiling a series of anthropological works, demonstrate how cognitive biases have influenced the evolution of cultural artifacts over several centuries. Kirby, Cornish, and Smith (n.d.; 2008) have shown how evolutionary pressures lead to the emergence of structured and expressive artificial languages in simulations and laboratory experiments. Such transmission chain experiments have also been explored in non-human primates by Claidière, Smith, Kirby, and Fagot (2014).

The theory of Epidemiology of Representations proposes a unifying framework for all these works by recasting them as questions of spread and transformation of representations: these are alternatively located in the mind ("mental representations" in Sperber's terminology), or in the outer world ("public representations") as expressions of mental representations in diverse cultural artifacts (pieces of text, utterances, pictures, building techniques, etc.). A human society is then modeled as a large dynamical system of people constantly interpreting public representations into mental representations, and producing new public representations based on what they have previously interpreted. Two key points are that (a) transmission is not reliable (representations change significantly each time they are interpreted and produced anew, as op-

Keywords: word recall; recollection bias; semantic network; cultural evolution; cultural attraction; data mining; big data; psycholinguistics.
Corresponding author: Sébastien Lerique, CAMS, EHESS, 190 av. de France, F-75013 Paris. Email: sebastien.lerique@normalesup.org.

posed to e.g. memetics), and (b) the reciprocal influences of cognition and culture can be captured by studying the evolution of public representations themselves, which is what the above-cited studies are doing.

The theory makes an additional strong hypothesis, which this paper focuses on: as transformations accumulate, some representations evolve to be very stable and spread throughout an entire society without changing any more (they are called "cultural representations", because they characterize a given culture). This process should manifest itself as attractors (called "cultural attractors") in the dynamical system that models cultural evolution, that is: there should be areas of the representation space where cognitive effects in transformations bring representations closer to a given stable asymptotic point.[1]

This hypothesis, a cornerstone of the theory because of the intelligibility it gives to cultural evolution, has been hard to test in concrete situations as quantitative data on out-of-laboratory cultural artifacts is not easy to collect. One approach, as mentioned above, has been the meta-analysis of large bodies of anthropological studies (see Miton et al., 2015, for instance). This paper exemplifies a second approach, taking advantage of the ever-increasing avalanche of available digital footprints since the 2000's. Indeed, tools and computing power to analyze such data are now widespread, and the body of research aimed at describing online communities and content is growing accordingly. For instance, the propagation of cultural artifacts across social networks has been studied in blogspace (Gruhl, Guha, Liben-Nowell, & Tomkins, 2004) and in emails (Liben-Nowell & Kleinberg, 2008); Cointet and Roth (2009) described the reciprocal influence between the social network topology and the distribution of issues; Leskovec, Backstrom, and Kleinberg (2009) detailed the characteristic times and diffusion cycles both within these social networks and with respect to the topical dynamics of news media, and Danescu-Niculescu-Mizil, Cheng, Kleinberg, and Lee (2012) studied the characteristics of particularly memorable quotes that circulate in those networks. We believe these works can connect the field of cultural evolution with psycholinguistics to advance the testing of cultural attractors.

To show this we analyze how quotes in blogs and media outlets are modified when they are copied from website to website. These public representations should normally not change as they spread on the Web (as opposed to more elaborate expressions or opinions, not identified as quoted utterances), but empirical observation shows that they are in fact occasionally transformed (Simmons, Adamic, & Adar, 2011): authors spontaneously transform quotes, not only cropping them but also replacing words. For instance the quote "we will not be scared of these cowards" (a substring of a quote from former Pakistani President Asif Ali Zardari) is also found as "we will not be **afraid** of these cowards". More meaningful changes often happen too, such as the transformation of McCain's "I admire Senator Obama and his accomplishments" during the 2008 US presidential campaign, into "I **respect** Senator Obama and his accomplishments". Since authors are implicitly required to copy quotes exactly, we can assume that most transformations, especially simple ones such as those shown above, are the result of automatic (i.e. hard to control) low-level cognitive biases of the authors.

We thus ask the following question: given such representations that seem to evolve precisely because of the kind of automatic cognitive biases evoked in the theory of epidemiology of representations, do cultural attractors appear, and if so how do cognitive biases participate in them? We chose to restrict our analysis to substitutions (i.e., one word being replaced by another), both to keep the analysis tractable and because of missing information in our data set.[2] While this limits the scope of our results to the particular data set we use, the methodological point we also make is left intact. By characterizing words using 6 well-studied features, we identify what makes a substitution more likely, and how a word changes when it is substituted. This exploratory approach uncovers a number of transmission biases consistent with known effects in linguistics. While the transformations we describe are not the only ones at work in this data set, our analysis also indicates that feature-specific attractors could exist because of the substitution process. This study can be viewed as analyzing part of the transmission step operating in transmission chains of artificial languages like those studied by Kirby et al. (2008), yet with natural language out of the laboratory.

The next section describes our hypotheses along with a review of the psycholinguistics literature. Then, we describe the data set and detail the various assumptions that were made in order to analyze it. Next, we introduce the measures we built to observe cognitive biases operating in quote transmission. Finally, we discuss the relevance of these results for the study of cultural evolution, followed with general guidelines for further work.

2 Related work

The study of cultural evolution on the part of cognitive science emerged only recently. While formal models of cultural transmission appeared with the development of dual inheritance theory (Boyd & Richerson, 1985; Cavalli-Sforza &

[1] Attractors need not be points in fact, they can also be sub-areas; in that case any transformation brings representations in the area closer to (or confined to) the target sub-area.

[2] As explained further down, source-destination links between quotes must be inferred from the data set, an operation which is much more reliable if we restrict our analysis to substitutions. This also impedes us from observing the effect of accumulated transformations in the long term, limiting our results to a view of the individual evolutionary step.

Feldman, 1981) and have included the notion of cultural attractor since then (Claidière, Scott-Phillips, & Sperber, 2014; Claidière & Sperber, 2007), collecting data to test and iterate over such models has been more challenging. The first above-mentioned method consists in rebuilding the history of a given type of representation by compiling anthropological or historical works on the subject (as for instance Morin, 2013, and Miton et al., 2015, have done). A second approach uses cultural evolution experiments in the laboratory, with an array of methods reviewed by Mesoudi and Whiten (2008). Transmission chains, in particular, have been used extensively to study the evolution of human language (see Tamariz & Kirby, 2016, for a review). Other recent examples include studies of the evolution of simple audio loops through consumer preference (MacCallum, Mauch, Burt, & Leroi, 2012), the emergence of structure in visual patterns transmitted by baboons (Claidière, Smith, et al., 2014), and the amplification of risk perception through chains of casual conversation (Moussaïd, Brighton, & Gaissmaier, 2015).

Research on online content points to a third approach to this question. By investigating the transformations of quotations in a large corpus of US blog posts and online news stories initially collected and studied by Leskovec et al. (2009), Simmons et al. (2011) and later Omodei, Poibeau, and Cointet (2012) show that even for quotations, a type of public representation that should change the least when transmitted on the Web, it is still possible to witness significant transformations. These studies focus on the influence of the quotation source (e.g. news outlet vs. blog) or of the surrounding public space (e.g. quotation frequency in the corpus), and suggest diffusion-transformation models to capture the dynamics of the population of quotations. But the cognitive features which may determine or, at least, influence these transformations, are overlooked. On the other hand cognitive and linguistic features have been used in diffusion studies not involving transformation: Danescu-Niculescu-Mizil et al. (2012), for instance, show that particularly memorable quotations (taken from movie scripts in this case) use more distinctive words and have more common syntax than less memorable quotations; they are also the quotes that adapt best to new contexts of use. One source of ideas to study the transformations of such quotes, then, might be the psycholinguistic literature studying word and sentence recall.

Potter and Lombardi (1990) suggest that immediate recall of sentences is based on the retention of an unordered list of words which is then regenerated as a sentence at the moment of production. Priming recall with other words can lead to replacement in the recalled sentence if the primed words are consistent with the overall meaning of the sentence. Regenerated syntax can also be influenced by priming recall with another syntactic structure (Potter & Lombardi, 1998), or with verbs whose category constraints call for a different structure (Lombardi & Potter, 1992).

Compared to full sentences, recall of word lists provides a situation that is easier to fully explore and has been extensively studied. In particular, the Deese, Roediger, and McDermott paradigm (introduced by Deese, 1959, and later popularized by Roediger & McDermott, 1995) has shown that it is possible to construct lists of words which reliably create the false memory of an external word related to those in the list. This is done by using lists of words produced by free association from the target intrusion word; the intruding recall then happens with probability nearly proportional to the average semantic association strength between the intruding word and the words in the list. A sizable literature studies this type of task with varying complexities in the design of the lists, a good review of which is given by Zaromb et al. (2006). One notable effect is that the semantic relations between words greatly influence, and correlate to, the order in which words are recalled (Howard & Kahana, 2002; Tulving, 1962), and that this reordering of items improves subjects' repeated recalls (Tulving, 1966). The frequency and type of intrusions in lists of random words are also influenced by associations created by the presentation of previous lists (Zaromb et al., 2006). Indeed, the question of how such temporal associations (contributing to contextual information retrieval in recall) interact with the prior semantic associations of subjects (contributing to associative information retrieval) is at the core of many of these studies.

These effects do not transpose simply to sentence recall however, as not only syntax but also effects of attention come into play for both retrieval and encoding. Jefferies, Lambon Ralph, and Baddeley (2004), for instance, show that attention is central to the encoding and retention of unrelated propositions, on top of more automatic syntactic and semantic processes. This involvement of executive resources also seems to contribute to the much greater memory span subjects exhibit for sentences compared to word lists (see Jefferies et al., 2004, again, for more details).

Given this complexity we decided to focus on more aggregate measures, where variations of the conditions in which sentences are read and produced have a chance of being statistically smoothed out.[3] If a cognitive bias in the substitution of words manifests itself with simple measures, then it will be worth applying predictive models of the substitution process in further research.

Lexical features, then, are obvious well-studied word measures that can be analyzed in aggregate. Indeed

[3] Aside from our lack of control on the precise conditions of encoding and recall in our data set, the analysis techniques mentioned above are better suited to data consisting of a high number of measures over a smaller number of lists (in which case it makes sense to ask e.g. what proportion of intrusions come from prior lists). As is explained further down however, our data set is shaped the opposite way: a great number of sentences, with only very few to no measures at all on each sentence.

word frequency (see Yonelinas, 2002, for a review), age-of-acquisition (Zevin & Seidenberg, 2002), number of phonemes (see for instance Nickels & Howard, 2004; Rey, Jacobs, Schmidt-Weigand, & Ziegler, 1998), and phonological neighborhood density (Garlock, Walley, & Metsala, 2001) to name a few, all have known effects on word recognition or production. More complex features based on word networks built from free association or phonological data have also been analyzed: Nelson, Kitto, Galea, McEvoy, and Bruza (2013) for instance, show the importance of clustering coefficient in such a semantic network by studying the role it plays in a variety of recall and recognition tasks (extralist and intralist cuing, single item recognition, and primed free association). Chan and Vitevitch (2010) show that pictures are named faster and with fewer mistakes when they have a lower clustering coefficient in an underlying phonological network. Griffiths, Steyvers, and Firl (2007) analyze a task where subjects are asked to name the first word which comes to their mind when they are presented with a random letter from the alphabet. The authors show that there is a link between the ease of recall of words and their authority position (pagerank) in a language-wide semantic network built from external word association data (Austerweil, Abbott, & Griffiths, 2012, further develop this tool to give a parsimonious account of the fact that related words are often retrieved together from memory).

On the whole, research on lexical features hints towards two antagonistic types of effects (also known as the "word-frequency paradox", Mandler, Goodman, & Wilkes-Gibbs, 1982). On one hand, part of the literature shows that recall is easier for the least "awkward" words; those whose age of acquisition is earlier, length is smaller, semantic network position is more central — this is particularly true in retrieval, that is in tasks where participants are asked to form spontaneous associations or utter a word in response to a given signal. On the other hand, when the task consists in recognizing a specific item in a list, "awkward" words are actually more easily remembered, possibly as they are more informative and plausibly more discernible (see again Yonelinas, 2002, for a review). The jury is still out as to whether reformulation alteration, that is spontaneous replacement of words when asked to rewrite a given utterance, is rather of the former or latter sort. We also aim to shed some light on this debate, considering oddness as a dimension of the purported fitness of utterances.

3 Methods

We rely on a text corpus made of quotations extracted from online blog posts, and focus on their evolution. Quotations appeared to be a perfect candidate to propose a first measure of automatic cognitive bias in cultural transmission. First, they are usually cleanly delimited by quotation marks, which greatly facilitates their detection in text corpora. Second, they stem from a unique original version, and are ideally traceable back to that version. Third, and most importantly, their duplication should *a priori* be highly faithful, apart from cases of cropping: not only should transformations be of moderate magnitude, but when specific words are not perfectly duplicated, it is safe to assume that the variation is due to involuntary cognitive bias — as writers may expect any casual reader to easily verify, and thus criticize, the fidelity to the original quotation.

We could therefore study the individual transformation process at work when authors alter quotations, by examining the modified words in each transformation. Since our approach is exploratory however, we do not know at the outset which precise effect of cognitive bias we are looking for. Indeed, the data we use does not come from a controlled experiment in the laboratory, designed to elicit a particular effect: they are recordings of real life interactions, with all the complexity and uncertainty of conditions this entails. Our goal, therefore, is to show that cognitive biases have measurable effects in this setting even if they are part of a larger complexity (the detailed prediction and deconstruction of the cognitive processes responsible for them being left to further research). If this is confirmed, we will have successfully tested fundamental cognitive biases with out-of-laboratory data, opening a path to explanations of actual (vs. simulated) cultural evolution with tools from cognitive science. Aiming to exhibit such subtle biases in complex data is the main reason we chose to use aggregate measures that have a chance of smoothing out the possible variations of experimental conditions in the data set.

To keep the analysis tractable, we focused on quotation transformations consisting of the *substitution* of a word by another word (and only those cases) in order to unambiguously discuss single word replacements. This restriction also allows us to more reliably infer the information that is missing in our data set, as explained in the "Substitution model" section. To quantify these substitutions we decided to associate a number of features to each word, the variation of which we can statistically study.

The next subsections describe the data set and the measures we used to assess this cognitive bias.

3.1 Corpus-based utterances

We used a quotation data set collected by Leskovec et al. (2009), large enough to lend itself to statistical analysis. This data set consists of the daily crawling of news stories and blog posts from around a million online sources, with an approximate publication rate of 900k texts per day, over a nine-month period of time from August 2008 to April 2009 (Leskovec et al., 2009).[4] The authors automatically extracted

[4]The original article (Leskovec et al., 2009) does not provide further details on the source selection methodology.

quotations from this corpus. Each quotation is a more or less faithful excerpt of an utterance (oral or written) by the quoted person; for instance:

> The Bank of England said, "these operations are designed to address funding pressures over quarter-end."

Then, the authors gathered quotations in a graph and connected each pair that differed by no more than one word or that shared at least ten consecutive words (they tested this procedure with a number of different parameters, see Leskovec et al., 2009, for more details). We find for example the following variation of the above quote:

> "these operations are **intended** to address funding pressures over quarter-end."

Next, they applied a community detection algorithm to that quotation graph to detect aggregates of tightly connected, that is sufficiently similar, groups of quotations (see again Leskovec et al., 2009, for more details). This analysis yielded the final data we had access to, with a total of about 70,000 sets of quotations; each of these sets ideally contains all variations of a same parent utterance, along with their respective publication URLs and timestamps (since the procedure cannot be perfect, sets of quotations contain occasional rogue unrelated variations that should have been discarded or assigned to another set).

Manual inspection of this data set revealed that it contains a significant number of everyday language quotations (such as "it was much better than I expected", "did that just happen", as well as many simple expletive-based sentences). Their presence is largely due to random variations around casual expressions, while we are interested in transformations of news-related quotes causally linked to an original, identifiable utterance. To filter them out, we exclude quotes with less than 5 words or whose occurrences span more than 80 days (indicating causally unrelated occurrences), as well as quotes not written in English. Clusters that are emptied by this procedure are therefore excluded. If, after this screening, a cluster's occurrences still span more than 80 days (because of short-lived but unrelated quotes far apart in time), we also exclude it. We eventually keep 50,427 clusters (out of 71,568; i.e. 70.5%), containing a total of 141,324 unique quotes (out of 310,457; i.e. 45.5%) making up about 2.60m occurrences (out of 7.67m; i.e. 33.9%).[5] Even if we lose some real event-related utterances which are present in clusters lasting more than 80 days (one such lost quote, for instance, is "the city is tired of me and the organization and I have run our course together"), we check that our filtering approach fulfills its goals by coding a random sub-sample of 100 clusters: 35 of them are rejected by the filter, with 15 false negatives (rejected clusters that should have been kept) and 9 false positives (clusters kept when they should have been rejected), giving a precision score of 0.862 and a recall score of 0.789. Furthermore, all but one of the 9 false positives are left with a single non-rejected quote, meaning those clusters are ignored by our substitution analysis; this brings the effective precision of our filter to 0.982.[6]

3.2 Word-level measures

3.2.1 Lexical features

We first introduce some lexical measures on words.

- **Word frequency**: the frequency at which words appear in our data set, known to be relevant for both recognition and recall (Gregg, 1976),
- **Age of Acquisition**: the average age at which words are learned (obtained from Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), known to have different effects than word frequency (Dewhurst, Hitch, & Barry, 1998; Morrison & Ellis, 1995),
- **Phonological** and **Orthographic Neighborhood Density** (obtained from Marian, Bartolotti, Chabal, & Shook, 2012), also known to be relevant for word production (Garlock et al., 2001),
- The average **Number of Phonemes** and **Number of Syllables** for all pronunciations of a word (obtained from the Carnegie Mellon University Pronouncing Dictionary, Weide, 1998)[7], as well as **Number of Letters**, as a proxy to word production cost,
- The average **Number of Synonyms** for all meanings of a word (obtained from WordNet, 2010) as an *a priori* indicator of how easy it would be to replace a word.

We also consider grammatical types within quotations by detecting *Part-of-Speech* (POS) categories with TreeTagger (Schmid, 1994); we distinguish between verbs, nouns, adjectives, adverbs, and closed class-like words.

Aside from these raw features, the systemic dimension of vocabulary (Cornish et al., n.d.) has led authors to develop measures based on the full topology of networks built from free association data or phonological similarity. Several such measures have been shown to be involved in recall, recognition, and naming tasks (Chan & Vitevitch, 2010; Griffiths et al., 2007; Nelson et al., 2013).

[5] The significantly larger loss in occurrences indicates that, on average, the clusters we lose contain more occurrences than those we keep, which is to be expected for everyday language utterances.

[6] A similar analysis was made for language detection, which is part of the cluster filtering: out of 100 randomly sampled quotes, 17 are rejected because their detected language is not English, with no false positives and 6 false negatives, giving a precision score of 1 and a recall score of 0.933. Of the 6 false negatives, 4 had less than 5 tokens and would have been excluded by the cluster filter anyway.

[7] The CMU Pronouncing Dictionary is included in the NTLK package (Bird, Klein, & Loper, 2009), the natural language processing toolkit we used for the analysis.

To compute these features we relied on the free association (FA) norms collected by Nelson, McEvoy, and Schreiber (2004), which record the words that come to mind when someone is presented with a given cue. As Nelson et al. explain, "free association response probabilities index the likelihood that one word can cue another word to come to mind with minimal contextual constraints in effect." Similar to what Griffiths et al. (2007) did, we first considered the directed weighted network formed by association norms, where nodes are words and edges are directed from cue to target word, with a weight equal to the association strength (that is the probability of that target word being produced when this particular cue is presented). This network is of particular interest since it lets us define features that reflect the associations driving false memories in word lists (Deese, 1959), a phenomenon which may be involved in the transformation of quotations.

We used three standard measures on the FA network:

- **Incoming degree centrality**, measured by the number of cues for which a given word is triggered as a target, and a corresponding generalized measure, node **Pagerank** (Page, Brin, Motwani, & Winograd, 1999), which has already been used on the FA network by Griffiths et al. (2007). In the present case these two polysemy-related measures are quasi-perfectly correlated.[8]
- **Betweenness centrality**, another measure of node centrality describing the extent to which a node connects otherwise remote areas of the network (Freeman, 1977). This quantity tells us if some words behave as unavoidable waypoints on association chains connecting one word to another.[9]
- **Clustering coefficient**, which measures the extent to which a node belongs to a local aggregate of tightly connected nodes (Watts & Strogatz, 1998), computed on the undirected weighted version of the FA network.[10] This tells us if a word belongs more or less to a group of equivalent words (from a free association point of view).

3.2.2 Variable correlations

Several of these features are strongly related and can be grouped together. To make correlation values as well as future comparisons more reliable, we log-transformed features that have marked exponential distributions (i.e. a few words valued orders of magnitude higher than the vast majority of other words).

The pairwise correlations in the initial set of features appears in Fig. 1. By looking at absolute values, three subsets of highly correlated features can be easily identified: (a) number of letters, phonemes, and syllables with pairwise correlations greater than .75; (b) orthographic and phonological neighborhood densities, with a correlation of .8; (c) age of acquisition, betweenness, degree, and pagerank centralities, with absolute pairwise correlations at .41, .59, .6, .61, .63 and .85. Applying a feature agglomeration algorithm targeted at 6 groups refined this observation by producing identical (a) and (b) groups, a (c) group without betweenness centrality which was instead assigned to a group (d) with clustering coefficient, and the remaining features (frequency and number of synonyms) as singletons.[11]

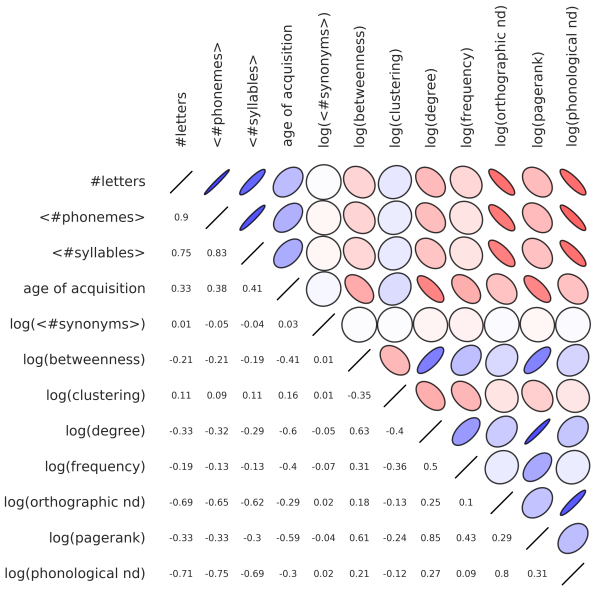


Figure 1. Spearman correlations in the initial set of features

Since our data is about written transformations, number of letters and orthographic neighborhood density are the natural representatives of groups (a) and (b) respectively. Given the importance of age of acquisition in the lexical feature literature, we chose it to represent group (c). Finally we used

[8] Note that in-degree does not take the weights of links into account, as it counts 1 for each incoming link. Pagerank on the other hand, does take the weights into account.

[9] For this measure, weights are interpreted as inverse cost: the stronger a link, the easier it is to travel across it. A stronger link will be favored over weaker links in the computation of the shortest path between two words.

[10] The Clustering coefficient is formally defined as the ratio between the number of actual versus possible edges between a node's neighbors; this is poorly defined in the case of directed networks, which led us to ignore the direction of links in the network for this measure (if two words are connected in both directions, the weights of both links are added to make the final undirected link's weight).

[11] Agglomerating into less than 6 groups merged groups (a) and (b), which we excluded to keep neighborhood densities in their own group; agglomerating into more than 6 groups separated age of acquisition from group (c), which we excluded given its high correlation values to the rest of group (c). We used scikit-learn's FeatureAgglomeration class for this procedure (Pedregosa et al., 2011).

clustering coefficient to represent group (d) since it has already been used in previous studies. The final set of features we will discuss in the rest of the paper, as well as their cross-correlations, can be seen in Fig. 2. [12]

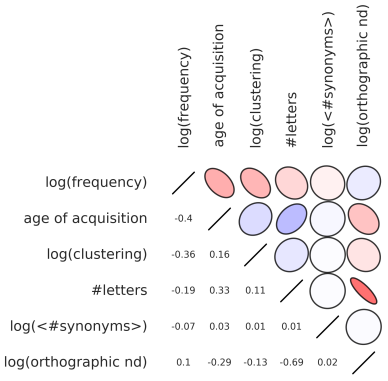


Figure 2. Spearman correlations in the filtered set of features

3.3 Substitution model

We finally need a substitution detection model, for the quotation data we use presents a challenge: quote-to-quote transformations and substitutions are not explicitly encoded in the data set. More precisely, each set of quotations bears no explicit information about either the authoritative original quotation, or the source quotation(s) each author relied on when creating a new post and reproducing (and possibly altering) that source. In other words we face an inference problem where, given all quotations and their occurrence timestamps, we must estimate which was the originating quotation for each instance of each quotation.

We therefore model the underlying quotation selection process by making a few additional assumptions. Given a particular occurrence of a quotation, the first issue is deciding whether that occurrence is a strict copy of an earlier occurrence, or a substitution from another quotation, or maybe a substitution or copy from quotes appearing outside the data set, that is from a source external to the data collection perimeter. The second issue is deciding which source originated such a substitution when several candidate sources are available.

Let us give an example: say the quotation "These accusations are false and **absurd**" (q) appears in two different blogs on January 19, and the slightly different quotation "These accusations are false and **incoherent**" (q') appears in another blog on the 20th of January. The second occurrence of q can safely be assumed to be a faithful copy of the first one the same day. And since q is fairly prominent when q' first appears, we could assume that the author of q' on the 20th based herself on q, as is shown with a dashed line in Fig. 3. Now say a third version, "These **allegations** are false and **incoherent**" (q'') also appears once on January 19 and once on January 20 after q'. Here, q and q'' differ by two substitutions, so we discard the possibility that one was written based on the other (see below for further details). q'' is only one substitution away from q' however, so we could also consider the first occurrence of q'' as a potential source for q' on the 20th. Conversely, the occurrence of q'' on the 20th could be considered as a substitution from q', or as a faithful copy from its initial occurrence on January 19. (Options shown in Fig. 3.)

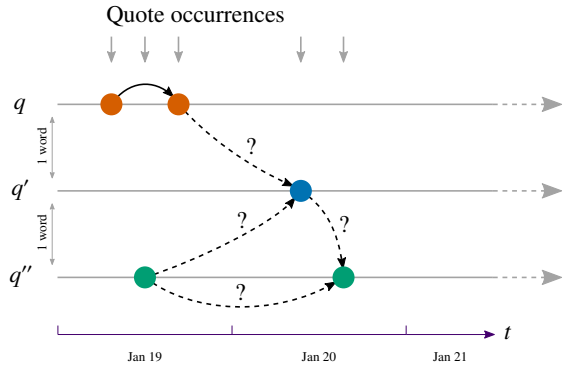
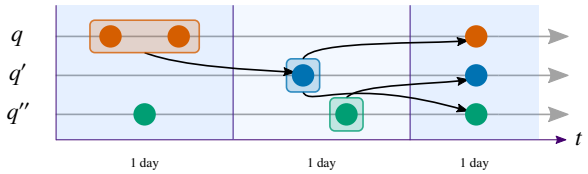


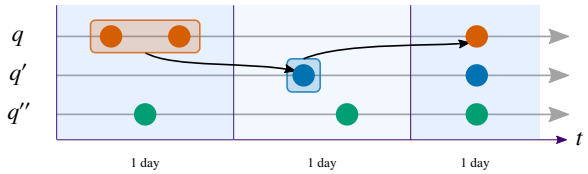
Figure 3. **Possible paths from occurrence to occurrence.** q, q' and q'' are three quotation variants belonging to the same cluster. q and q'' differ by two words, but q' differs from both q and q'' by one word. The second occurrence of q can safely be considered a faithful copy of the first, but the occurrences of q' and q'' are uncertain: while the first occurrence of q' is most likely a substitution from q, it could also stem from q''; conversely, the second occurrence of q'' could also be a substitution from q' instead of being a faithful copy of its first occurrence.

One way to settle these questions is the following: group quote occurrences into fixed bins spanning Δt days (1 day in the implementation), each one representing a unit of time evolution; when a quotation q' appears in bin $t+1$, it is counted as a substitution if it differs from the most frequent quote of the preceding bin t (or a substring thereof) by only one word; if not, q' is not considered to be an instance of substitution. Fig. 4a shows the inferences made by such a model. The assumptions it embeds, however, are a subset of a much wider set of possibilities, each leading to alternative inferences.

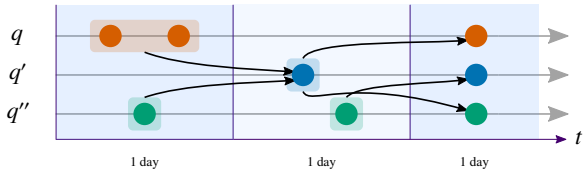
[12] Note that feature values stem from different data sets which do not always encode the same words. Indeed, we have data on frequency for about 33.5k words, on age of acquisition for 30.1k words, on clustering coefficient for 5.7k words, number of synonyms 111.2k, and orthographic density 17.8k words. Quite often then, not all features are available for a given word in our data set; however this is not problematic since the analysis is done on a per-feature basis, and not all words need be encoded in all features.



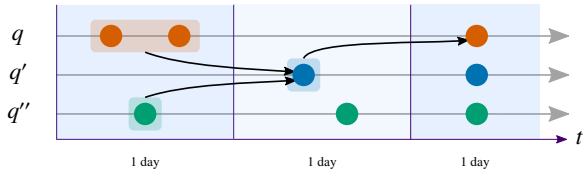
(a) Source must be majority in preceding bin, destination can be anything



(b) Source must be majority in preceding bin, destination must not appear in preceding bin



(c) Source can be anything, destination can be anything



(d) Source can be anything, destination must not appear in preceding bin

Figure 4. **Substitution models.** Substitutions inferred by four models in the situation introduced by Fig. 3. Each of these models uses bins spanning 1 day aligned to midnight (see the main text for a complete description of parameters). In the top left panel (a), q holds the majority in the first bin and is considered the unique basis for q' in bin 2. q' and q'' have equal maximum frequency in bin 2 however, so both are sources of substitutions towards bin 3. In the top right panel (b), quotes that appear in the preceding bin cannot be the target of a substitution; this removes two substitutions compared to panel (a). In the bottom left panel (c), the majority constraint is lifted compared to panel (a), making q'' in bin 1 a candidate source for q' in bin 2. In the bottom right panel (d), the majority constraint is also lifted compared to panel (a) (adding the same $q'' \to q'$ substitution as in panel (c)), and the excluded-past constraint is added as in panel (b) (removing the two same substitutions from bin 2 to bin 3 as in panel (b)). If the bins were extended to the beginning of the quotation family, the excluded-past constraint would also remove the $q' \to q$ substitution from bin 2 to bin 3. In all four panels, a background rectangle or square indicates the quotation is the source of a substitution. A thick border on that rectangle or square indicates the quotation was selected because it has maximum frequency.

We identified four binary parameters that differentiate potential models, such that the resulting 16 combinations cover most of the reasonable answers to inference uncertainties. The first two parameters define the preceding time bin from which authors could have drawn a source when producing a new occurrence: (1) **bin positions**, which can be aligned to midnight (as in the model presented above) or kept sliding (for each occurrence, use a bin that ends precisely at that occurrence); (2) **bin span**, which can be $\Delta t = 1$ day (as in the model above) or can be extended up to the very first occurrence in the quotation family. The other two parameters configure rules on the selection of source and destination quotes of a substitution: (3) **candidate sources** can be restricted to the most frequent quotations in the preceding time bin (as in the model above), or not (in which case all quotations in the preceding bin are candidate sources); (4) **candidate destinations** can be restricted to quotations that do not appear in the preceding bin, or without restriction (as in the model above). A substitution model, then, is the given of a value for each of those parameters; it considers valid all the substitutions (and only those) where the source and destination follow the rules set out by the parameters. If a destination has substitutions from multiple sources we count a single effective substitution where, for each feature, the value for the effective source word is the average of the values of the candidate source words.

Put shortly a model defines how many times, and under what source and destination conditions, quote occurrences can be counted as substitutions. Fig. 4 shows the inferences made by the four models that use bins spanning 1 day aligned to midnight: later occurrences of q' and q'' are counted as substitutions in Fig. 4a and Fig. 4c, whereas in Fig. 4b and Fig. 4d they are not.

The results reported and discussed in the following sections are valid for all 16 models, and the graphics we present were produced by the model first introduced above. Finally, note that this inference procedure is one of the reasons we restricted our analysis to single-substitutions: looking for more complex transformations would (a) exponentially increase the number of candidate sources for a destination occurrence, which correspondingly reduces the confidence in inferences made, and (b) greatly increase the complexity of the transformation models used to make these inferences.[13]

[13] We checked that this restriction does not bias the results discussed below by extending our protocol to two-substitution trans-

In practice for the model first introduced above, from the 2.60m initial occurrences spread over 50,427 quotation families, with significant redundancy (many quotes are indeed simple duplicates), we mine 40,868 substitutions. From these substitutions we remove those featuring stopwords, minor spelling changes (e.g. center/centre), abbreviations (e.g. November/Nov or Senator/Sen), spelled out numbers, words unknown to WordNet, and deletions in substrings (which can appear as substitutions of non-deleted words); this eventually yields 6,177 valid substitutions.[14]

4 Results

Substitutions usually replace a word with another semantically related word: manual observation of a random subset of 100 substitutions shows that, compared to the word it replaces, the new word often achieves a similar meaning in the context of its sentence while still slightly changing the implications or the attitude expressed by the author.[15] The following examples illustrate this phenomenon:

- "This is {socialism → welfare} for the rich",
- [The] "perverse logic of {clashes → confrontation} and violence",
- "This {crisis → problem} did not develop overnight and it will not be solved overnight".

Our question concerns the low-level properties of these substitutions: we ask (a) which words are targets of the substitutions and (b) what change these words are subjected to. To this end, we build the following two observables for each word feature. First, we measure which word features are more or less substituted compared to how often they would be if the process were random, in order to capture the susceptibility for words to be the target of a substitution in a quote. Second, we measure the change in word feature upon substitution, looking at the variation of a given feature between start and arrival words. Since sentence context is also central to this process, we extend these two observables by applying them to feature values relative to the distribution of feature values in the sentence in which a word appears.

Note that since we only consider substitutions and not faithful copies, we measure the features of an alteration *knowing that there has been an alteration*, that is we do not take invariant quotations into account. Indeed, in the former case we know there has been a human reformulation, whereas in the latter case we cannot know whether there has been perfect human reformulation or simply digital copy-pasting of a source ("Ctrl-C/Ctrl-V"). Moreover, perfect human reformulation possibly involves different practices than those involved in alteration — for instance drafting before publishing, double-checking sources, proof-reading — and may not be representative of the cognitive processes at work during alteration. The two situations are different enough to be studied separately, and we focus here on the latter.

4.1 Susceptibility

We say that a word is *substitutable* if it appears in a quote which undergoes a substitution, whether the substitution operates on that word or on another one. For a given group of words g, say all nouns, or all words in a particular range of values for a feature (e.g. words 2 to 4 letters long), susceptibility is computed as the ratio of s_g, the number of times words of that group are substituted, to s_g^0, the number of times words of that group would be substituted if substitutions fell randomly on substitutable words.[16] That is:

$$\sigma_g = \frac{s_g}{s_g^0}$$

In other words, susceptibility measures how much more or less a group of words g actually gets substituted compared to picking targets at random in quotes undergoing substitutions. By applying this measure to Part-of-Speech (POS) categories and feature bins (e.g. for a feature ϕ and a bin $[a;b]$, $g = \{w|\phi(w) \in [a;b]\}$), susceptibility measures the bias in the selection of start words involved in substitutions, i.e. it measures the preferential selection of some word properties for substitution.

formations; the results were unchanged.

[14] Manually coding a random subset of 100 substitutions to evaluate this last filter showed that 84 were true negatives, 5 were false positives, and 11 true positives, giving a recall score of .688. Precision was evaluated over a random subset of 100 *kept* substitutions, showing a score of .87. Finally, note that excluding minor spelling changes does not bias our use of orthographic neighborhood density as a feature: out of the first 100 substitutions coded for recall, those with Levenshtein distance equal to 1 (which is what orthographic neighborhood density codes, Marian et al., 2012) were all typos or UK/US spelling changes, neither of which are relevant for this study.

[15] However, the substituted and substituting words are not so often direct synonyms: only a third of all substitutions travel less than 3 hops on the hyponym-hypernym network defined by WordNet (direct synonyms count as 0 hops on this network), meaning that at least two thirds involve non-synonyms. A similar phenomenon is observed on the FA network, where about 104 clusters have substitutions traveling only 1 hop, 110 traveling 2 hops, 137 traveling 3, 72 traveling 4, and 13 traveling 5.

[16] s_g^0 is computed by summing, over all quotes undergoing a substitution, the ratio of the number of words in a quote that are from group g to the number of words in the same quote that could have been the substitution's target. If, for instance, half the content words of a given quote are nouns, such a quote contributes .5 to the total s_{nouns}^0. Further, to avoid possible autocorrelation effects due to substitutions belonging to the same cluster (which are likely not statistically independent and may lead to overly optimistic confidence intervals), we scale s_g and s_g^0 to count one for each cluster. That is, each quote cluster has a maximum contribution of 1, computed as the average contribution of all substitutions in that cluster.

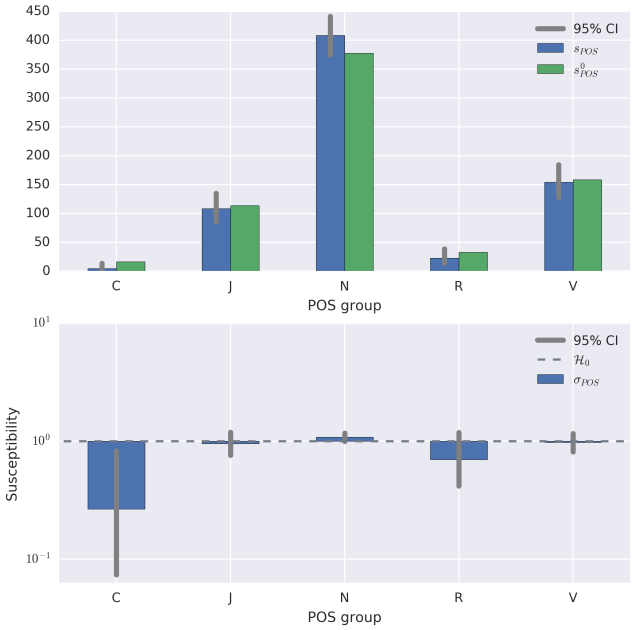


Figure 5. **POS-related results:** categories are simplified from the TreeTagger tag set: *C* means *Closed class-like* (see main text for details), *J* means adjective, *N* noun, *R* adverb, and *V* means verb. The top panel shows the actual s_{POS} and s^0_{POS} counts. The bottom panel shows the substitution susceptibility σ_{POS}, which is the ratio between the two previous counts. Confidence intervals are computed with the Goodman (1965) method for multinomial proportions.

Fig. 5 gathers the results for POS groups. A Goodman-based multinomial goodness-of-fit test (Goodman, 1965) shows that these categories have a significant effect on susceptibility ($p < .05$ in all substitution models), but this seems mostly due to the *Closed class-like*[17] and *Adverb* categories. Indeed, detailing which categories are out of their confidence region under $\mathcal{H}_0$ shows that susceptibility for closed class-likes is significantly below 1 (confirmed in all substitution models), as is that for adverbs in 13 of the 16 substitution models; none of the other categories are significantly different from $\mathcal{H}_0$ (except nouns which appear significantly above 1 in a single substitution model). While we acknowledge the low susceptibilities of adverbs and closed class-likes, these categories concern less than 7% of all substitutions under $\mathcal{H}_0$ (and even less in the actual data); it seems, then, that POS categories do not capture any strong bias in the selection of substitution targets.

The results for word features presented in Fig. 6, on the other hand, show marked effects for several features. Word frequency, Age of acquisition, and Number of letters each exhibit significant susceptibility variations (Goodman goodness-of-fit with $p < .05$ in all substitution models, $p < .001$ in most) consistent with known effects of those features on recall. High-frequency words, much easier to recall, are substituted about half as much as they would be at random; conversely low-frequency words, harder to recall, are substituted about 50% more than random. Age of acquisition and Number of letters show the opposite pattern, consistent with their negative correlation to word frequency ($-.4$ and $-.19$): words learned before 5 or 6 years old, or made of less than 5 letters, are substituted less than random, whereas words learned after 10 years old, or made of more than 8 letters, are substituted far more than random. Orthographic neighborhood density also shows a slight effect (significant at $p < .05$ in 15 of the 16 substitution models): words with very sparse neighborhoods are more substituted than random (which may seem counter-intuitive, but is probably because over 70% of those words have 7 letters or more). Clustering coefficient shows no effect on susceptibility, and neither does Number of synonyms: in particular, words with many synonyms do not attract substitutions more than random (in fact, half the substitution models show they have a slight tendency to be substituted less than random).

On the whole, the trends observed are consistent with known effects of word frequency, age of acquisition, and number of letters, indicating that the triggering of a substitution could behave quite similarly to word recall in standard tasks.

4.2 Variation

We now examine how words are modified when they are substituted, that is how their features change upon substitution. Considering a word w substituted for w', we measure how a feature ϕ of w varies when it is replaced with w', that is we look at $\phi(w')$ as a function of $\phi(w)$. Averaging this value over all start words such that $\phi(w) = f$ yields the mean variation for that feature value f, that is:[18]

$$\nu_\phi(f) = \langle \phi(w') \rangle_{\{w \to w' | \phi(w) = f\}}$$

[17] The *Closed class-like* category gathers all the POS groups representing closed class words (coordinating conjunctions, prepositions, subordinating conjunctions, modals and possessive endings). These groups, essentially made of stopwords, feature very low counts for both s (substitutions falling on stopwords are filtered out) and s^0 (stopwords are never counted as substitutable). While the susceptibility reported for the remaining words is left unbiased (as s and s^0 are equally affected), they represent a very small portion of all substitutions, which led us to group them together. Finally, we added to this meta-category the few POS groups that cover words entirely excluded from the analysis (foreign words, punctuation symbols and interjections), only sporadically present because of tagging fluctuations; hence the name *Closed class-like*.

[18] Similarly to what we do for susceptibility, we avoid possible autocorrelation effects by averaging start- and arrival-word features of substitutions from the same cluster into a single aggregate substitution per cluster.

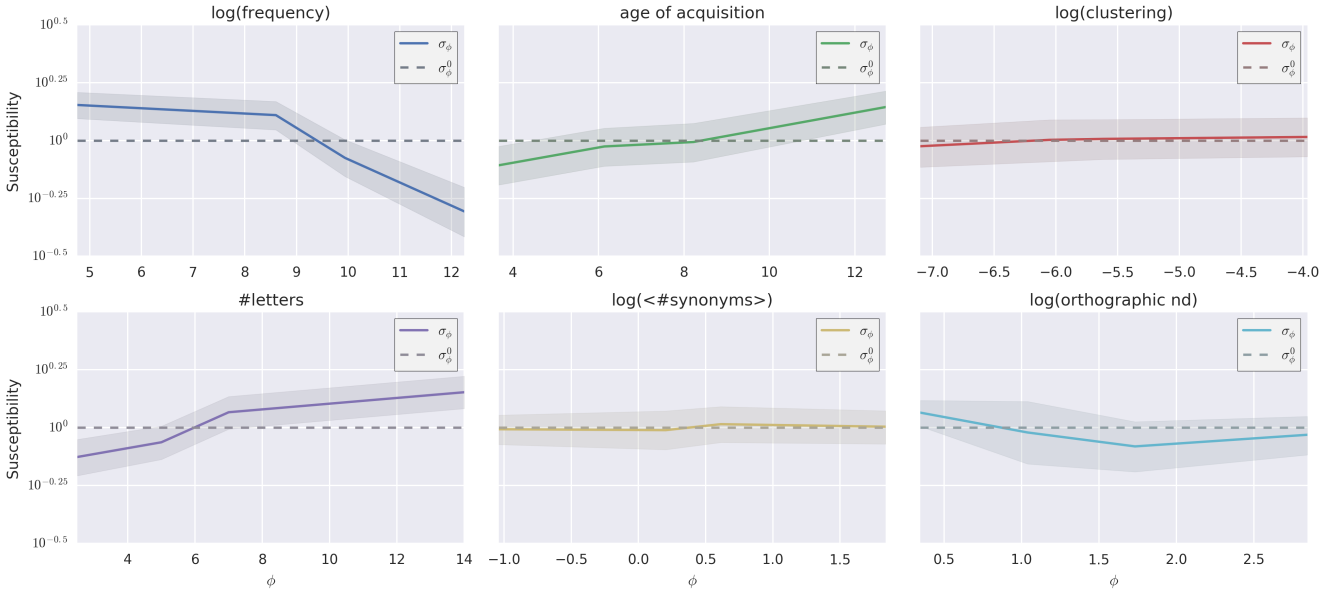


Figure 6. **Substitution susceptibility for feature values:** susceptibility to substitution versus feature value of a candidate word for substitution (binned by quartiles), with 95% asymptotic confidence intervals (Goodman-based multinomial).

We are interested in comparing the value of $\nu_\phi(f)$ to f itself, as this shows whether there is an attraction (or a repulsion) effect towards (respectively from) some values of each feature. In other words, plotting the $y = x$ line, we can see if substitutions tend to attract words towards some typical feature value or not — a standard procedure in the study of dynamical systems.

We also introduce two null hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_{00}$, to compare the actual variation of a word's feature to expected variations under unbiased transformations. $\mathcal{H}_0$ models the situation where the arrival word w' is randomly chosen from the whole pool of words available in the data set for that feature.[19] In this case, since $\phi(w')$ becomes a constant value in the above averaging (by definition w' does not depend on w anymore), the baseline variation under $\mathcal{H}_0$ may be rewritten as:

$$\nu^0_\phi(f) = \langle\phi\rangle$$

$\mathcal{H}_{00}$ models the situation where the arrival word w' is chosen *among immediate synonyms of the start word* w, i.e. an arrival word chosen among semantically plausible though still random words. In this case ν^{00}_ϕ does depend on f:[20]

$$\nu^{00}_\phi(f) = \left\langle \langle\phi(w')\rangle_{w'\in syn(w)} \right\rangle_{\{w|\phi(w)=f\}}$$

This approach yields a fine-grained view of how word features evolve upon substitution, on average, with respect to (a) the original feature (vs. $y = x$), (b) a random arrival (vs. ν^0_ϕ), and (c) an unbiased semantically plausible arrival (vs. ν^{00}_ϕ).

Results are gathered in Fig. 7. A first observation is that all graphs show the existence of a unique intersection of ν_ϕ with $y = x$, and the slope of ν_ϕ is smaller than 1 in absolute value, independently of the feature considered. This means that for each feature ϕ, whichever the value $\phi(w)$ of the disappearing word, the appearing word's feature value $\phi(w')$ will, on average, be closer to that feature's intersection of ν_ϕ with $y = x$.[21] In other words, beyond individual variation patterns, the substitution process exhibits a unique attractor for each feature. Note that this is also true under $\mathcal{H}_0$ or $\mathcal{H}_{00}$ (both null hypoth-

[19] For instance, when considering the feature "Clustering coefficient", the arrival word is randomly chosen among words present in the data set of FA norms.

[20] The actual implementation has an additional level of averaging since WordNet, used to get a word's synonyms, defines several meanings for a single given word, which we have no means of disambiguating. Therefore:

$$\nu^{00}_\phi(f) = \left\langle \left\langle \langle\phi(w')\rangle_{w'\in syn(m)} \right\rangle_{m\in meanings(w)} \right\rangle_{\{w|\phi(w)=f\}}$$

[21] This reasoning is standard in the analysis of dynamical systems (where the same transformation is applied to the whole system over and over), and becomes obvious when one manually simulates a substitution on the graph by picking a start value, using the ν_ϕ curve to obtain the corresponding arrival value, and comparing it to the start value: the arrival value is always closer to the intersection with $y = x$, meaning that that intersection is an attractor point for the substitution process. If the slope of ν_ϕ were greater than one (in absolute value), the arrival value would always be *farther* from the intersection than the start value was, making the intersection with $y = x$ a *repulsor* point. This is how the number of intersections with $y = x$ and the slope of ν_ϕ at those intersections characterize the behavior of substitutions.

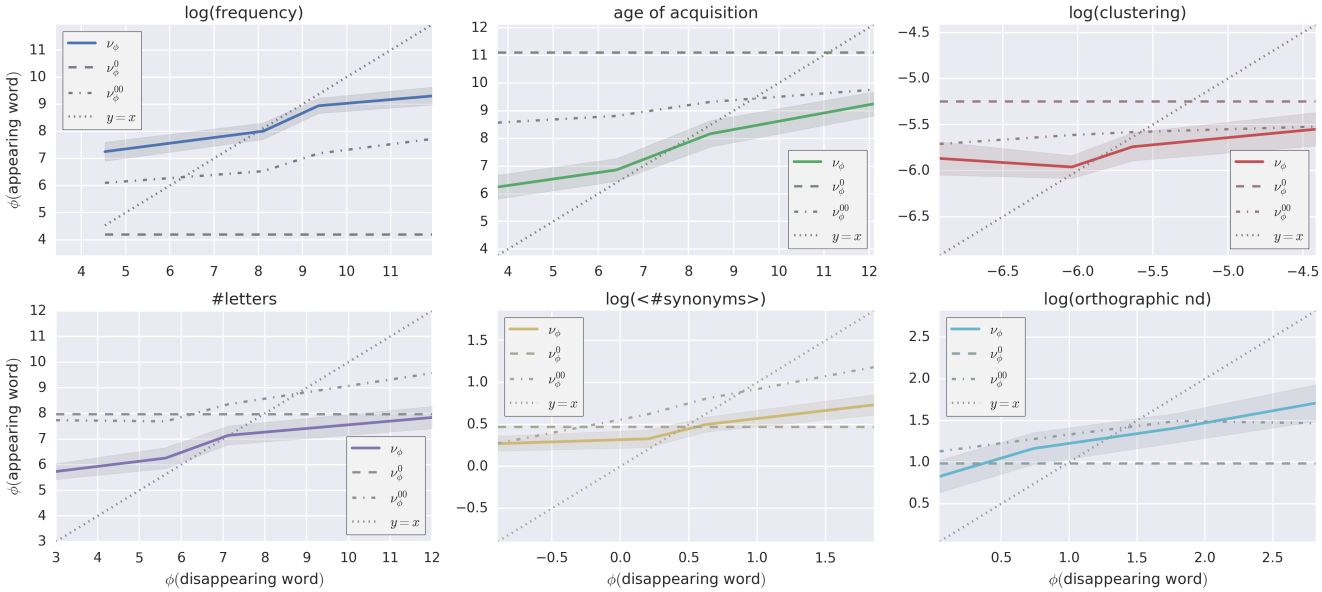


Figure 7. **Feature variation upon substitution:** ν_ϕ, average feature value of the appearing word as a function of the feature value of the disappearing word in a substitution (binned by quartiles), with 95% asymptotic confidence intervals based on Student's t-distribution. The overall position of the curve with respect to the dashed line representing $\mathcal{H}_0$ (constant ν_ϕ^0) indicates the direction of the cognitive bias compared to a purely random variation. The position with respect to the dash-dotted line representing $\mathcal{H}_{00}$ (ν_ϕ^{00}) indicates the bias compared to a semantically plausible random variation obtained by choosing a random synonym of the disappearing word. The intersection with $y = x$ marks the attractor value. The fact that all curves have slopes smaller than 1 in absolute value means that the substitution operation is contractile on average: it brings each feature closer to its own specific asymptotic range.

esis curves have single intersections with $y = x$ with slopes smaller than 1): the substitution process naturally leads to an attraction even under reasonable random conditions.

Second, the comparison with ν_ϕ^0 and ν_ϕ^{00} shows that there are two classes of attractors, depending on whether:

1. there is a triple intersection (of $y = x$, ν_ϕ, and ν_ϕ^0 or ν_ϕ^{00});
2. or ν_ϕ always remains above or below ν_ϕ^0 and ν_ϕ^{00}.

The first class (Number of synonyms and Orthographic neighborhood density) are features for which the substitution process only brings words slightly closer to ν_ϕ^0 (for Number of synonyms) or ν_ϕ^{00} (for Orthographic neighborhood density), and no uniform bias can be observed. The second class (comprising Word frequency, Age of acquisition, Clustering coefficient, and Number of letters) are features for which the substitution process has a clear bias, positive or negative, with respect to both the purely random situation ($\mathcal{H}_0$) and the semantically plausible random situation ($\mathcal{H}_{00}$).

Word frequency, with ν_ϕ always significantly above ν_ϕ^0 and ν_ϕ^{00}, exhibits a strong bias towards more frequent words. This, in turn, is consistent with the hypothesis that substitution is a recall process, since common words are favored over awkward ones. Age of acquisition, Clustering coefficient and Number of letters, on the other hand, exhibit a clear negative bias for the substitution process (except for high clustering values or very high number of letters). The three curves are significantly below their respective ν_ϕ^0 and ν_ϕ^{00} curves for most start values, which is consistent with the literature on recall: words learned earlier, with lower clustering coefficient or with fewer letters are easier to produce than average (Baddeley, Thomson, & Buchanan, 1975; Nelson et al., 2013; Zevin & Seidenberg, 2002). All these effects are significant with two-tailed t-tests at $p < .05$ (and more often $p < .001$) and were verified across the 16 substitution models.

To make sure our observations are not the product of correlations or interactions, we model the variations of the 6 features as a linear function of the start word's feature values:

$$\boldsymbol{\phi}(w') - \boldsymbol{\phi}(w) = \boldsymbol{A} + \boldsymbol{B} \cdot \boldsymbol{\phi}(w)$$

where $\boldsymbol{\phi}$ is the vector of all 6 features of a word, $\boldsymbol{A}$ is an intercept vector, and $\boldsymbol{B}$ is a 6×6 coefficients matrix. This regression achieves an overall R^2 of .33. The corresponding matrix of coefficients $\boldsymbol{B}$ is shown in Fig. 8: aside from Age of acquisition and Clustering coefficient on which word frequency has a slight effect, the variation of all other features depends solely on the disappearing word's same feature. In other words there is little to no interaction between a disappearing word's features in determining the variations that that word will undergo when substituted.

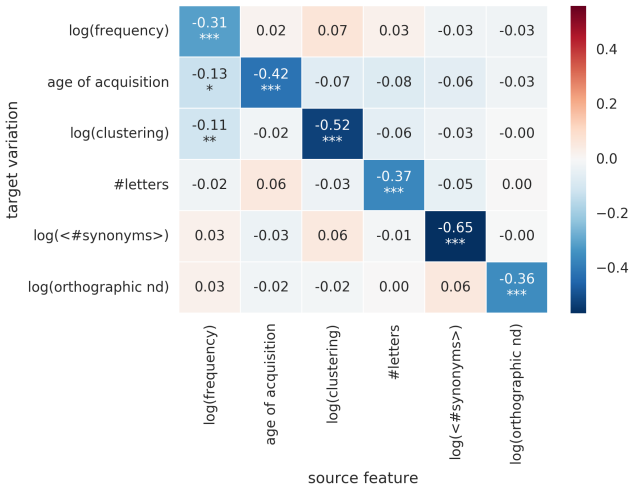


Figure 8. **Feature variations regression coefficients**: source feature values (columns) and feature variations (rows) were normalized to [0; 1] to ensure the coefficients are comparable. Significance levels for individual t-tests against the hypothesis of a null coefficient are denoted by stars below the corresponding coefficient (*** for $p \leq .001$, ** for $p \leq .01$, * for $p \leq .05$, and nothing when $p > .05$). Frequency has a slight effect on Age of acquisition and Clustering coefficient, with small coefficients compared to the respective diagonal ones. Aside from those two, only diagonal values are significantly non null.

To make things concrete, here is an example substitution taking place in the data set. Around mid-November 2008 several media websites reported the following quote from Burmese poet Saw Wai (arrested for one of his poems),

> "Senior general Than Shwe is foolish with power."

and a smaller number of media websites, and blogs, reported the following,

> "Senior general Than Shwe is **crazy** with power."

The word *foolish* is acquired at an average of 8.94 years old, appears 675 times in the data set, has a Clustering coefficient of 8.2×10^{-3} and is 7 letters long. The word it was replaced with, *crazy*, is acquired on average at 5.22 years old, appears about 4.1k times in the data set, has a Clustering coefficient of 1.7×10^{-3}, and is 5 letters long. Such a change, though minor in appearance, is a typical example of alteration along the lines shown by our results.

4.3 Sentence context

The alterations we study are always made in a context, that is while the author is writing. We wish to ask, therefore, if

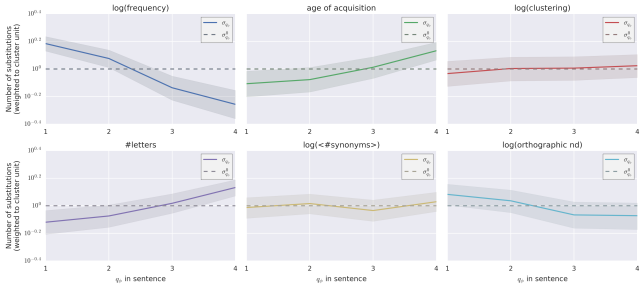


Figure 9. **Substitution susceptibility for in-quote feature quartiles:** susceptibility to substitution versus quartile of the feature distribution in the originating quote, with 95% asymptotic confidence intervals (Goodman-based multinomial).

taking that context into account can provide more insight into the substitution process. To do so we adapt the two observables presented above to capture some of the relationships between a word and the sentence it appears in.

Let us start with the first one: given a feature ϕ, we define the context-relative susceptibility to substitution with the following three steps. (1) For each quote in which a substitution appears, compute the distribution of ϕ values in that quote (excluding stopwords) and divide it into quartiles. (2) Count how many times each quartile (first, second, third or fourth) contains a word that is substituted. This procedure tells us, for $i \in \{1; 2; 3; 4\}$, how many times substitutions fall in the i-th quartile of each in-quote distribution of ϕ; in other words it gives us the numerator s_{q_i} for the computation of susceptibility, where q_i represents the i-th quartile of the distributions of ϕ in the quotes. (3) Finally divide each quartile count by its corresponding $s^0_{q_i}$, that is the number of times the i-th quartile would receive substitutions if targets were picked at random; since the random situation would equally distribute a fourth of all substitutions to each quartile, we divide by the number of substitutions divided by 4. Taking for instance word frequency, this measure tells us if words that have high- or low-frequency *compared to the quote they appear in* are more or less substituted than at random.

Surprisingly the results for this measure are no different from the context-free measure, as can be seen in Fig. 9: low-frequency words compared to the sentence are substituted much more than higher-frequency words, words learned earlier than the rest of the sentence are substituted less than words learned later, shorter words less than longer words, and words with scarce neighborhoods slightly more than words with denser neighborhoods. Clustering coefficient and Number of synonyms are, here again and across all substitution models, not significantly different from $\mathcal{H}_0$: with or without context, they do not seem relevant to the selection of substitution targets.

Feature variation is more easily extended to the context-

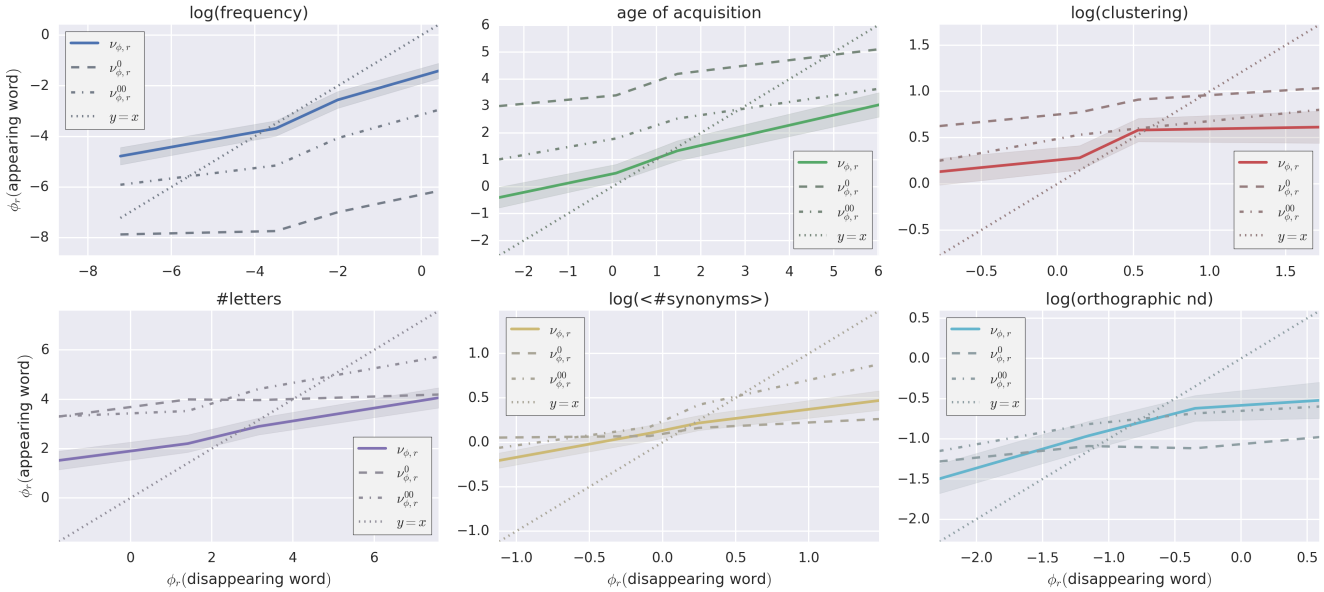


Figure 10. **Sentence-relative feature variation:** ν_{ϕ_r}, average sentence-relative feature value of the appearing word as a function of the sentence-relative value of the disappearing word (binned by quartiles), with 95% asymptotic confidence intervals based on Student's t-distribution. $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ are similarly converted to be sentence-relative. Attraction, magnitude and direction of bias with respect to null hypotheses are similar to Fig. 7. However, attractors are always positioned between sentence median ($y = 0$) on one side and $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ on the other side. Clustering coefficient, Number of synonyms and Orthographic neighborhood density are limit cases, with triple intersections with one of the null hypothesis curves.

relative case. To do so we consider all feature values relative to the median word feature in the sentence. That is, in all the equations of the previous subsection we replace $\phi(w)$ with:

$$\phi_r(w) = \phi(w) - \text{median}\,\{\phi(w)|w \in \text{sentence}\}$$

ν_ϕ, ν^0_ϕ and ν^{00}_ϕ each transpose to ν_{ϕ_r}, $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ (note that $\nu^0_{\phi_r}$ now depends on w since it is sentence-relative, whereas ν^0_ϕ did not).

The results for sentence-relative feature variations are gathered in Fig. 10. Here too, the behavior is strikingly similar to the context-free measure: a single attractor is visible for each feature, and the magnitude and direction of biases are near-identical to those for the previous measure. The values of the appearing words give an additional insight into the process, however: the attractor value of a feature, at the intersection of ν_{ϕ_r} and $y = x$, is always between the sentence median, on one side, and $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ on the other side (for Number of synonyms it is a triple intersection with $\nu^0_{\phi_r}$; for Clustering coefficient and Orthographic neighborhood density, a triple intersection with $\nu^{00}_{\phi_r}$). Substitutions, therefore, seem to attract words closer to the sentence median than what a random process would do. This is true with respect to both null hypotheses (semantically plausible or not) for Frequency, Age of Acquisition and Number of letters, and true with respect to at least one of the two null hypotheses for the remaining features.

On the whole, we observe a clear attraction pattern for each feature, with two different classes corresponding to the psychological relevance of each feature for the substitution process. More awkward words along relevant features (less frequent, learned later, or made of more letters), both globally and with respect to the sentence they appear in, are substituted more often than what would happen if targets were picked randomly in the sentences; conversely, more common words are substituted less. Finally, across all features, substituted words are attracted towards a point closer to the sentence median than what a random process, semantically plausible or not, would do.

5 Discussion

We initially aimed to connect the field of cultural evolution with psycholinguistics by asking if cultural attractors appear in a corpus of online news-related quotes gradually transformed by low-level biases. The data set we used imposed a few constraints on our analysis: first, it was necessary to infer source-destination links, an operation made more reliable when restricting the scope of transformations to very simple cases, which we did by focusing on single word substitutions. Second, contrary to laboratory experiments which produce data made of many repeated measures on a small number of cases (e.g. a given list of words), we have a great number of different cases (one case per cluster in

which substitutions are found, i.e. 698 cases), with very few measures on each of them (average 9, median 5). This rendered the prediction of individual words impractical: if we cannot compute a percentage of explained data for a given case, any approximate prediction will be heavily underestimated. This last factor, added to the potential for variation of external conditions when authors wrote the quotes, led us to use word features to analyze the transformations by aggregating over individual cases.

By characterizing substitutions with 6 features on the disappearing word, we show that authors preferentially substitute words known for being harder to recall: most prominently words with low frequency (Gregg, 1976), learned later (Dewhurst et al., 1998), or made up of more letters (Nickels & Howard, 2004), both globally and in comparison to the sentence they appear in. Further characterizing the substitutions by examining the variation of word features from disappearing to appearing words, we show: (a) that the operation is contractile on average, that is words are brought closer to an attractor point on each feature; (b) that authors produce words that are easier to remember than the average of synonyms of the disappearing word (a fact that is reflected in the position of the attraction point).

We do not actually observe quotes converging on a global scale towards attractors in their various dimensions. Indeed the limits of the data set do not allow us to infer chains of substitutions, and substitutions themselves are not the only type of transformation at work in the data set. Nonetheless, these findings (a) bring light to this simple type of transformation, and (b) are consistent with known psycholinguistic effects, with the hypothesis of cultural attractors in representations from everyday life, and with the lineage specificity discussed in the iterated learning literature (Claidière, Smith, et al., 2014; Cornish et al., n.d.). They are obtained by successfully applying knowledge from cognitive science to real-life complex data, a task that remains a challenge in the study of cultural evolution. More broadly, we believe that applying such data mining tools to manage the complexity of real-life data is a promising approach for the joint analysis of cognitive science and culture.

In the simple case presented here, however, much remains to be explored. Since it is clear that observing cognitive biases in such data is now possible, questions addressed in controlled laboratory situations could be opened by further research. One question concerns the influence of the context surrounding a quote, be it in terms of other quotes preceding it temporally or of text surrounding it in a post. A first step could be the application of results from Zaromb et al. (2006) who have shown, in the simpler task of recall of random word lists, that the source of prior-list intrusions can be predicted based on the associations those preceding lists have formed: in our case, a substitution could be triggered and directed by associations formed by preceding context. A further step would be to follow what Cornish et al. (n.d.) have shown about reciprocal influences between context and transformations (in their case, with transmission chains of artificial content). Indeed substitutions, and more generally all transformations, also participate in creating the context for later quotes. One can ask, therefore, what are the reciprocal effects between, on one side, the corpus-level evolution of quotes through iterated transformations, and on the other side, a gradual change in the properties of transformations operated because of the evolution of surrounding context. Such interactions have been shown to underlie the lineage specificity observed in transmission chains (Claidière, Smith, et al., 2014). Exploring how similar loop interactions happen in real-life data could indeed be the next step in understanding the coevolution of cultural content and the ways in which it is transformed. In our particular case, such insight could shed some light on how the feature attractors examined in this paper actually emerge, and help assess their potential role on this coevolution.

6 Concluding remarks

The theory of Epidemiology of Representations proposes a unifying framework for the study of cultural evolution. One of its core claims, the existence of cultural attractors, has been both a challenge to test empirically and a fruitful line to pursue in the study of cultural evolution. We aimed to contribute to testing this hypothesis by studying a simple everyday-life task where individuals are implicitly trying to reproduce quotations. To some extent, our work amounts to an out-of-laboratory experiment where we examine the influence of well-known word features on the accuracy of reproduction of short sentences. Our analysis of substitutions shows that words are attracted, in each dimension, to feature-specific values. Furthermore, the features' known effects in psycholinguistic experiments are reflected in the biases of these attraction points, meaning that the evolution of such quotations can be partially explained by known low-level cognitive biases. We believe that such an approach, which combines psycholinguistic knowledge and data mining tools, can be fruitfully developed to improve the study of cultural attractors and explore the reciprocal influences of cognition and culture.

Let us conclude by noting that the question of short- and long-term cultural evolution, and the approaches to study them, are becoming increasingly relevant to other fields. In biology in particular, work on evo-devo and non-genetic inheritance has accumulated evidence that is poorly accounted for by the modern synthesis of biological evolution, and is creating a demand for new or extended approaches to joint cultural and biological evolution (see Gilbert, Bosch, & Ledón-Rettig, 2015, for instance). Such an approach has long been called upon by anthropologists like Ingold (1998, 2004), in line with Mauss' initial works (Mauss,

1936), and the question is not entirely foreign to the enactive-representational debate in cognitive science. The study of cultural evolution will most likely benefit greatly from the growing interactions between these disciplines.

Acknowledgements

We are warmly grateful to Ana Sofia Morais for her precious feedback and advice on this research, and to Telmo Menezes, Jean-Philippe Cointet, Jean-Pierre Nadal, Sharon Peperkamp, Nicolas Claidière and Nicolas Baumard for useful suggestions and comments.

This work has also been partially supported by the French National Agency of Research (ANR) through the grant Algopol (ANR-12-CORD-0018).

Software colophon

Finally, this paper was developed using Python's scientific computing ecosystem (Millman & Aivazis, 2011). In particular, we directly used NumPy and SciPy (Walt, Colbert, & Varoquaux, 2011), Matplotlib (Hunter, 2007), Pandas (McKinney, 2010), scikit-learn (Pedregosa et al., 2011), NetworkX (Hagberg, Schult, & Swart, 2008), NLTK (Bird et al., 2009), IPython (Pérez & Granger, 2007), and many other libraries from the Python ecosystem. The software and analyses written for the paper are documented and published under a Free Software license. They can be found at `https://github.com/wehlutyk/brainscopypaste`.

References

Aunger, R. (2000). *Darwinizing culture: the status of memetics as a science*. Oxford; New York: Oxford University Press. (OCLC: 44518383)

Austerweil, J. L., Abbott, J. T., & Griffiths, T. L. (2012). Human memory search as a random walk in a semantic network. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 3041–3049). Curran Associates, Inc.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975, December). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589. doi: 10.1016/S0022-5371(75)80045-4

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python.* O'Reilly Media.

Bloch, M. (2000). A well-disposed social anthropologist's problems with memes. In *Darwinizing culture: the status of memetics as a science* (pp. 189–204). Oxford ; New York: Oxford University Press.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press. (OCLC: 11496588)

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: a quantitative approach*. Princeton, N.J.: Princeton University Press. (OCLC: 6863128)

Chan, K. Y., & Vitevitch, M. S. (2010, May). Network Structure Influences Speech Production. *Cognitive Science*, *34*(4), 685–697. doi: 10.1111/j.1551-6709.2010.01100.x

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014, May). How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1642). doi: 10.1098/rstb.2013.0368

Claidière, N., Smith, K., Kirby, S., & Fagot, J. (2014, December). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1797). doi: 10.1098/rspb.2014.1541

Claidière, N., & Sperber, D. (2007, March). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1), 89–111. doi: 10.1163/156853707X171829

Cointet, J. P., & Roth, C. (2009, August). Socio-semantic Dynamics in a Blog Network. In *International Conference on Computational Science and Engineering, 2009. CSE '09* (Vol. 4, pp. 114–121). doi: 10.1109/CSE.2009.105

Cornish, H., Smith, K., & Kirby, S. (n.d.). Systems from Sequences: an Iterated Learning Account of the Emergence of Systematic Structure in a Non-Linguistic Task. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 340–345). Austin, TX: Cognitive Science Society.

Danescu-Niculescu-Mizil, C., Cheng, J., Kleinberg, J., & Lee, L. (2012, March). You had me at hello: How phrasing affects memorability. *arXiv:1203.6360 [physics]*. (arXiv: 1203.6360)

Dawkins, R. (1976). Memes, the new replicator. In *The selfish gene* (pp. 189–201). Oxford: Oxford University Press.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. doi: 10.1037/h0046671

Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 284–298. doi: 10.1037/0278-7393.24.2.284

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35–41. doi: 10.2307/3033543

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001, October). Age-of-Acquisition, Word Frequency,

and Neighborhood Density Effects on Spoken Word Recognition by Children and Adults. *Journal of Memory and Language*, *45*(3), 468–492. doi: 10.1006/jmla.2000.2784

Gilbert, S. F., Bosch, T. C. G., & Ledón-Rettig, C. (2015, October). Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics*, *16*(10), 611–622. doi: 10.1038/nrg3982

Goodman, L. A. (1965, May). On Simultaneous Confidence Intervals for Multinomial Proportions. *Technometrics*, *7*(2), 247–254. doi: 10.1080/00401706.1965.10490252

Gregg, V. (1976). Word frequency, recognition and recall. In *Recall and recognition.* Oxford, England: John Wiley & Sons.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007, December). Google and the Mind Predicting Fluency With PageRank. *Psychological Science*, *18*(12), 1069–1076. doi: 10.1111/j.1467-9280.2007.02027.x

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information Diffusion Through Blogspace. In *Proceedings of the 13th International Conference on World Wide Web* (pp. 491–501). New York, NY, USA: ACM. doi: 10.1145/988672.988739

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008, August). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)* (pp. 11–15). Pasadena, CA USA.

Howard, M. W., & Kahana, M. J. (2002, January). When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language*, *46*(1), 85–98. doi: 10.1006/jmla.2001.2798

Hunter, J. D. (2007, May). Matplotlib: A 2d Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

Ingold, T. (1998, January). From complementarity to obviation: on dissolving the boundaries between social and biological anthropology, archaeology and psychology. *Zeitschrift für Ethnologie*, *123*(1), 21–52.

Ingold, T. (2004). Beyond biology and culture. The meaning of evolution in a relational world. *Social Anthropology*, *12*(2), 209–221. doi: 10.1111/j.1469-8676.2004.tb00102.x

Jefferies, E., Lambon Ralph, M. A., & Baddeley, A. D. (2004, November). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language*, *51*(4), 623–643. doi: 10.1016/j.jml.2004.07.005

Kirby, S., Cornish, H., & Smith, K. (2008, August). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. doi: 10.1073/pnas.0707835105

Kroeber, A. L. (1952). *The nature of culture.* Chicago: University of Chicago Press. (OCLC: 487751)

Kuper, A. (2000). If memes are the answer, what is the question? In *Darwinizing culture: the status of memetics as a science* (pp. 175–188).

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012, May). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. doi: 10.3758/s13428-012-0210-4

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the Dynamics of the News Cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 497–506). New York, NY, USA: ACM. doi: 10.1145/1557019.1557077

Liben-Nowell, D., & Kleinberg, J. (2008, March). Tracing information flow on a global scale using Internet chain-letter data. *Proceedings of the National Academy of Sciences*, *105*(12), 4633–4638. doi: 10.1073/pnas.0708471105

Lombardi, L., & Potter, M. C. (1992, December). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6), 713–733. doi: 10.1016/0749-596X(92)90036-W

MacCallum, R. M., Mauch, M., Burt, A., & Leroi, A. M. (2012, July). Evolution of music by public choice. *Proceedings of the National Academy of Sciences*, *109*(30), 12081–12086. doi: 10.1073/pnas.1203182109

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, *10*(1), 33–42. doi: 10.3758/BF03197623

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012, August). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLOS ONE*, *7*(8), e43230. doi: 10.1371/journal.pone.0043230

Mauss, M. (1936). Les techniques du corps. *Journal de Psychologie*, *32*(3-4).

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).

Mesoudi, A., & Whiten, A. (2008, November). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1509), 3489–3501. doi: 10.1098/rstb.2008.0129

Millman, K. J., & Aivazis, M. (2011, March). Python for

Scientists and Engineers. *Computing in Science & Engineering*, *13*(2), 9–12. doi: 10.1109/MCSE.2011.36

Miton, H., Claidière, N., & Mercier, H. (2015, July). Universal cognitive mechanisms explain the cultural success of bloodletting. *Evolution and Human Behavior*, *36*(4), 303–312. doi: 10.1016/j.evolhumbehav.2015.01.003

Morin, O. (2013, May). How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution. *Evolution and Human Behavior*, *34*(3), 222–229. doi: 10.1016/j.evolhumbehav.2013.01.004

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116–133. doi: 10.1037/0278-7393.21.1.116

Moussaïd, M., Brighton, H., & Gaissmaier, W. (2015, May). The amplification of risk in experimental diffusion chains. *Proceedings of the National Academy of Sciences*, *112*(18), 5631–5636. doi: 10.1073/pnas.1421883112

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013, May). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition*, *41*(6), 797–819. doi: 10.3758/s13421-013-0312-y

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. doi: 10.3758/BF03195588

Nickels, L., & Howard, D. (2004, February). Dissociating Effects of Number of Phonemes, Number of Syllables, and Syllabic Complexity on Word Production in Aphasia: It's the Number of Phonemes that Counts. *Cognitive Neuropsychology*, *21*(1), 57–78. doi: 10.1080/02643290342000122

Omodei, E., Poibeau, T., & Cointet, J.-P. (2012, September). Multi-Level Modeling of Quotation Families Morphogenesis. In *Proceedings of the ASE/IEEE 4th Intl. Conf. on Social Computing.* Amsterdam, Netherlands. (arXiv: 1209.4277)

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web* (Tech. Rep.). Stanford InfoLab.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Pérez, F., & Granger, B. E. (2007, May). IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, *9*(3), 21–29. doi: 10.1109/MCSE.2007.53

Potter, M. C., & Lombardi, L. (1990, December). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654. doi: 10.1016/0749-596X(90)90042-X

Potter, M. C., & Lombardi, L. (1998, April). Syntactic Priming in Immediate Recall of Sentences. *Journal of Memory and Language*, *38*(3), 265–282. doi: 10.1006/jmla.1997.2546

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998, September). A phoneme effect in visual word recognition. *Cognition*, *68*(3), B71–B80. doi: 10.1016/S0010-0277(98)00051-1

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. doi: 10.1037/0278-7393.21.4.803

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing.* Manchester, UK.

Simmons, M. P., Adamic, L. A., & Adar, E. (2011, July). Memes Online: Extracted, Subtracted, Injected, and Recollected. In *Fifth International AAAI Conference on Weblogs and Social Media.*

Sperber, D. (1996). *Explaining culture: a naturalistic approach*. Oxford, UK; Cambridge, Mass.: Blackwell.

Tamariz, M., & Kirby, S. (2016, April). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43. doi: 10.1016/j.copsyc.2015.09.003

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*(4), 344–354. doi: 10.1037/h0043150

Tulving, E. (1966, April). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *5*(2), 193–197. doi: 10.1016/S0022-5371(66)80016-6

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. (2011, March). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, *13*(2), 22–30. doi: 10.1109/MCSE.2011.37

Watts, D. J., & Strogatz, S. H. (1998, June). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. doi: 10.1038/30918

Weide, R. (1998). *The CMU Pronouncing Dictionary.* Retrieved 2016-06-30, from `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`

WordNet. (2010). *Princeton University "About WordNet".* Retrieved 2016-06-30, from `https://wordnet.princeton.edu/wordnet/`

Yonelinas, A. P. (2002, April). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. doi: 10.1006/jmla.2002.2864

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 792–804. doi: 10.1037/0278-7393.32.4.792

Zevin, J. D., & Seidenberg, M. S. (2002, July). Age of Acquisition Effects in Word Reading and Other Tasks. *Journal of Memory and Language*, *47*(1), 1–29. doi: 10.1006/jmla.2001.2834