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, UMIFRE 14
CNRS/MAEE/HU
Friedrichstr. 191, D-10117 Berlin

Camille Roth
CNRS
Centre Marc Bloch Berlin, UMIFRE 14 CNRS/MAEE/HU
Friedrichstr. 191, D-10117 Berlin

We describe reformulation processes within a large distributed system such as blogspace; showing how some specific features of public representations may be altered by bloggers when they freely reproduce them. To deal with robust and simple cultural representations, we focus on the evolution of quotations. In particular, we uncover some of the semantic and structural characteristics of individual words and the substitutions they undergo. Our work amounts to a large *in vivo* experiment where we appraise the impact of classically-influent psycholinguistic variables in the accuracy of the reproduction. We show that all variables remarkably exhibit a single attractor and are generally contractile. Even though the observed convergence patterns only partially explain quotation evolution, we shed light on a class of phenomena which are prone to constitute a key element of a broader empirically-grounded, attractor-based theory of cultural evolution. Update after introduction rework.

Keywords: word production; recollection bias; semantic network; cultural evolution; cultural attraction; data mining; big data; *in vivo* psycholinguistics

Introduction

#20: check text/flow/definitions for clarity against Gureckis' edited pdf

TOO GENERAL OR GIVES THE IMPRESSION OF TOO HIGH AMBITIONS. COMPARE TO OTHER INTROS TO JUST SAY WE'LL TALK ABOUT THIS.

Since the very beginnings of both social science and psychology, scientists have tried to capture the way cognition and culture influence each other. While this has been the subject of intense debate in the social sciences in the 20th century (starting with Durkheim's initial works, 1912, later tackled in earnest by e.g. Mauss' *Techniques of the Body*, 1936, Giddens' *Structuration Theory*, 1984, and Bourdieu's *Sens Pratique*, 1980), today's discussion is mostly structured by proponents from cognitive science.

These construe culture as an evolutionary process analogous and parallel to biological evolution (and especially the modern synthesis' account of it). That analogy can be traced a long way back in the 20th century and earlier, with milestones such as Kroeber's works (1952), Dawkins' *Memetics* (2006), 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 the theory of memetics cannot account for the levels of transformation culture undergoes as it is transmitted. Mesoudi and Whiten (2008) have discussed 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, show how cognitive biases have influenced the evolution of cultural artifacts over several centuries. Kirby, Cornish, and Smith (2013; 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

Correspondence should be directed to lerique@cmb.hu-berlin.de and roth@cmb.hu-berlin.de

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 opposed 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 studies cited above 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 representation 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 the email network (Liben-Nowell & Kleinberg, 2008); Cointet and Roth (2009) have described the reciprocal influence between the social network topology and the distribution of issues; Leskovec, Backstrom, and Kleinberg (2009b) 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) have studied the characteristics of particularly memorable quotes that circulate in those networks. We believe those works can connect the field of cultural evolution with psycholinguistics to advance the testing of cultural attractors.

add examples

To show this we analyze the way 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, when in fact they are implicitly required to copy them exactly. We can therefore assume that most transformations, especially the simple ones, are the result of automatic (i.e. hard to control) low-level cognitive biases of the authors.

Our question is as follows: given such representations that seem to evolve precisely because of the kind of automatic cognitive biases referred to in the theory of epidemiology of representations, do cultural attractors appear and how do cognitive biases participate in them? We chose to restrict our analysis to substitutions (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 observations to the particular data set we use, the methodological point we also make is left intact. By characterizing words using 6 well-known 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), but 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 describe the measures built to observe the 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.

[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 maintained inside) 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.

Related work

#20: check text/flow/definitions for clarity against Gureckis' edited pdf

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 & 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 method mentioned above 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).[3] 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. (2009b), 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 support 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 that 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 strength of semantic association 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 center 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 effects, where variations of the conditions in which sentences are read and produced have a chance of being statistically smoothed out.[4] If a cognitive bias in the substitu-

[3]Critics like Ingold (2007), however, have noted that the quantitative use of data imported from the social sciences risks overlooking the ontological debates in history and anthropology over the way to interpret such data.

[4]Aside from our lack of control on the precise conditions of en-

tion 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 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.

Methods

#20: check text/flow/definitions for clarity against Gureckis' edited pdf

We rely on a text corpus made of quotations extracted from online blog posts, and focus on their evolution. Indeed 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 such effects exist and are measurable 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 applied laboratory analyses to out-of-laboratory data, opening a path to explanations of actual (vs. simulated) cultural evolution with tools from cognitive science. As explained in the previous section, this is the reason we chose to use measures that can aggregate over all the transformations in the data set. this last sentence is unclear or not useful

To keep the analysis tractable, we focused on quotation transformations consisting in 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 further down (see "Substitution model"). To quantify those substitutions we decided to

coding 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.

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.

In vivo utterances

We used a quotation data set collected by Leskovec et al. (2009b), 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, Backstrom, & Kleinberg, 2009a).[5] The authors automatically extracted 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., 2009b, 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., 2009b, 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%).[6] 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 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.[7]

Word-level measures

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)[8], as well as **Number of Letters**, as a proxy to word

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

[6]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.

[7]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.

[8]The CMU Pronouncing Dictionary is included in the NTLK

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 stopword-like words.

Aside from these raw features, the systemic dimension of vocabulary (Cornish et al., 2013) 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).

To compute those 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. (2004) 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.[9]
- **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.[10]
- **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.[11] This tells us if a word belongs more or less to a local aggregate of equivalent words (from a free association point of view).

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 (a few words valued orders of magnitude higher than the vast majority of other words).[12]

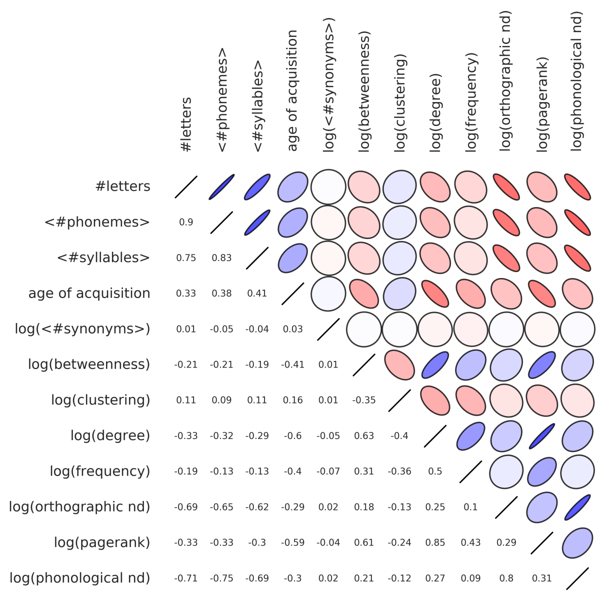


Figure 1. Spearman correlations in the initial set of features

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

package (Bird, Klein, & Loper, 2009), the natural language processing toolkit we used for the analysis.

[9]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.

[10]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.

[11]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).

[12]The distributions of original and log-transformed features appears in the supplementary material, along with scatter plots for each feature pair, summarized below with correlation values.

acquisition, betweenness, degree, and pagerank centralities, with absolute pairwise correlations at .41, .6, .59, .63, .61 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.[13]

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 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 (the analysis on the complete set of features can be found in the appendix).[14]

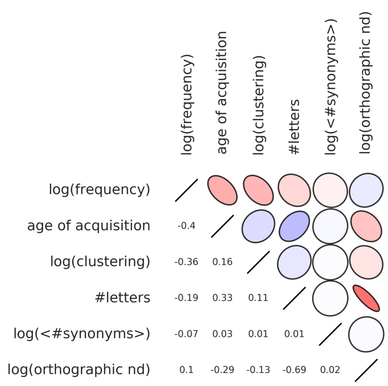


Figure 2. Spearman correlations in the filtered set of features

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 of 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'. q and q'' differ by two substitutions, so we discard the possibility that one was written based on the other (this 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.)

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.

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 discretized

[13] 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).

[14] 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 all words 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.

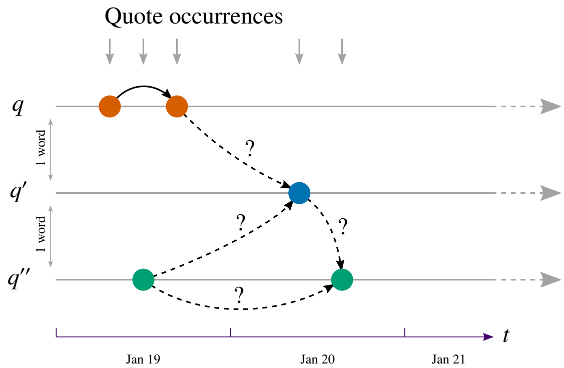


Figure 3. **Possible paths from occurrence to occurrence.** q, q' and q'' are three quotation variants belonging to the same cluster. q and q'' differ by two words, but q' differs from both q and q'' by a word. The second occurrence of q can safely be considered a faithful copy of the first, but the 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.

(aligning the end of a bin to midnight, as in the model presented above) or kept continuous (for each occurrence, use a bin that ends precisely at that occurrence); (2) **bin span**, which can be 24 hours (as in the model above) or can be extended to start at the very first occurrence in the quotation family. The other two parameters define 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 not (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 discretely positioned bins spanning 1 day: 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.[15]

In practice for the model first introduced above, from the 2.60m initial occurrences spread into 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 stop words, minor spelling changes (e.g. center/centre, November/Nov, Senator/Sen), abbreviations, spelled out numbers, words unknown to WordNet, and deletions in substrings (which can appear as substitutions of non-deleted words); this eventually yields 6,318 valid substitutions (before merging substitutions that share the same destination).[16]

Results

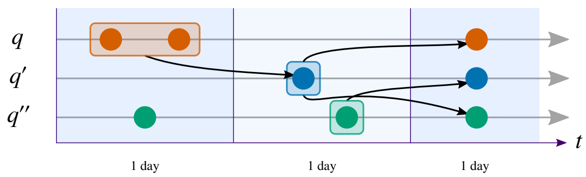
#16: add a quick description of what types of substitutions we see, then moving on to the observables since it's not what principally interests us.

order of presentation: - distance in substitution - susceptibility of POS - susceptibility of bins of absolute value - variation of bins of absolute value - susceptibility of quartiles in sentence - variation of quartiles in sentence (bins->bins or quartile->bins or quartile->quartile) - PCA of susceptibilities absolute, quartile, absolute+quartile

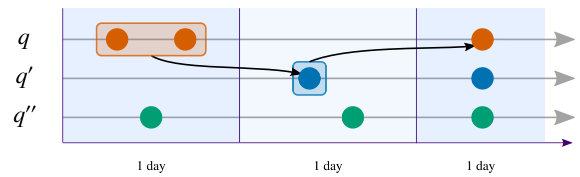
We may now use this substitution model to formulate a family of psycholinguistic hypotheses describing the role of each feature in the accuracy of the reformulation. To this end, we build two main observables for each word feature. First, we measure the susceptibility for words to be the target of a substitution in a quote, knowing that there has been a variation, in order to show which semantic features are the

[15] We checked that this restriction does not bias the results discussed below by extending our protocol to two-substitution transformations. The results and graphics for the 16 additional models involved are available in the code repository for this paper: `https://github.com/wehlutyk/brainscopypaste`.

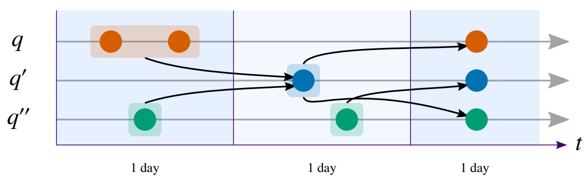
[16] Manually coding a random subset of 100 substitutions to evaluate this last filter showed that 84 were true negatives, 5 were false positives, and 11 true positives, giving a recall score of .688. Precision was evaluated over a random subset of 100 *kept* substitutions, showing a score of .87. Finally, 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.



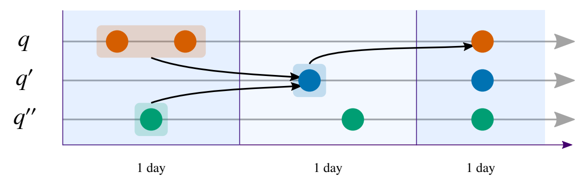
(a) Source must be majority in preceding bin, destination can be anything



(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 discretely positioned bins spanning 1 day (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 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.

most likely to "attract" a substitution under this condition. Second, we measure the change in word feature upon substitution, looking at the variation of a given feature between start and arrival words.

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*, and 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 it is impossible to know whether there has been perfect human reformulation or simply digital copy-pasting of a source ("Ctrl-C/Ctrl-V"). Furthermore, 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.

Susceptibility

#19: explain if we're susceptible to bias from single-substitution

We say that a word is *substitutable* if it appears in a quote which undergoes a substitution, whether that substitution operates on that word or on another one. Word substitution susceptibility is computed as the ratio of the number of times s_w a word is substituted to the number of times p_w that word appears in a substitutable position, that is $^{s_w}/_{p_w}$. In other words, it measures how often a word w actually gets substituted, compared to how often it could have been substituted (because it appears in quotes undergoing substitution).

Now, for a given feature ϕ, we obtain the mean susceptibility $\sigma_\phi(f)$ for the feature value f by averaging this ratio over all words such that $\phi(w) = f$, that is:

$$\sigma_\phi(f) = \left\langle \frac{s_w}{p_w} \right\rangle_{\{w|\phi(w)=f\}}$$

Put shortly, susceptibility focuses on the selection of start words involved in substitutions, measuring the effect of features at the moment preceding the substitution when it is not yet known which word in the quotation will be substituted.

#16: (1) show POS don't differ, (2) note that stopwords are excluded, which excludes POS open/closed discussion

#16: explain we do this on quartiles, rework results discussion

#16: explicit all significance claims, referring to the CIs

#17: show other feature variations and other models in annex

#15: relate to missed literature

Results for this measure are gathered in Fig. **??**. They first show an obvious strong effect of Word frequency: the more frequent a word, the less likely it is to attract substitutions. Indeed, susceptibility goes from .33 for low-frequency words down to nearly 0 for very high-frequency words. To make things clear, this value of .33 means that low-frequency words, when present in a quote undergoing a substitution, are the ones being substituted 33% of the time on average.

The other features — Age of acquisition, Number of phonemes, Clustering coefficient and Number of synonyms — do not seem to exhibit any particularly significant effect on susceptibility. If we set aside the values for low Number of phonemes, for each of these features it is indeed possible to draw a constant line which always remains within the respective confidence intervals. If these variables have an effect, it is by no means as strong as it is for Word frequency. This is remarkably clear for Clustering coefficient and Age of acquisition, where susceptibility values remain within quite small intervals (respectively [.13 – .18] and [.16 – .20]). We may notice a slight effect for the lowest values of Number of synonyms and Number of phonemes, where the mean susceptibility is almost half as high as the average of the other values (respectively .09 vs. .16, and .11 vs. .17). Keeping in mind the poor statistical significance of this effect, we could still wonder if the shortest words and words with fewest synonyms are significantly less susceptible to substitution. To further examine this phenomenon, we plotted the two-dimensional map of susceptibility values for these two features (see heatmap at the bottom right of Fig. **??**). Even if there are a few outlier cells, values tend to navigate around the mean value (.16) with little obvious regularity (except for a low number of synonyms, consistent with the unidimensional graph). On the whole, this makes it relatively hard to draw any conclusion as regards the direction of an effect, except for the least populated value ranges (which as a result are also less significant).

All in all, apart from Word frequency and despite some local tendencies, in general these results do not allow us to conclude to a marked effect of the selected psycholinguistic features on substitution susceptibility. We may therefore globally assume that substitution targets are chosen in a more or less uniform way with respect to these features.

Variation

We can thus show how words are modified once we know they are substituted, that is how their features are modified by said substitution. Considering a word w substituted for w', we measure how the feature 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:[17]

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

Of prime interest is the comparison of the value of $\nu_\phi(f)$ with respect to f, as it 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 converge towards some typical value of a word feature or not — as is classically done in the study of dynamical systems.

We also introduce a null hypothesis $\mathcal{H}_0$ to compare the actual variation of a word's feature to its expected variation, assuming the arrival word w' was randomly chosen from the whole pool of words available in the data set for that feature.[18] 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:[19]

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

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

#16: (1) add H00 lines to account for semantic similarity (nothing more fancy since it involves a threshold), (2) rework results discussion

#17: show other feature variations and other models in annex

#17: explain binning

#18: tone down claims about contractile: it's a possible hypothesis if this were the only process, but not observed with the mix of all other processes

#15: relate to missed literature

[17] 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 first average substitutions over each cluster, by considering the average of arrival word features for a given start word.

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

[19] We additionally considered an alternative null hypothesis, denoted $\mathcal{H}_{00}$, where the arrival word is randomly chosen *among immediate synonyms of the start word*, that is an arrival word chosen among semantically plausible though still random words. In this case w'_{00} does depend on w. Our conclusions hold under this second null hypothesis, so for the sake of clarity we chose to keep the simpler $\mathcal{H}_0$.

Results are gathered in Fig. **??**. We can do a first striking observation: all graphs show the existence of a unique intersection of ν_ϕ with $y = x$, while the slope of ν_ϕ is smaller than 1, independently of the feature considered. In other words, beyond individual variation patterns, the substitution process is contractile for all the features, and each of them therefore exhibits a unique attractor. Second, the comparison with ν^0_ϕ shows that there are two classes of attractors, depending on whether:

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

The first class (Number of phonemes and Number of synonyms) are features for which the substitution process only brings words slightly closer to ν^0_ϕ, and no uniform bias can be observed.

On the other hand, the second class (comprising Word frequency, Age of acquisition, and Clustering coefficient) are features for which the substitution process has a clear bias, positive or negative, with respect to the purely random situation ($\mathcal{H}_0$).

Word frequency, with ν_ϕ always significantly above ν^0_ϕ, 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, while it goes against the idea that it could be a familiarity process, where awkward terms would be favored.

Age of acquisition and Clustering coefficient, on the other hand, exhibit a clear negative bias for the substitution process. Both curves are significantly below their respective ν^0_ϕ values, which is consistent with the literature on recall: words learned earlier and words with lower clustering coefficient are easier to produce than average (**??**). Clustering coefficient has the additional particularity that, on average, the destination word does not depend on the start word; that is on average, substitutions will always produce words with a clustering coefficient around $\exp(-2.4) \simeq .1$.

To make things concrete, here is an example substitution taking place in the data set. At the end of January 2009, many media websites reported the following quote,

> "The massive economic upheaval being experienced across the globe is sparing no one in the consumer electronics world."

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

> "The massive economic upheaval being experienced across the **world** is sparing no one in the consumer electronics world."

The word *globe* is acquired at an average of 6.5 years old, appears about 3.5k times in the data set, and has a Clustering coefficient of .24. The word it was replaced with, *world*, is acquired on average at 5.3 years old, appears about 146k times in the data set, and has a Clustering coefficient of .05. (Both words have four phonemes.) Such a change, though minor in appearance, is a typical example of alteration along the lines shown by our results.

#16: add the taking-context-into-account parts, potentially moving the non-relative results discussions to here: (1) explain and add sentence-relative susceptibility by quartiles, which has a perfectly clear interpretation + results discussion, (2) explain and add sentence-relative variations + results discussion (3) link those results to non sentence-relative values

#17: show other feature variations and other models in annex

#15: relate to missed literature

We thus observe a clear convergence pattern for each feature, with two different classes corresponding to the psychological relevance of each feature for the substitution process. Taken as a dynamical system where substitutions are repeatedly applied, Number of phonemes and Number of synonyms will simply converge towards their average value in the FA corpus (i.e. ν^0_ϕ), while Word frequency, Age of acquisition and Clustering coefficient, consistent with the literature, will converge towards significantly biased values indicated by the intersection with $y = x$ (respectively, a frequency of $\exp(9.1) \simeq 9000$, an acquisition age slightly below 8, and a Clustering coefficient of .1).

Discussion

Discuss related to introduction. Attractors, lineage with specification, what we couldn't observe, how it fits into Kirby.

#15: relate to missed literature

ADDTHIS: We also chose exploratory vs. predictive to give a detailed view of what happens and because there's too many possible things to predict.

ADDTHIS: By characterizing substitutions with 6 features on the disappearing and appearing words, we identify what makes a substitution more likely, and how a word changes when it is substituted. Consistent with known effects in linguistics, we observe that low-frequency words and words learned later in development are more susceptible to substitution than other words. Looking at the context those words appear in, we observe a marked effect for substitution of extreme words in a sentence (either very high-valued or very low-valued features compared to sentence average, except for word frequency). Focusing on how words are transformed, we see that the appearing words have significantly higher frequency and lower age-of-acquisition than synonyms of the disappearing word. Finally, the patterns we observe are also consistent with an attraction of each of the features towards a (feature-specific) asymptotic value.

ADDTHIS: It is possible however, that these attractors appear due to an interaction between biases and sentence context, making it a contingency rather than a rule. This is not really dealt with (context, aside from relevance) by Sperber.

Attraction can also be defined on any number of dimensions. It can be on the structure, on anything, so saying there could be an attraction while not specifying the dimension is really meaningless. What's more important is to look at a specific dimension, and see if there's attraction on that one, as we did here for features.

We could've done also on semantic grouping, predicting the new word based on semantic similarity (or on frequent dyads, i.e. collocation with previous word the same way Zaromb et al. 2006 explain PLIs), and predicting the disappearing word based on the cost of doing such a substitution (lower cost -> higher prob of substitution). The point is, there's decades and many fields of psycholinguistics, and we can connect each of them with this question.

All of this is possible with our software that we published.

ADDTHIS: Taking context into account is more than what we did. For instance, building on Zaromb 2006, you could imagine that the substitutions appear because the word preceding the substituted one appears in another dyad a lot more than this one, triggering a substitution (Zaromb's associative vs. contextual retrieval processes in recall).

ADDTHIS: Soooooo... following literature on word lists (Zaromb and DRM): - we could predict the new word in substitution? Take the strongest average association to words in the sentence. - we could predict substituted word? Take the word following the one that triggers the new word. Problems: - we probably won't find the exact word, but one similar to it (even Zaromb can't predict the exact word, they don't try, they just check it comes from the right list). How to evaluate that? - if computing LSA/LDA on the corpus (which probably isn't adapted because of the short sentence nature of the data -> topics = quote families), it's tautological unless you suppose substitutions have a negligible effect. Explain that to the reviewer.

Again justify our approach w/ features by the shape of our data: few substitutions per cluster (avg. 9), and substitutions on relatively few clusters overall -> opposite situation to a few lists repeated through 50 subjects where frequency of transition has meaning. Here for each word, it's nearly one shot. So we have to categorize words (by using low number of features -> known features best) to get frequencies. From there, you can predict many many things, so better describe.

Concluding remarks

#14: link to introduction discussion: (1) this can be a model system, (2) convergence can be looked for in any dimension, but that doesn't make a theory, so Epidemiology of Representations is all nice, but: (3a) taken simplistically it predicts obviously simplistic stuff (all quotes CV. to a single quote), (3b) taken more realistically (many dimensions in life) it gives some ideas, but it's not clear it's a core principle (3c) we need more controlled investigation to fuel the discussion and see how relevant it is.

ADDTHIS: On the other side an enactive proposition which anthropologists like Ingold, in line with Mauss' works, are calling for [Citation needed], is being developed by Froese, Di Paolo, and De Jaegher among others [Multiple citations needed].

The question is also gaining relevance in other fields, as work in evo-devo and non-genetic inheritance is accumulating evidence not accounted for by the modern synthesis [Citation needed]; these discoveries are creating demand for new or extended approaches to life evolution that unify its different levels, as well as creative empirical methods to test the predictions these approaches make [Citation needed].

We aimed to contribute to the empirical understanding of representation transformation processes by studying a simple task where individuals are *implicitly* trying to reproduce textual content. To some extent, our work amounts to a large *in vivo* experiment where we appraise the impact of classically-influent psycholinguistic variables in the accuracy of the reproduction. In more detail, we describe the joint properties of the substituted and substituting terms in the reformulation by individuals of a specific type of utterances (quotations).

#18: tone down claims about contractile: it's a possible hypothesis if this were the only process, but not observed with the mix of all other processes

For each of the selected psycholinguistic variables, we demonstrate the existence of attractor values in the underlying variable spaces. More precisely, beyond the interpretation of our results for each variable, we notice that all variables remarkably exhibit a single attractor and are generally contractile — as such, even though the observed convergence patterns only partially explain quotation evolution, we shed light on a class of phenomena which are susceptible to constitute a key element of a broader empirically-grounded, attractor-based theory of cultural evolution.

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).

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. Retrieved 2016-06-28, from http://papers.nips.cc/paper/4761-human-memory-search-as-a-random-walk-in-a-semantic-network.pdf

Bird, S., Klein, E., & Loper, E. (2009). *NLTK Book*. Retrieved 2016-06-30, from http://www.nltk.org/book/

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).

Bourdieu, P. (1980). *Le sens pratique*. Paris: Editions de Minuit. (OCLC: 299354015)

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. Retrieved 2016-06-28, from http://onlinelibrary.wiley.com.gate3.inist.fr/doi/10.1111/j.1551-6709.2010.01100.x/abstract

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), 20130368. Retrieved 2015-03-31, from http://classic.rstb.royalsocietypublishing.org/content/369/1642/20130368

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), 20141541. Retrieved 2015-03-31, from http://rspb.royalsocietypublishing.org/content/281/1797/20141541

Claidière, N., & Sperber, D. (2007, March). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1), 89–111. Retrieved 2016-05-18, from http://booksandjournals.brillonline.com.gate3.inist.fr/content/journals/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).

Cornish, H., Smith, K., & Kirby, S. (2013). 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.*

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]*. Retrieved 2016-02-12, from http://arxiv.org/abs/1203.6360 (arXiv: 1203.6360)

Dawkins, R. (2006). *The selfish gene*. Oxford; New York: 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.

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.

Durkheim, E. (1912). *Les formes élémentaires de la vie religieuse le système totémique en Australie*. Paris: F. Alcan. (OCLC: 489968385)

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35–41. Retrieved 2016-07-01, from http://www.jstor.org/stable/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. Retrieved 2016-06-28, from http://www.sciencedirect.com/science/article/pii/S074

Giddens, A. (1984). *The constitution of society: outline of the theory of structuration*. (OCLC: 11029282)

Gregg, V. (1976). Word frequency, recognition and recall. In *Recall and recognition* (pp. x, 275). 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. Retrieved 2016-06-28, from http://pss.sagepub.com/content/18/12/1069

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. Retrieved 2016-06-22, from http://doi.acm.org/10.1145/988672.988739

Howard, M. W., & Kahana, M. J. (2002, January). When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language*, *46*(1), 85–98. Retrieved 2016-06-28, from http://www.sciencedirect.com/science/article/pii/S0749596X01927986

Ingold, T. (2007). The trouble with 'evolutionary biology'. *Anthropology Today*, *23*(2), 13–17. Retrieved 2015-02-13, from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8322.2007.00497.x/abstract

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. Retrieved 2015-07-30, from http://www.sciencedirect.com/science/article/pii/S0749596X04000853

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. Retrieved 2016-06-18, from http://www.pnas.org/content/105/31/10681

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. Retrieved 2016-06-30, from http://link.springer.com.gate3.inist.fr/article/10.3758/s13428-012-0210-4

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009a). *MemeTracker: tracking news phrases over the web.* Retrieved 2016-06-29, from http://www.memetracker.org/

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009b). 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. Retrieved 2016-02-21, from http://doi.acm.org/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. Retrieved 2016-06-22, from http://www.pnas.org/content/105/12/4633

Lombardi, L., & Potter, M. C. (1992, December). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6), 713–733. Retrieved 2015-07-30, from http://www.sciencedirect.com/science/article/pii/0749

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. Retrieved 2016-05-18, from http://www.pnas.org/content/109/30/12081

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, *10*(1), 33–42. Retrieved 2016-06-30, from http://link.springer.com.gate3.inist.fr/article/10.37

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. Retrieved 2016-03-18, from http://journals.plos.org/plosone/article?id=10.1371/j

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

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. Retrieved 2016-06-18, from http://rstb.royalsocietypublishing.org/content/363/15

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. Retrieved 2016-06-25, from http://www.sciencedirect.com/science/article/pii/S109

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. Retrieved 2016-05-18, from http://www.ehbonline.org/article/S1090513813000056/ab

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.

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. Retrieved 2015-06-10, from http://www.pnas.org/content/112/18/5631

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. Retrieved 2016-06-28, from http://link.springer.com.gate3.inist.fr/article/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. Retrieved 2016-06-30, from http://link.springer.com.gate3.inist.fr/article/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. Retrieved 2016-06-28, from http://dx.doi.org/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. Retrieved 2016-06-27, from http://arxiv.org/abs/1209.4277 (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. Retrieved 2016-07-12, from http://www.jmlr.org/papers/v12/pedregosa11a.html

Potter, M. C., & Lombardi, L. (1990, December). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654. Retrieved 2015-07-30, from http://www.sciencedirect.com/science/article/pii/0749596X9090002Z

Potter, M. C., & Lombardi, L. (1998, April). Syntactic Priming in Immediate Recall of Sentences. *Journal of Memory and Language*, *38*(3), 265–282. Retrieved 2015-07-30, from http://www.sciencedirect.com/science/article/pii/S0749596X97925468

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. Retrieved 2016-06-28, from http://www.sciencedirect.com/science/article/pii/S001

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.

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.* Retrieved 2016-02-12, from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper

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. Retrieved 2016-06-18, from http://www.sciencedirect.com/science/article/pii/S235

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

Tulving, E. (1966). Subjective Organization and Effects of Repetition in Multi-Trial Free-Recall Learning.

Watts, D. J., & Strogatz, S. H. (1998, June). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. Retrieved 2016-07-01, from http://www.nature.com.gate3.inist.fr/nature/journal/v

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. Retrieved 2014-12-12, from http://www.sciencedirect.com/science/article/pii/S074

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.

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. Retrieved 2016-06-28, from http://www.sciencedirect.com/science/article/pii/S074