Livrable L3.2-4

“Report on content mutations”

(M36)

Projet Algopol

Programme ANR CONTINT 2012

Partenaire responsable: CAMS

This deliverable takes place within the context of subtask 3.2 of the original submission document. Subtask 3.2 partly consisted of the understanding of curation processes and, more importantly, *content evolution*, ideally with respect to source authority. This issue is admittedly broad and strongly connects with contemporary research topics in cognitive science, mainly in relation to *cultural epidemiology*. In a nutshell, this research program aims at describing the reasons behind the contagiosity of ideas and representations: both intrinsically (with respect to the cognitive characteristics of the content itself) and extrinsically (with respect to the property of sources, users, including authority phenomena).

Eventually, most of the work under this topic within AlgoPol has been devoted to the preliminary and necessary understanding of content evolution itself, at a sentence level, essentially in online communities. The project expired before we could explore further the relationship between authority and content diffusion and, perhaps more importantly, content transformation, which we now have to leave for future research.

Here, we principally exploited the recent possibility of tracking very similar content in large online datasets (Leskovec et al., 2009) in order to empirically describe how information is being altered by being (iteratively) re-formulated by individuals, in a decentralized manner.

The present document thus includes four related manuscripts:

1. "Multi-Level Modeling of Quotation Families Morphogenesis", by Elisa Omodei (CAMS), Thierry Poibeau (ENS) and Jean-Philippe Cointet (INRA SenS / as part of CAMS), on the evolution of quotations in blogspace (such as "In a speech, X said: 'We should do our best to (...)'"). This first work had started shortly before the official beginning of Algopol and has been completed in the first months of the project. It features results on the stability of quotations at a meso-level, for example with respect to their length and popularity. It has been presented at the 4th ASE/IEEE Intl. Conf. on Social Computing "SocialCom 2012", in September 2012.

2. "How do we copy and paste? The semantic drift of quotations in blogspace", by Sébastien Lerique and Camille Roth, currently in (major) revision for a publication in *Cognitive Science*. This study makes use of the same dataset as above to provide a finer description of psycholinguistic biases in the reformulation of quotations. We focused on a more micro level than in the first manuscript: we looked at the evolution of words to show for instance that some words are more likely to be replaced with others based on their age of acquisition, popularity, length. This paper more broadly demonstrates the existence of cognitive attractors in these reformulation processes.

3. "Pour une étude du contexte d'interprétation", by Sébastien Lerique, provides an epistemological reflection on the prospect of studying content mutation by taking context into account, thereby taking some hindsight on the above studies focused on the sentence level. It has been submitted to a special issue of "Travaux de Linguistique" upon the solicitation of the editors.

4. "The Gistr Platform", by Sébastien Lerique, is a white paper describing the functioning of an experimental platform called "Gistr" meant at studying the *in vivo* transformation of content, in order to provide a reference point on the *in vitro* observations on empirical datasets stemming from online communities. Gistr, as such, is also a software deliverable of the project, which will be used to further refine the understanding of reformulation processes.

Multi-Level Modeling of Quotation Families Morphogenesis

Published in the Proceedings of the ASE/IEEE 4th Intl. Conf. on Social Computing "SocialCom 2012"
Sep. 3-5, 2012, Amsterdam, NL

Elisa Omodei
LaTTiCe (CNRS, ENS, Paris 3)
ISC-PIF
Email: elisa.omodei@ens.fr

Thierry Poibeau
LaTTiCe (CNRS, ENS, Paris 3)
Email: thierry.poibeau@ens.fr

Jean-Philippe Cointet
INRA-SenS
INRA, IFRIS, ISC-PIF
Email: jean-philippe.cointet@polytechnique.edu

***Abstract*—This paper investigates cultural dynamics in social media by examining the proliferation and diversification of clearly-cut pieces of content: quoted texts. In line with the pioneering work of Leskovec et al. [1] and Simmons et al. [2] on meme dynamics we investigate in deep the transformations that quotations published online undergo during their diffusion. We deliberately put aside the structure of the social network as well as the dynamical patterns pertaining to the diffusion process to focus on the way quotations are changed, how often they are modified and to what extent these changes shape diverse families and sub-families of quotations. Following a biological metaphor, we try to understand in which way *mutations* can transform quotations at different scales and how mutation rates depend on various properties of the quotations.**

I. Introduction

"Memeticists", whose forefather Richard Dawkins [3] was the first to coin the term *meme* in 1976 in analogy to gene, defend the thesis that culture is subjected to evolutionary processes in the same way that living beings are. This trend of research has suffered strong oppositions especially from ethnographers and anthropologists [4] criticizing in particular the assumption that culture could be divided into individual objects called cultural entities [5], and more generally the lack of empirical validation [6]. Social media offer at last this opportunity for studying social and cultural dynamics *in-vivo*. In this paper, we claim that tracking the transformations of quotations in the blogosphere is a good way to tackle such empirical endeavour. If quotations admittedly cannot catch the complex properties found in every cultural traits, their atomic structure is, by nature, an opportunity that should be seized to put the memetic program, at least partially, into practice.

We are then interested in the diffusion of quotations which builds on the notion of "intertextuality", frequently used in political discourse studies. Intertextuality refers to the fact that fragments of discourse are repeated, re-used and progressively modified in different ways. It is thus possible to track the stability or the progressive distortion of an utterance over time. Following Kristeva [7], we assume that these distortions are not neutral and reflect the way ideas diffuse in different communities.

Diffusion studies have attracted much interest from pioneering sociological approaches [8] to more contemporary studies observing diffusion dynamics in online media. Whatever the nature of the diffusive entity: drugs [9], books recommandations [10], citations [11], or URLs [12], the process at stake in each of these studies assumes that objects are perfectly replicated, their stability being a necessary condition for studying their diffusion. Like the previously mentioned objects, quotations can be tracked because whatever their changes they still refer to the same singular external event. But contrarily to them, quotations can undergo transformations, opening the way to the systematic quantitative analysis of regular patterns underlying these changes.

When it comes to characterizing changes that can affect quotations, it may be useful to follow the biological metaphor provided by memetics. A sequence of genes can be altered by mutations which may affect only one nucleotide or a large sequence of nucleotides. Small-scale mutations encompass *point mutations* (substitute a single nucleotide with another), *deletions* and *insertions*. We claim that such an ontology distinguishing between small-scale and large-scale changes is fruitful for addressing quotation transformation dynamics. We then introduce a different typology than the one described by Simmons et al. in their analysis of the same dataset. In [2] authors discriminate between "reframing" and "alteration" events: if a phrase is transformed into a superstring or substring, then reframing takes place, otherwise one should talk about alteration. The relation of inclusion between two quotations then defines the type of transformation. We prefer to use a different typology directly inspired from the biological mutation process. We will then simply consider on the one hand *micro-mutations* affecting only one word whatever the type of transformation (*i.e.* a word can be added, deleted or even replaced by another one) and on the other hand *macro-mutations* affecting more deeply the composition of phrases. Intuitively small-scale and large-scale mutation events stand for different underlying cognitive processes. Micro-mutations are small changes in the original quotation which can be introduced voluntarily or simply by error with no special intention to alter the original meaning of the replicated quote.

In the other case, macro-mutations are more probably due to voluntary changes by bloggers or journalists that only want to stress the attention of readers toward a subpart of the original quoted text.

Our goal will then be to describe how micro-scale and macro-scale mutations progressively transform quotations during their diffusion. The first part of the article will be devoted to a very short description of our empirical dataset. An original algorithm for detecting coherent families of quotations is then introduced. Based on these families, we will then introduce stability and diversity indexes which help us describing the transformation process at different levels (words, quotes and families). In the last part we will empirically measure mutation rates according to different properties of the quotations and will propose a morphogenesis model for building realistic families of quotations.

II. Dataset description

We analyze the MemeTracker corpus which is made of quotations automatically extracted from 90 millions news and blog articles collected over the final three months of the 2008 U.S. Presidential Election and the following three months [1]. More precisely, we downloaded the MemeTracker dataset from the publicly available website *memetracker.org*, that consists of $310\,457$ distinct quotations collected from news and blog articles from August 2008 till the end of January 2009. Each quotation had to be mentioned at least 5 times in order to be included in the corpus. As we will primarily focus on characterizing how quotes are being transformed, the effects stemming from the underlying social network are out of the scope of this article. We then decided to neglect all the hyperlinks between articles and concentrate only on the textual data, i.e. the quotes themselves, and their number of mentions.

III. Building quotation families

In order to analyze the MemeTracker corpus, it is necessary to identify families of quotations, which means grouping together the different quotations in relation with a same "seed" quotation (i.e. an original quoted text that can be subsequently re-used, duplicated or modified). This analysis is done in three steps: (i) all the quotations are linguistically analyzed and normalized (by lemmatizing the quotes and removing stop words); (ii) similarity between every pair of quotations is calculated and the quotes whose similarity is above a given threshold are linked so as to obtain a graph of quotations; (iii) a clustering algorithm is applied to detect communities (i.e. cohesive subgraphs in the graph) that will correspond to our families. We detail this process in the following subsection, followed by an evaluation of our results and a discussion.

A. A hybrid linguistic and structural approach

While the clustering method of Leskovec and his colleagues [1] builds on structural relations between phrases mainly defined according to their potential string inclusion, we tried to design a proximity measure between quoted phrases following more linguistic hypotheses.

A domain of interest regarding our objective is the paraphrase detection task, which is useful for various natural language applications, including information extraction, automatic summarization and machine translation. Paraphrase detection is highly difficult since it theoretically requires both a semantic and a syntactic analysis of sentences to give valuable results. In practice, most approaches are based on the identification of similar words between couples of sentences, which makes it possible to calculate a similarity value (using a similarity measure like cosine) [13]. Various refinements can be explored in order to get more accurate results, like trying to calculate word similarity (using for example a resource like Wordnet for English) or trying to identify relations between words. For example, Qiu et al. [14] use the Charniak parser to get a syntactic analysis of the sentence and try to map predicate-argument patterns (for example, a verb with its arguments) between sentences, which makes the method more precise.

In this study, we preferred to design a simple strategy for building quotation families which features basic text processing techniques and makes use of a refined proximity measure. First we substituted every word with its lemma using the TreeTagger software [15] and eliminated all the stop words. This step is supposed to conserve only the chore semantic part of each quoted text so that our proximity measure only focuses on the most informative part of each phrase. Lemmatization allows to unify into one single class simple variations of the same word like singular/plurals, or verbs at different tenses. Stop words, also called "empty words", are usually considered as noise when comparing the semantic content of two phrases.

We make use of the traditional Levenshtein distance to assess the dissimilarity between two cleaned quotes. But we still need to add some sophistication to take into account word frequency in our measure, considering that rare words are more informative than frequent ones. We then computed the tf*idf score [16] for every word w of quotation q defined as tf*idf(w,q,Q) = tf$(w,q) *$ idf(w,Q), where tf(w,q) is the word w frequency in the quotation q and idf$(w,Q) = \log|Q|/\{q \in Q : w \in q\}$ where $|Q|$ is the cardinality of the set Q of all the quotations. The first term gives more weight to frequent words (in the quotation) and the second one adjusts this value by penalizing words that are too frequent in the dataset since these words are considered to be not discriminative enough.

Then for every couple of quotations q and q' (with $|q| \geq |q'|$) we computed an adjusted Levenshtein distance treating words as tokens and weighting them with their tf*idf scores. Classically, the Levenshtein distance - also called edit-distance - computes the miminum number of additions, deletions or substitutions necessary to transform an ordered sequence of object into another. Our weighted Levenshtein distance $\mathcal{L}$ then allows to compare two quotations, defined as two ordered sequences of words following this formula:

$$\mathcal{L}(q,q') = \frac{\sum_{\text{min edit path}} f(w,q,w',q')(1-\delta(w,w'))}{\sum_{\text{min edit path}} f(w,q,w',q')}$$

where "min edit path" is the minimum edit path found by the algorithm to compute the Levenshtein distance, $\delta(w, w') = 1$ if $w = w'$, 0 otherwise, and

$$f(w, q, w', q') = \begin{cases} \max(\text{tf*idf}(w, q, Q), \text{tf*idf}(w', q', Q)) \\ \quad \text{if } w = w' \text{ or } w \text{ substituted } w' \text{ or vice-versa} \\ \text{tf*idf}(w, q, Q) \\ \quad \text{if } w \text{ was inserted or } w' \text{ was deleted} \end{cases}$$

The rationale behind this method is to use a proximity measure based on sequences of words since word order clearly matters (as opposed to bag-of-word approaches where words are considered independently of their order of appearance). But we also give more weight to more discriminating words with their tf*idf score.

After this pre-processing step, we constructed a similarity network with the set of quotations, in which every quotation is a node. We assign a weighted edge between two quotations q and q' if they have at least two (full) words in common and if their similarity score, calculated as $1 - \mathcal{L}(q, q')$, is greater than 0.35, a value that we empirically found to be an appropriate threshold. The weight of the edge equals $1 - \mathcal{L}(q, q')$.

The final step was to apply an algorithm for community detection in networks in order to identify different quotation families. For this purpose we chose the Infomap algorithm by Rosvall and Bergstrom [17], an information theoretic approach algorithm which uses the probability of flow of random walks on a network as a proxy for information flow in the real system and decomposes the graph into communities by compressing a description of the probability flow. Lancichinetti and Fortunato tested various community detection algorithms [18] and found that Infomap has an excellent performance combined with low computational complexity, which enables to analyze large systems like our dataset.

B. *Result evaluation and comparison*

Clustering methods are widely used for natural language processing applications that require grouping different sets of elements. However, evaluating the output of clustering methods remains challenging since gold standards[1] are rarely available and different partitions of the data may often make sense depending on the task and the context.

As for our experiment, no gold standard was available but it is possible to use the result of the MemeTracker project experiment as a comparison point. We chose to rely on a manual evaluation of a relevant sample of clusters randomly selected from those produced by our method and those produced by the MemeTracker method. We randomly selected 30 of our families, and for each family we also selected every MemeTracker family that had at least one quotation in common with the original family. Then for each family we made two lists: the first one containing all the quotations in the family, and the second one containing all the quotations which belonged to the corresponding selected families of the MemeTracker project. Then we did the opposite with 30 families of the MemeTracker project. The size of the initial clusters used for evaluation varies from 3 to 150 text snippets/sentences.

The list was then assessed by two judges, who were told that every first list represents a subset of closely related quotations and to mark any quotation that they thought should not belong to the family. Then they had to look at the second list and mark if any of the quotations should be added to the family.

Before detailing the result of this evaluation, it is necessary to quickly examine some methodological issues. First, a number of text snippets were not real quotations but titles ("high school musical"), short expressions with no clear meaning out of context ("a little bit") or foreign words ("la vida no vale nada") between brackets. The corresponding clusters were excluded from the evaluation[2]. Second, the identification of the main information expressed in a set of snippets is a difficult task, especially given the variation in length of the different snippets. The instruction given to the evaluators was to first have a look at the whole set of snippets before determining the prominent information, which seems to have worked pretty well. Lastly, the instruction was to tag as equivalent snippets that were reporting the same main information even if some secondary information was missing. It was possible to tag a snippet as uncertain.

Despite the minimal set of instructions given to our evaluators, we obtained interesting and reliable results. We compared the evaluation produced by two annotators and obtained a high inter-annotator agreement (Cohen's kappa is 0.69, which is surprisingly high given the relative subjectivity of the task and the scarcity of instructions provided to the evaluators).

We obtained the following results for precision and relative recall (the recall is relative since we performed an evaluation based on a comparison between two methods and not with respect to a gold standard), where precision is calculated as the fraction of quotations considered relevant in the first list, recall as the number of relevant quotations in the first list over the same number plus the number of relevant quotations found in the second list.

Clustering Method	Precision	Relative Recall	F-measure
Ours	.58	.90	.70
MemeTracker	.47	.78	.58

Our method outperforms Leskovec et al.'s approach both in precision and relative recall. This increase in precision is probably due to the linguistic preprocessing step that makes the whole process more precise (our analysis is focused on content words that are themselves weighted according to their discriminative power). The increase in relative recall is probably due to the fact that the MemeTracker clusters contain also a lot of snippets that are very small fractions of the seed quotation and were thus considered by the judges not to carry enough information to be unambiguously part of the family.

[1] In natural language processing, a gold standard is a set of manually annotated data. Most of the time, the data is annotated by several annotators to ensure a reliable result based on a high inter-annotator agreement.

[2] Note that our families will be filtered in a second phase in order to delete these kinds of pathologic families.

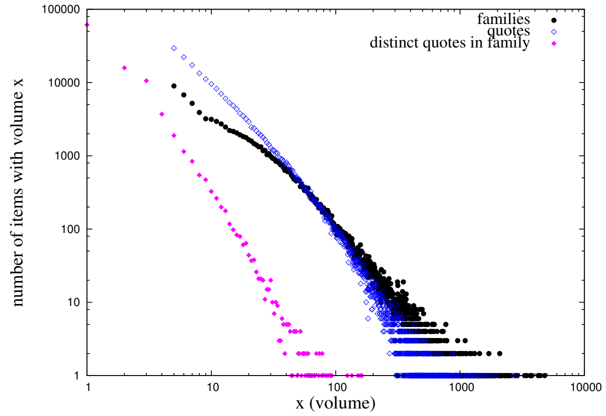


Fig. 1. Distribution of family sizes defined as the total number of mentions of the quotations in the family (black dots), distribution of the number of distinct quotations in a family (pink full diamonds) and distribution of the number of mentions per quotation (blue diamonds).

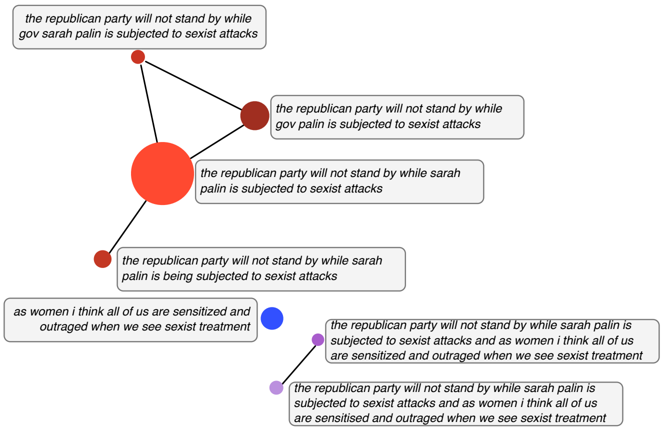


Fig. 2. Empirical family detected by our algorithm and its associated edit-distance graph. Quotes are connected when their edit distance is at most 1. Sub-families (colored in red, blue and purple in this example) are made of connected components of the edit graph. Node sizes scale with the observed total number of mentions of the corresponding quote.

We also measured the statistical significance between the two F-measure values through the SIGF V2 test [19], which implements an assumption-free randomization framework. It allows to assess whether the difference in performance between two sets of predictions is significant. We found a p-value of 0.049, which means that the F-scores are different with a 5% significance level.

C. Family filtering

Since the dataset contains many quotations that are either not in English either too short to convey a real unit of meaning, we first decided to filter it by considering only quotations containing at least 5 words in English[3].

D. Family description

We plotted the distribution of family sizes defined as the total number of quotations mentioned in a family, the distribution of the number of distinct versions found in each family and the distribution of the number of mentions per quotation (Fig. 1). The shapes of all the three distributions can be approximated as a power-law and are comparable to distributions found in [1] although the families were defined differently.

IV. Multi-level transformation analysis

Before going into further details, let's first have a look at an actual family identified as a family gathering 7 different quotes from the MemeTracker dataset. On Sep 3, 2008, Carly Fiorina, former boss of computer-making giant Hewlett-Packard, told a press conference: "*The Republican party will not stand by while Sarah Palin is subjected to sexist attacks ... and as women, I think all of us are sensitized and outraged when we see sexist treatment*". The quote is genuinely replicated 16 times in the dataset but it can also be found in 6 alternative forms, the most frequent of them being mentioned 56 times in a much shorter form than the original one: "*the republican party will not stand by while sarah palin is subjected to sexist attack*".

Observing the diversity of quotes within the family, it appears that 3 main sub-families clearly emerge gathering either the first, the second or both parts of the original quote. Within each sub-family, we also observe small variations between different versions of the quotations due to different spellings of the same word (*sensitised/sensitized*), words trimmed or added (*sarah palin/gov palin/gov sarah palin*, *is being subjected/is subjected*).

As we wish to take into account the diversity of possible changes, we will distinguish in the following between *small-scale transformations* observed when a single word is changed, added or removed, and *large-scale transformations* occurring when the quote is significantly trimmed into a smaller version.

A. Definitions

Before providing a more formal definition of these transformations, we first define the edit-distance graph as the network connecting quotes whose edit-distance is no larger than 1. Edit distance is defined as a Levenshtein distance between two quotes considering words as single characters. The edit distance between two quotes is then the minimum number of edits needed to transform one quote into the other, with the allowable edit operations being insertion, deletion, or substitution of a single term. The maximum number of allowed edits is fixed to one[4]. Please note that in this case and in all the following analysis no preliminary linguistic processing is applied to quotes before computing their edit-distance.

[3] For this purpose we used the *guess-language* software available at http://code.google.com/p/guess-language/.

[4] We have chosen to consider only the edit-distance graph connecting quotes at the smallest possible edit-distance *i.e.* 1. However, we have checked that our results are qualitatively unchanged when considering larger transformations (edit distance at most 2 or 3).

Contrarily to the strategy we adopted for detecting families, we naturally do not wish to alter quotations at this step as we are interested in identifying every possible transformation.

For every family, we then build the edit-distance graph G connecting its quotations, and extract its connected components. These connected components define the sub-families, *i.e.* sets of quotes which can only be differentiated by small-scale changes. Applied to our former example, edit distance graph indeed allows us to exhibit three different sub-families including various micro-level variants (see Fig. 2). The edit distance graph is not only useful to define sub-families since it will also be used to identify - given a target quote q - which quotes are found in its immediate neighbourhood $\mathcal{V}_q$[5] (i.e. quotes that are directly competing for the attention of bloggers or journalists). We then introduce three different measures that will help us to assess the transformation dynamics at different levels.

a) Term level: we are first interested in the relative stability of terms[6] found in a quotation. Given a quote q and a term $t \in q$, we define the stability of this term in this quote $s(t|q)$ as the proportion of quotes in $\mathcal{V}_q$ that share the same term[7]. The global stability of a term t is then defined as the weighted average of term stabilities computed for every quote it belongs to, that is to say: $s(t) = \sum_{q,t\in q} w_q s(t|q)$, where w_q stands for the total number of mentions of q.

b) Quote level: at the quote level, we simply define the stability of a quote q as the proportion between the number of mentions of q and the total number of mentions of quotes in its immediate neighborhood: $S(q) = \frac{w_q}{\sum\limits_{q'\in\mathcal{V}_q} w_{q'}}$. At this stage, we voluntarily focus on micro changes, excluding large-scale transformations that may occur when copying only a subpart of a quotation for example.

c) Family level: we also wish to appraise how much a family or a sub-family composition is heterogeneous. We compute the entropy of the distribution of number of mentions for every quote in the family / sub-family: $H_F = -\sum_{q\in F} p_q log(p_q)$ where p_q holds for the proportion of mentions of quotes q in its family / sub-family: $p_q = \frac{w_q}{\sum_{q\in F} w_q}$.

B. Term level

The observation of the list of the most unstable words exhibits some well known patterns, the first one being the slight orthographic variations that exist for certain words in English. We thus observe the following alternations: "defence/defense", "programme/program" and, among many others, "behaviour/behavior". Other variations include words with a dash ("cease-fire/ceasefire"), abbreviations ("gov/governor") and foreign words ("al-qaeda/al-qaida"). Lastly, slang words

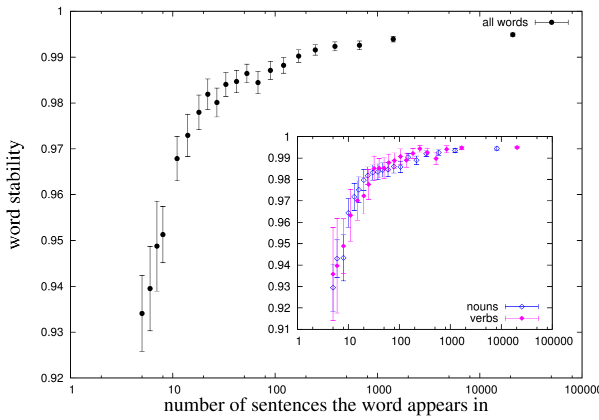


Fig. 3. Stability in function of term frequency. The figure is obtained by creating 20 equally populated quantiles and averaging the stability values corresponding to each quantile. Error bars stand for confidence interval (5%). *Inset*: stability of verbs and nouns only.

are frequently omitted, which makes them more subject to variation than ordinary words (for example, "fuck" is in the top 20 most instable words in the corpus; it can be either suppressed or replaced by a simple "f" in indirect quotes).

Besides these qualitative observations, more systematic patterns emerged when analyzing term stability according to different properties. We also wish to describe which features can provide quotes or terms higher-fidelity replication rates. Put differently, we are asking which properties at the term or quote level may systematically account for a higher or lower mutation rates? Formally, the stability of any feature attached to a term/quote is defined as the weighted average of every terms/quotes sharing this property[8].

Figure 3 shows that word frequency significantly affects term stability. More frequent terms are significantly more likely to be stable. More precisely terms with more than 100 occurrences in our quote corpus have a stability higher than 99%, this value reaching up to 99.5% for the most frequent terms. The rarer the term the more dramatically its stability falls. We checked that this pattern is still present after removing stop-words from our set of terms. The same pattern can be observed even when selecting only certain grammatical types of terms, suggesting that frequency of terms plays a central role when it comes to memory issues or when one has to decide whether a given term should undergo a change. This observation may seem counter-intuitive if one considers that less frequent terms may convey more specific meaning. Yet rare terms may also be more prone to change as they may be misspelled or simply more difficult to spell precisely because of their scarcity. The same kind of observation has actually been made in studies examining the long-term evolution of language. For example, in [20],

[5] $\mathcal{V}_q$ will define the set of quotes whose edit-distance with q is less than or equal to 1 and that belong to q's family. Note that $q \in \mathcal{V}_q$.

[6] In this study we define terms as simple words.

[7] More formally the stability of t in a given quote q is defined as: $s(t|q) = \frac{\sum_{q'\in\mathcal{V}_q, t\in q'} w_{q'}}{\sum_{q'\in\mathcal{V}_q} w_{q'}}$, where w_q stands for the total number of mentions of q.

[8] For example for assessing the stability of terms with a given total frequency in the corpus n we will compute: $s(f=n) = \frac{\sum_{t,f(t)=n} \sum_{q,t\in q} w_q s(t|q)}{\sum_{t,f(t)=n} \sum_{q,t\in q} w_q}$

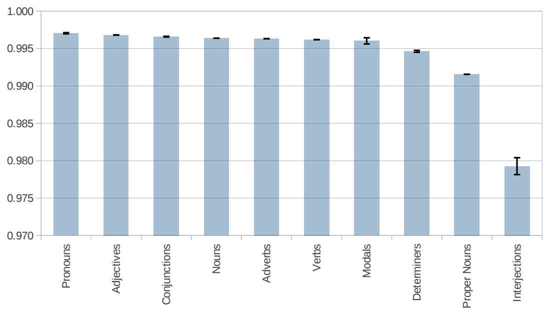


Fig. 4. Stability values for different grammatical types.

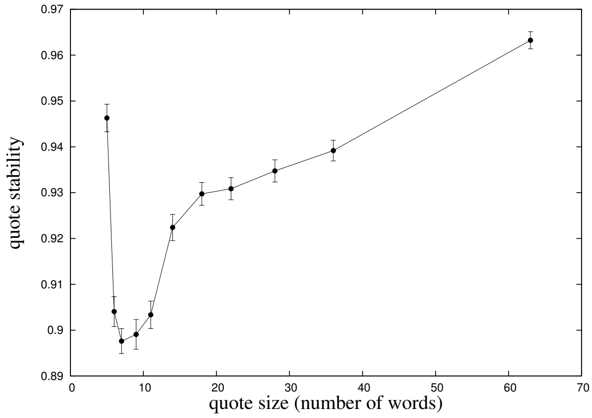


Fig. 5. Quote stability in function of quote length (number of words). The figure is obtained by creating 10 equally populated quantiles and averaging the stability values corresponding to each quantile. Error bars stand for confidence interval (5%).

authors showed that the regularization rate of irregular English verbs was rapidly decreasing with their frequency, indicating that low frequency irregulars verbs are subject to more errors, leading to their "rapid" regularization. Besides, specificity is not synonymous with stability. A recent study analyzing the same dataset used *Wordnet*[9] to rank terms according to their genericity. They showed that the more specific the terms the more likely they are to be replaced, especially by more generic ones [21]. This "natural preference" for more generic terms may also account for the shape of the curve we observe as it is probable that more specific terms occur less frequently than more polysemic ones.

We also computed the average stability of terms according to their grammatical type (see Fig. 4). Results are as expected. Most common grammatical types approximately have the same level of stability. Yet we note that besides interjections which we could expect to feature lower stability, proper noun tend to be more than twice more unstable than average. Indeed proper nouns may produce more mutations as they can be misspelled or undergo more transformations as illustrated in our example in Fig. 2.

C. *Quote level*

We computed quote stability and plotted stability against quote length, i.e the number of words that it contains (see Fig. 5). Quote stability is minimal for a certain length (around 8 words). Smaller quotes are less keen to change, while longer quotes also feature higher stability. On such a digital medium, two processes may be in competition when it comes to editing quotes: *i*) a blogger may read/hear a quote in a newspaper, Twitter, or more broadly catch it from any media and try to replicate it by memory or *ii*) he can simply copy & paste the quote from a digital source. The first copy process is probably more used for quotes that are not too long (a 10-15 words long quote seems already quite difficult to memorize), and is also more keen to introduce variations than a pure copy & paste process (note that variations depend on the length of the quote: very short quotes are easier to memorize and are thus quite stable; longer quotes are more unstable, 8 words long quotes being the most unstable ones). On the other hand, it seems plausible that longer quotes have greater chances to have been replicated from an existing source (for quotes over 8 words long, copy & paste is probably the preferred option, hence the increase in stability for longer quotes). This competition between low and high-fidelity replication along with the uneven probability to introduce errors according to the size of the quote may explain the particular shape of the curve.

We also investigated whether a quote stability is affected or not by its total frequency. In Figure 6 we plotted the weighted average stability of quotes according to their frequency and observe that the curve increases logarithmically. Two explanations may account for the higher stability of high frequency quotes. People trying to replicate them may produce less errors simply because they are inherently better replicators (this property also explaining their popularity). Another cause of their stability may simply be that they are being copied - because of their spread - significantly more often than their alternative forms, increasing in the long-run the disparity between their frequency and alternative forms which are less and less likely to get copied. The two processes may also be taking place at the same time: popular quotes tend to be naturally copied more frequently, while their number may decrease the chances to introduce mistakes in the copying process as it would seem more unlikely for someone to alter a quote she/he has already read several times.

D. *Family / sub-family level*

To better understand how families are composed and more precisely their inner diversity we also measured their entropy. We both measured the entropy of sub-families and of families as a whole. We recall that families group togther every quote related to an original quote whatever the scale of transformations it may have undergone, while sub-families gather quotes from the family that can be connected through micro-

[9]Wordnet is a lexical database of English featuring synonymous relations between words (http://wordnet.princeton.edu/).

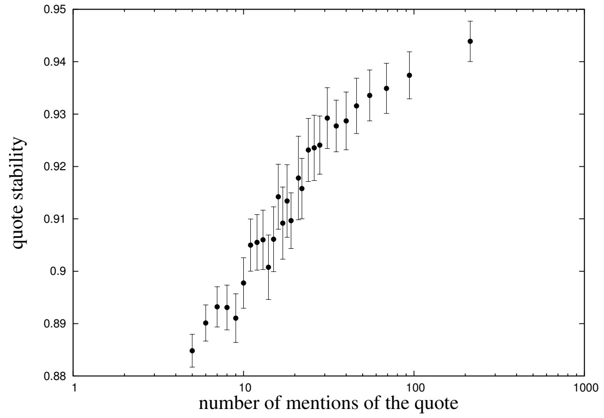


Fig. 6. Quote stability in function of their number of mentions. The figure is obtained by creating 30 equally populated quantiles and averaging the stability values corresponding to each quantile. Error bars stand for confidence interval (5%).

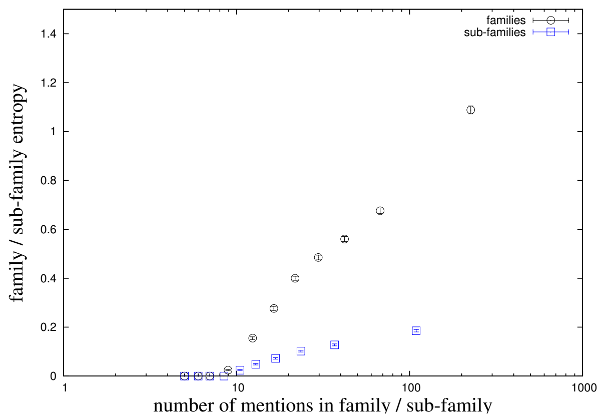


Fig. 7. Family and sub-family entropy in function of the total number of mentions of their quotes. The figure is obtained by creating 10 equally populated quantiles and averaging the entropy values corresponding to each quantile. Error bars stand for confidence interval (5%).

level changes. The Shannon entropy, which was originally applied to letters [22], is classically used as a diversity index. Applied to quotes, the entropy measures the diversity of quotes composing a family or a sub-family. Entropy at the sub-family and family levels exhibits very different patterns. Figures 7 shows how entropy correlates with the family / sub-family size. We observe that the value of the entropy for the families with a certain number of mentions is always much higher than the value of the entropy of the sub-families of the dataset gathering the same number of mentions. Put differently sub-families exhibit less diversity than families, suggesting that - at the micro-level - the competition among different versions of the same phrase eventually leads to a situation in which there is one version that is predominant regarding the number of mentions with respect to other versions, whereas - at the macro-level - there is more heterogeneity due to the coexistence of different relatively independent sub-families.

V. Modeling Quotation Family Generation

We now propose a model of quotation family morphogenesis that accounts for their composition in terms of sub-family size distribution, and regarding their diversity at both levels. To design a realistic model, we still need to precisely define how quotations are being changed when the family is growing. In this section we will then have a closer look at the temporal evolution of families, examining how many and which type of mutations are introduced during the process.

A. *Mutation rates*

We have shown that two types of mutations could occur when copying a quote. When the quotation is not perfectly replicated, one may observe a macro-level mutation - only a subpart of the original quotation is selected - or a micro-level mutation - small changes affecting only one word in the quotation. In the first case, we will consider that a new sub-family is produced, in the second case that the sub-family the original quote belongs to is simply enriched by a new version. We make the assumption that the chances that a mutation produces a version that had been already published before is negligible.

As we assume that every new version is triggered by a mutation event, we can assess mutation rates *a posteriori* at both levels by comparing the number of different versions in a given subset and the total number of mentions these different versions received. More precisely, the average mutation rate can be computed at both micro / macro level as the ratio between the number of micro / macro-mutation events (number of versions in the sub-family minus one / number of sub-families in the family minus one), and the total number of copying steps (mentions in the sub-family / family minus one).

But accessing average mutation rates is not enough to realistically reconstruct family and sub-family morphogenesis: we also need to take into account the relevant properties affecting mutation rates. In the previous section, we showed that quote stability is modified according to their length and their popularity, which suggests that mutation rates could strongly differ according to those two properties. Besides, those properties may critically depend on the diffusion dynamics. That is the reason why we should *dynamically* assess the rate at which new versions are being produced in the empirical process according to those different conditions.

We then define the following strategy to dynamically measure mutation rates. Each family is considered as a growing set progressively populating the various sub-families with new quotes. Each time a new quotation is produced, we record whether it is a perfect copy of a previously mentioned quotation, or a new version that had never been observed before. In the latter case we also record whether the original quotation is enriching an existing sub-family or creating a completely new sub-family. We compile those events according to the original

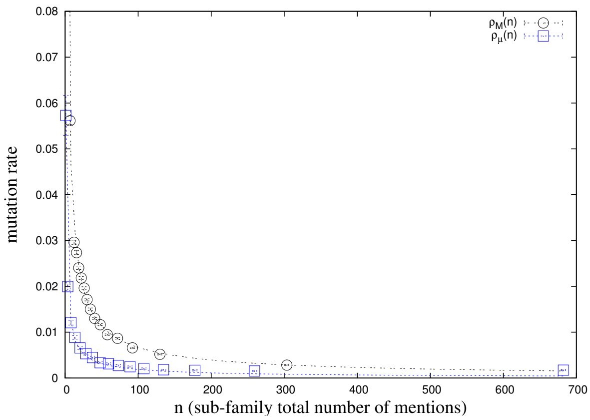


Fig. 8. Micro and Macro mutation rates according to the sub-family total number of mentions, along with their fitted model. $\rho_M(n) = 0.225n^{-0.763}$ and $\rho_\mu(n) = 0.057n^{-0.739}$. The figure is obtained by creating 15 equally populated quantiles and averaging the mutation rate values corresponding to each quantile. Error bars stand for confidence interval (5%).

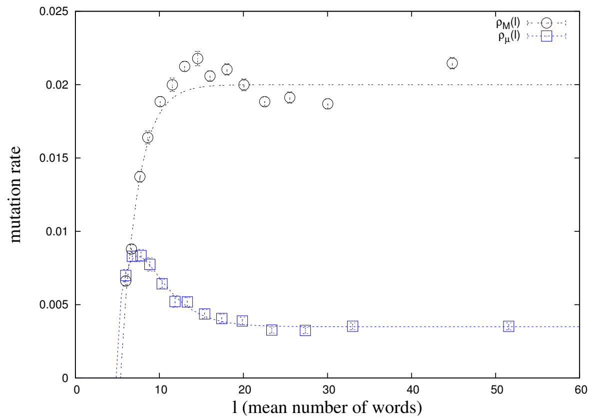


Fig. 9. Micro and Macro mutation rates according to quotation lengths, along with their fitted model. $\rho_M(l) = 0.020 - 0.292\exp(-0.499l)$ and $\rho_\mu(l) = 0.004 + 0.046(l-5)\exp(-0.423l)$. The figure is obtained by creating 15 equally populated quantiles and averaging the mutation rate values corresponding to each quantile. Error bars stand for confidence interval (5%).

state of the family and sub-family[10], *i.e.* we enumerate micro-changes, macro-changes and perfect copying events according to the average sub-family size and average quotation length. From there on, it is straightforward to define the micro / macro-mutation rate according to a given average length or a given total number of mentions as the proportion of replication events producing a micro / macro change. We will call the so computed micro and macro mutation rates ρ_μ and ρ_M respectively.

From figure 6 we can suspect that the number of mentions is crucial for determining the precise mutation rate of a quotation. Therefore we plotted (Fig. 8) the micro (and respectively macro) mutation rates according to the total number of mentions observed in the sub-family. As expected we find that both mutation rates decrease with the number of mentions (see figure caption for further details about the fitting functions used). This behaviour confirms our hypothesis that more popular quotations are less keen to changes. Very popular quotations may be so ubiquitous in the environment that the probability to introduce micro-mutations by error is lowered (many copies can recall the agents how the correct version should be spelled) or that "successful" quotations have such high "fitness" that any further refinements is unnecessary.

Figure 5 showed that quotation stability is sensibly modified according to their length l. We plotted in Fig. 9 the mutation rates according to the average length of the family / sub-family quotations. We observe that the macro-mutation rate is growing with quotation length. While small quotations (less than 5 words) can naturally not undergo any macro mutation given the definition of our family categorization, the macro-mutation rate reaches a threshold for quotations over 20 words. In our model, we use an exponential function to express the dependence of the macro mutation rate with quotation size (see Fig. 9 caption for further details). A plausible explanation is that a quotation can hardly be trimmed before reaching a certain critical length. Above 20 words the quotation is certainly made of different phrases or sentences that individually carry some autonomous meaning even when separated from their original environment.

The correlation between l and the micro mutation rate seems more complex. A peak of the micro mutation rate is reached for mid-size quotations around 8 words. After the maximal value is reached, we observe an exponential decrease until the curve plateau at the minimal mutation rate. As already hypothesized when commenting the shape of quotation stability with length, micro-change dynamics seems to be driven by two processes acting in different directions regarding the number of words. First, it seems clear that mistakes introduced during the copying process are possible only when the quotation is not simply copy & paste. It seems reasonable to postulate that the probability for a quotation to be retrieved from memory rather than copy & paste is exponentially decreasing with quotation length. If the quotation was retrieved from memory then chances are that some mistakes will be introduced. If we refer to classical works in psychology [23], human brain can hold up to a certain number of objects or chunks in memory. This so-called "magic number" below which short term memory is almost perfectly accurate is precisely around 5 for words. This is the reason why we chose to fit the correlation between micro mutation rate and length with a more complex equation made of the product of two probabilities: the probability that the quoted phrase is not replicated by a copy & paste event (which is exponentially decreasing with l) and the probability that an error is introduced by chance (which is assumed to be linear for quotations larger than the magical number 5). This product models the probability that the quotation is replicated with an error. We also add a baseline in the fitting function to account for the constant probability that a blogger or a journalist

[10]We make the hypothesis that a new quotation enriching a pre-existing sub-family was necessarily copied from one member of this sub-family

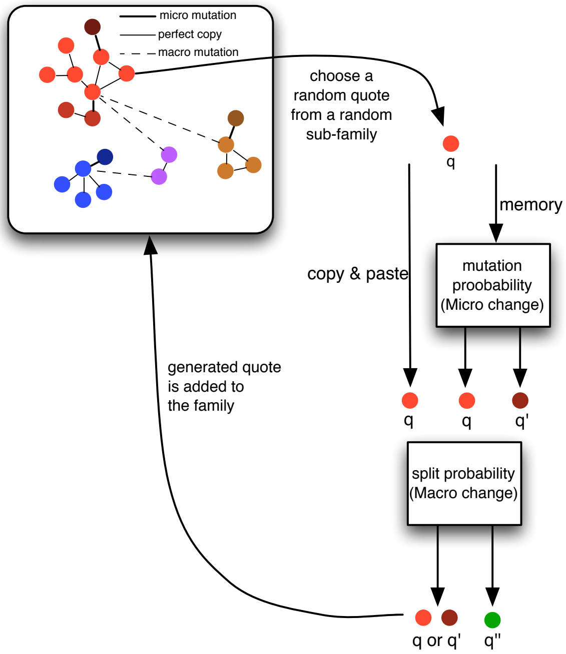


Fig. 10. Schematic representation of the model dynamics.

voluntarily introduces a single change to the quotation (see Fig. 9 caption for further details).

B. Model Design

We now propose a generative process producing families of quotations. Our aim is to find a realistic agent-based process that accounts for the distribution of the size of subfamilies as well as for the shape of the increase of diversity at the family and sub-family level in time. We rely on a classical Polya urn principle like in [2]. We assume that each sub-family is centered around a specific kernel of meaning and is then characterized with its own autonomous dynamical process. This is the reason why we randomly select a sub-family from which a quotation is picked for replication. The quotation may then undergo a mutation according to its length l and its number of mentions n.

More formally, a family F is initialized with a quotation q of a given length l ($F = \{q\}$). This first quotation is also assigned a sub-family f. The simulation then iterates over every time step as follows (see Figure 10 for an illustration):

1) One randomly chooses a sub-family f of F and then a quotation $q \in f$ with probability proportional to the number of mentions of q
2) With a probability given by the combination of the two micro mutations rates $\rho_\mu(l)$ and $\rho_\mu(n)$[11] the quotation undergoes a micro-change, resulting in a new quotation q' that differs from the original one by only one edit (deletion, insertion or substitution). If above this probability the quote is not modified,
3) Then, if long enough, the quotation can also be trimmed into a smaller quotation with a probability given by the the macro mutation rates which depend on the quotation length l and number of mentions n according to the fitted values of $\rho_M(l)$ and $\rho_M(n)$[11]. If so, a new shorter quotation q'' is created.
4) If the quotation did not undergo any mutation on steps 2 and 3, it is perfectly replicated (copy&paste) and a new quotation q is produced.
5) The possibly mutated (q' or q'') or unchanged (q) version of the original quotation is added to the family.

The process is repeated from step 1 until the family is considered complete, *i.e.* when it has received the total number of quotations we observed in the empirical distribution.

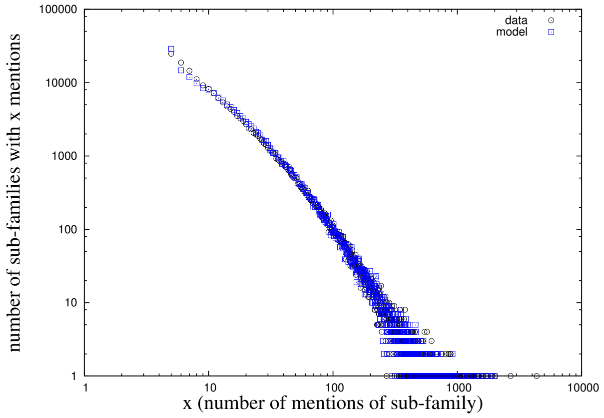


Fig. 11. Comparison of the distribution of sub-family sizes in the empirical data (black dots) and produced by the model (blue diamonds).

C. Model Results

Our simulation almost produced the same number of distinct versions and sub-families than in our original dataset (less than 1% error). We also observe that the proposed model accounts for the size distribution of sub-families and for the diversity of families and sub-families. Figure 11 shows a very good fit between the empirical and the simulated sub-family size distributions, suggesting that our model succeeds in reproducing the

[11] Precisely the global probability to observe a mutation given the total number of mentions and average length is given by $\rho = \frac{\rho(l)\rho(n)}{<\rho>}$. As one can not simply infer which was the original quote from which a new quote was replicated, we computed the stylized behaviour linking mutation rates with l and n considering average lengths and estimated cumulated number of mentions in the sub-family. We are now making the hypothesis that mutation rates can be directly computed based on the quotation length and number of mentions. Lengths being homogeneously distributed, the approximation seems reasonable. Since the distribution of mentions is heterogeneous and given that our process preferentially selects the most cited quote, it is very likely that the number of mentions of a random quote is well approximated by the total number of mentions in the sub-family.

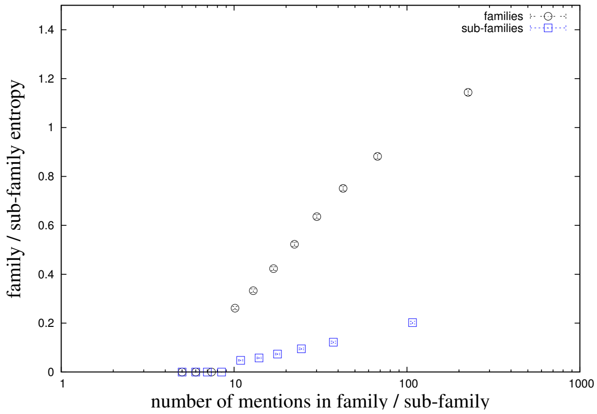


Fig. 12. Family and sub-family entropy produced by the model in function of the total number of mentions of their quotations. The figure is obtained by creating 10 equally populated quantiles. Error bars stand for confidence interval (5%).

observed feature that families are composed of sub-families of more similar quotation versions. Moreover, Figure 12 shows that our model is also able to reproduce the difference that we observed in the values of entropy between families and sub-families. The model creates quite homogeneous sub-families that put together into families account for their larger diversity.

VI. Conclusions

In this paper we introduced a new algorithm for quotation clustering as a first step towards the analysis of quotation family structure and transformations. We showed how these families can be characterized at a meso level by retrieving the corresponding sub-families, i.e. the connected components of a graph of edit distances of at most one edit. This multi-level analysis allowed us to find new interesting results, such as the difference in the entropy level at the two scales, suggesting that the strong competition among very similar quotations leads to a more homogeneous situation with respect to the co-existence of the different sub-families. Moreover, we presented a model that attempts to describe the morphogenesis of these families of quotations and accounts for their composition in terms of sub-family size distribution and for the difference of diversity measured at both levels. The model relies on the analysis of quotation stability, through which we showed that quotations undergo a different number of mutations according to their length and their number of mentions, concluding that quotations that are already very popular have less chances to be transformed. In future work we would like to take into account the underlying social network in order to enrich the analysis of the driving forces determining quotations transformation. Moreover, our model could be significantly more realistic with a finer description of temporal patterns pertaining to quotation diffusion.

Acknowledgment

The authors would like to thank Tommaso Brotto and Zorana Ratkovic for serving as judges for the clustering method evaluation. We also warmly thank the reviewers as well as Sébastien Lerique, Camille Roth, Andreï Mogoutov, Benjamin Fagard, Isabelle Tellier and Jonathan Platkiewicz for fruitful comments. Elisa Omodei's work is supported by a PhD grant from the Région Ile-de-France. This work has also been partially supported by the French National Agency of Research (ANR) through the grant Algopol (ANR-12-CORD-0018).

References

[1] J. Leskovec, L. Backstrom, and J. M. Kleinberg, "Meme-tracking and the Dynamics of the News Cycle," *Proceedings of The Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-09)*, 2009.

[2] M. Simmons and L. Adamic, "Memes Online: Extracted, Subtracted, Injected, and Recollected," *ICWSM 2011*, 2011.

[3] R. Dawkins, *The selfish gene*. Oxford University Press, USA, 2006.

[4] D. Sperber, *La contagion des idées*. Jacob, 1996.

[5] R. Aunger, "Cultural transmission and diffusion," *Encyclopedia of cognitive science*, 2003.

[6] B. Edmonds, "Three challenges for the survival of memetics," *Journal of Memetics-Evolutionary models of information transmission*, vol. 6, no. 2, pp. 45–50, 2002.

[7] J. Kristeva, "Word, dialogue and novel," in *The Kristeva Reader*, T. Moi, Ed. Oxford: Blackwell, 1966.

[8] E. M. Rogers, "New Product Adoption and Diffusion," *The Journal of Consumer Research*, 1976.

[9] J. S. Coleman, E. Katz, and H. Menzel, "The Diffusion of an Innovation Among Physicians," *Sociometry*, 1957.

[10] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *portal.acm.org*, 2007.

[11] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, "Cascading behavior in large blog graphs," *SIAM International Conference on Data Mining (SDM 2007)*, 2007.

[12] E. Adar, L. Zhang, L. A. Adamic, and R. Lukose, "Implicit structure and the dynamics of blogspace," *Workshop on the Weblogging Ecosystem, 13th International World Wide Web Conference*, 2004.

[13] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *Proceedings of the American Association for Artificial Intelligence (AAAI 2006)*, Boston, 2006.

[14] L. Qiu, M.-Y. Kan, and T.-S. Chua, "Paraphrase recognition via dissimilarity significance classification," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia, 2006.

[15] H. Schmid, "Probabilistic part-of-speech tagging using decision trees," in *Proceedings of International Conference on New Methods in Language Processing, Manchester, UK*, 1994.

[16] G. Salton and M. McGill, *Introduction to modern information retrieval*. McGraw-Hill, 1983.

[17] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *PNAS*, 2008.

[18] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E*, 2009.

[19] S. Padó, *User's guide to `sigf`: Significance testing by approximate randomisation*, 2006.

[20] E. Lieberman, J. Michel, J. Jackson, T. Tang, and M. Nowak, "Quantifying the evolutionary dynamics of language," *Nature*, vol. 449, no. 7163, pp. 713–716, 2007.

[21] S. Lerique and C. Roth, "How do brains copy and paste? the semantic drift of quotes in blogspace." *Forthcoming*, 2012.

[22] C. Shannon, "A Mathematical Theory of Communication (pt. 1), 27 Bell Sys. Tech. J. 379 (1948); CE Shannon."

[23] G. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological review*, vol. 101, no. 2, p. 343, 1956.

How do we copy and paste? The semantic drift of quotations in blogspace

Sébastien Lerique, Camille Roth

How do we copy and paste? The semantic drift of quotations in blogspace

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

Author Note

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

Abstract

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.

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

How do we copy and paste? The semantic drift of quotations in blogspace

Introduction

The understanding of knowledge transmission mechanisms has led to a sizable literature in the recent past, spanning over numerous research fields ranging from cultural anthropology to social network analysis and complex systems modeling; from social cognition to data mining. These works are diversely labeled as studies on “opinion dynamics”, “cultural evolution”, or “information diffusion” and, for the most part, investigate phenomena pertaining to both cognitive science and social science, both at the individual and social levels.

Broadly speaking, we may distinguish two main research streams, depending on whether the focus lies on cognitive processes or on social dynamics. A first stream is largely structured around cultural anthropology and essentially addresses cultural similarity, diversity and its evolution. It features several theories mixing social and individual cognition including, to cite a few, the debated “memetic” program initiated by Dawkins (1976) (for which the collection of works by Aunger, 2000, provides a solid overview), the development of evolutionary models of norms (see for instance Ehrlich & Levin, 2005) following the seminal work of Boyd and Richerson (1985); or the “cultural epidemiology” program proposed by Sperber (1996), which links the concept of mental representation to the concept of public representation (the latter being the counterpart of the former outside of the brain, i.e. in all kinds of cultural artifacts: texts, utterances, etc.).

One of the core claims of this literature consists in emphasizing that not all knowledge is equally fit for being reproduced, although the various approaches have a different take on how exactly this notion of fitness should be operationalized. Sperber’s cultural epidemiology classically opposes Dawkins’ memetics by insisting that representations are not being replicated through a high-fidelity copy process, but are being interpreted and produced anew, and are thus greatly subject to change. Cultural epidemiology postulates that this conceptual evolution can be appraised through the notion of “cultural attractor”, seen as the attraction domain of an underlying socio-semantic dynamical system.[1] Despite some recent modeling attempts (e.g. Claidière & Sperber, 2007), the development of quantitative measurements relying on the concept of cultural attractors has remained a relatively hard task and, to our knowledge, this hypothesis has not yet been empirically analyzed in an extensive manner.

Another research stream deals with rather macroscopic studies of knowledge diffusion. Here, one of the focal points is that not all knowledge gets propagated identically along the same routes, within the same communities, at the same pace. The various approaches usually feature a minimalistic description of cognitive processes, strongly reminiscent of biological epidemiology (a single, atomic piece of information may or may not be adopted by each individual). This research program nonetheless exhibits a particularly interesting empirical track record — largely owing to a recent avalanche of observable *in vivo* data which, for a good decade now, have mainly come from online interaction contexts. While these information trails are not records of “physical” inter-individual interactions (in the sense of “real life” interactions), they still constitute a wealth of observations on the dynamics of public – albeit online – representations. Some authors could describe for instance the propagation of cultural artifacts across social networks such as blogspace (Gruhl, Guha, Liben-Nowell, & Tomkins, 2004) or the email network (Liben-Nowell & Kleinberg, 2008), the characteristic times and diffusion cycles both within these social networks and with respect to the topical dynamics of news media (Leskovec, Backstrom, & Kleinberg, 2009b), or the reciprocal influence between the social network topology and the distribution of issues (Cointet & Roth, 2009).

These latter studies are at the interface between data mining, complex systems and quantitative sociology (first and foremost social network analysis) and are relatively remote from cognitive science; for a significant part, they rely rather marginally on specific social

[1]Works such as Atran (2003) argue that this approach is anthropologically better suited than memetics, and some of the main issues in this debate are further detailed by Kuper (2000) and Bloch (2000).

science theories. They nonetheless show us the added value of using these rapidly growing records towards radically improving the empirical understanding of (individual-level) cultural evolution processes.

Stepping back, we thus observe a gap between, on one side, empirical studies of diffusion dynamics in social systems and, on the other side, more theoretical works focused on knowledge transformation processes. Our research lies at the intersection of these two programs, aiming to shed light on micro-level information transformation by leveraging the empirical wealth of (*in vivo*) social diffusion phenomena. More precisely, we hope to describe reformulation processes within a large distributed system such as blogspace; showing how some specific types and features of public representations may be altered by bloggers when they freely reproduce them.

We focus on simple linguistic modifications, thereby connecting our research to the broader psycholinguistic literature. To deal with robust and simple cultural representations, we paid attention to the evolution of quotations. While these verbatim public representations should in theory not suffer any alterations when they are produced anew (as opposed to more elaborate expressions and opinions, not identified as quoted utterances), empirical observation shows that they are occasionally transformed. We will in particular exhibit a non-trivial process by which individual words in quotations are replaced. We will uncover some of the semantic and structural characteristics of these words and the substitutions they undergo. In a way using this type of data is equivalent to a large-scale psycholinguistic experiment and at the same time constitutes a first step towards building empirically realistic models of cultural evolution.

The next section describes our hypotheses along with the relevant state-of-the-art on this psycholinguistic matter. Then, we detail the empirical protocol and the various assumptions that were made in order to deal with the available empirical material. We further describe the significant psycholinguistic biases observed during *in vivo* quotation reformulation as well as their epidemiological setting, followed by a discussion and general guidelines for further work in the final section.

Related work

The practical study of the transformation of public representations has emerged only recently. For one, models involving evolution and representations to study the notion of "cultural attractor" have appeared only a few years ago (see Claidière & Sperber, 2007, and Claidière, Scott-Phillips, & Sperber, 2014, as well as a hybrid empirical-theoretical protocol in MacCallum, Mauch, Burt, & Leroi, 2012). Among the empirical approaches, some of the most relevant studies to date consist in a series of papers investigating *quotation* transformations in a large corpus of US blog posts, initially collected and studied by Leskovec et al. (2009b) and further analyzed by Simmons, Adamic, and Adar (2011) and Omodei, Poibeau, and Cointet (2012). One of the main observations in these works is that even for quotations, a type of public representation that should be among the most stable, it is still possible to witness significant transformations. They essentially examine the effect of some properties of the quotation source (e.g. news outlet vs. blog) or of the surrounding public space (e.g. quotation frequency in the corpus). Some diffusion-transformation models have been proposed, yet the very cognitive features which may determine or, at least, influence these transformations, are overlooked; which may appear to be relatively unsatisfying from a cognitive viewpoint.

At this level, we have to turn to the broader psycholinguistic literature which provides one of the main cognitive foundations for public representation evolution by studying the influence of word features on the ease of recall. This field is well developed and details the impact that classical psycholinguistic variables such as word frequency (see Yonelinas, 2002, for a review), age-of-acquisition (Zevin & Seidenberg, 2002), number of phonemes or number of syllables (see for instance Nickels & Howard, 2004; Rey, Jacobs, Schmidt-Weigand, & Ziegler, 1998), have on this type of task.

Less classical linguistic variables, based on the study of semantic network properties, have recently started to be used, in the context of connectionism and its normative processual models (see for instance Collins & Loftus, 1975). Let us mention four interesting studies on that matter, which demonstrate in a strictly *in vitro* framework and at the vocabulary level that properties computed on a word network are important factors for the cognitive processes and reproduction of those words. First, 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 exists a link between the ease of recall of words and one of their semantic features, namely their authority position (pagerank) in a language-wide semantic network built from external word association data. Austerweil, Abbott, and Griffiths (2012) further develop this idea by showing that random walk on such a semantic network, that is the exact process measured by the pagerank index, gives a parsimonious account of some semantic retrieval effects (namely, related items being retrieved together). A third psycholinguistic study by Chan and Vitevitch (2010) shows, in a picture-naming task, that words are produced faster and with fewer mistakes when they have a lower clustering coefficient in an underlying phonological network (which, again, is defined from external phonological data). D. L. Nelson, Kitto, Galea, McEvoy, and Bruza (2013), finally, show the importance of clustering coefficient in 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 assocation).

On the whole, the current psycholinguistic state-of-the-art seems to hint towards two antagonistic types of results. On one hand, part of the literature tends to show 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 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 repeat a given utterance, is rather of the former or latter sort. We also aim here at shedding some light on this debate, considering oddness as a dimension of the purported fitness of utterances.

Methods

Quotations appeared to be a perfect candidate to propose a first *in vivo* measure of low-level cognitive bias in a reformulation task. 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 could ideally be 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. To keep the analysis palatable, 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. To quantify those 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 dataset and measures we used to assess this cognitive bias.

***In vivo* utterances**

We used a quotation dataset collected by Leskovec et al. (2009b), large enough to lend itself to statistical analysis. This dataset 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).[2] Quotations were then automatically extracted 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.”

Quotations were then gathered in a graph and connected according to their similarity: either because they differ by very few words (in that case, no more than one word) or because they share a certain sequence of words (in that case, at least ten consecutive words). We find for example the following variation of the above quote:

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

A community detection algorithm was applied to that quotation graph to detect aggregates of tightly connected, that is sufficiently similar, groups of quotations (see 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 allegedly contains all variations of a same parent utterance, along with their respective publication URLs and timestamps.

Manual inspection of this dataset 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 all quotes having less than 5 words long or lasting more than 80 days (as well as quotes not written in English). If an entire cluster still lasts more than 80 days

[2]Unfortunately, the original article (Leskovec et al., 2009b) does not provide additional details on the source selection methodology.

after this screening (because of short-lived but unrelated quotes far apart in time), we also exclude it. We eventually keep 45,749 clusters (out of 71,568; i.e. 63.9%), containing a total of 127,778 unique quotes (out of 310,457; i.e. 41.2%) making up about 2.43m occurrences (out of 8.16m, i.e. 29.8%).[3] Even if we lose some real event-related utterances which are present in clusters lasting more than 80 days (such as "the city is tired of me and the organization and I have run our course together"), we check that our approach essentially fulfills its goals by manually coding a random subsample of 100 excluded clusters: a solid 71% appear to be entirely irrelevant to our analysis (everyday language rather than quotations), and all but one of the remaining clusters were of relevance to the protocol set out below.

Word-level measures

Psycholinguistic indices. We first introduce some of the most classical psycholinguistic measures on words.

- **Word frequency**: the frequency at which words appear in our dataset, 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),
- 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)[4] 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.

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

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

The number of synonyms is related to a notion of the word connectivity in a semantic network. To go a bit further in this direction, we appraise the possible role of network-based variables which have received special attention in the recent related literature, following the blooming interest in networks from many disciplines over the last decade.

We relied on the “free association” (FA) norms collected by D. Nelson, McEvoy, and Schreiber (2004) which naturally embed information on the idea association process underlying transformation of quotations. FA norms record the words that come to mind when someone is presented with a given cue. As D. 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.” Following Griffiths et al. (2007), we first build a directed unweighted network based on association norms, where nodes are words and edges are directed from cue to target word whenever the considered target word was produced in response to the considered cue word. This network is of particular interest since it measures the *in-vitro forced-choice* version of a substitution whereas the data we analyze is the *in-vivo spontaneous* version of what we otherwise hypothesize to be the same process.

Three standard network-based measures are to be used on the FA network:

• **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.

• **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 like unavoidable waypoints on association chains connecting one word to another.

- **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 version of the FA network.[5] 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. An important question arises concerning the possible correlations between all the variables we use.

The number of phonemes and the number of syllables naturally exhibit a strong linear correlation (.8). Our analysis showed clearer results with number of phonemes over number of syllables, which is consistent with Nickels and Howard (2004), and we therefore chose to only present results for the former.

Age of acquisition is a key variable which appears as a usual suspect in psycholinguistic studies. Despite it being usually difficult to disentangle from many of the other variables, it is known to have independent effects, which is consistent with what we see on Fig. 1: age of acquisition has a limited correlation to the other variables (absolute value not above .39 if we exclude the number of syllables and the network properties), leading us to keep the variable in the rest of the analysis.

Frequency and number of synonyms both have relatively low levels of correlation to the other variables (excluding again the network properties); we therefore also keep them in the rest of the analysis.

Network centrality properties, on the other hand, are strongly dependent on one another. As mentioned earlier, degree centrality and pagerank have a very strong correlation (.85), and are also redundant with betweenness centrality (with correlation levels at .75 and .68 respectively). Furthermore, the three variables are also strongly related to age of acquisition, which leads us to keep the latter as the sole indicator for centrality. This may trigger a chicken-and-egg issue where a strong centrality may be due,

[5]The Clustering coefficient is formally defined as the ratio between the number of actual versus possible edges between a node's neighbors.

or be the result, of an early age of acquisition; in any case, the age of acquisition seems to partially capture centrality-based network properties.

Conversely, clustering coefficient exhibits low correlation levels with all the variables we kept (maximum absolute value .38), leading us to include it in the rest of the analysis.

The final set of variables we consider, as well as their cross-correlations, can be seen in Fig. 2.[6]

Substitution model

We finally need a substitution detection model, for the utterance data we use presents a challenge: quote-to-quote transformations, and much less substitutions, are not explicitly encoded in the dataset. 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. We thus face an inference problem where, given all quotations and their occurrence timestamps, we should 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. The main issue is deciding whether a later occurrence is a strict copy of an earlier occurrence, or a substitution of an even earlier occurrence, or perhaps even a substitution or copy from quotes appearing outside the dataset, that is from a source external to the data collection perimeter.

Let us give an example: say the quotation "These accusations are false and **absurd**" (q) appears in a blog on January 19, and the slightly different quotation "These

[6]Note that feature values stem from different datasets which do not always encode the same words. Indeed, we have data on frequency for about 22.6k words, on age of acquisition for 30.1k words, on number of phonemes for 123.4k words, number of synonyms 111.2k, and clustering coefficient 5.7k words. Quite often then, not all features are available for all words in our dataset; 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.

accusations are false and **incoherent**" (q') appears in other blogs twice on the 20th and once on the 21st of January. If q was sufficiently prominent when q' first appeared, we can safely assume that the first author of q' on the 20th based himself on q as is shown in Fig. 3a. But what about the second and third occurrences of q', on the 20th and 21st? Should we consider them to be substitutions based on q or accurate reproductions of the previous occurrences of q'? (Options shown in Fig. 3a.)

To settle this question we 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 q 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. Note that these assumptions are admittedly a subset of a much wider set of possibilities, each leading to alternative substitution inferences.[7] It is however not feasible to try them all and, for the sake of simplicity, we decided to go with a sensible set of assumptions, and stick to them without trying alternative options.

Put shortly, such a model defines how many times quote occurrences can be counted as substitutions: in Fig. 3b, occurrences of q' on the 20th are counted as substitutions, whereas the occurrences on the 21st are not. In practice, from the 2.43m initial occurrences spread into 45,749 classes of quotes, with significant redundancy (many quotes are indeed simple duplicates), we manage to mine 6,172 real substitutions obeying to this model. From these substitutions we remove those featuring stop words, minor spelling changes (e.g. center/centre, November/Nov, Senator/Sen), abbreviations, spelled out numbers; this eventually yields 1,051 valid substitutions.

[7] In particular, the criterion of the most frequent quote in the preceding bin may be replaced with the most frequent quote overall, or the oldest quote; time can be sliced into fixed bins as is done here, or kept fine-grained by using sliding bins.

Results

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 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

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.

Results for this measure are gathered in Fig. 4. 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. 4). 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:[8]

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

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

[8]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.

the whole pool of words available in the dataset for that feature.[9] 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:[10]

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

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).

Results are gathered in Fig. 5. 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 ($\mathcal{H}_0$).

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

[10]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 $\mathcal{H}_0$.

Word frequency, with ν_ϕ always significantly above ν_ϕ^0, 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_ϕ^0 values, which is consistent with the literature on recall: words learned earlier and words with lower clustering coefficient are easier to produce than average (D. L. Nelson et al., 2013; Zevin & Seidenberg, 2002). 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 dataset. 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 dataset, 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 dataset, 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.

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_{ϕ}^{0}), 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).

Concluding remarks

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).

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, and Nicolas Baumard for useful suggestions and comments.

References

Atran, S. (2003). Théorie cognitive de la culture. *L'Homme*, *166*(2), 107-143.

Aunger, R. (Ed.). (2000). *Darwinizing Culture: The Status of Memetics as a Science.* Oxford: Oxford University Press.

Austerweil, J. L., Abbott, J. T., & Griffiths, T. L. (2012). Human memory search as a random walk in a semantic network. In *Advances in Neural Information Processing Systems* (pp. 3041–3049).

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python.* O'Reilly Media, Incorporated.

Bloch, M. (2000). A well-disposed social anthropologist's problems with memes. In R. Aunger (Ed.), *Darwinizing Culture: The Status of Memetics as a Science* (pp. 189–203). Oxford University Press.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process.* University of Chicago Press.

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cogn Sci*, *34*(4), 685-97.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1642), 20130368.

Claidière, N., & Sperber, D. (2007). The role of cultural attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1), 89–111.

Cointet, J.-P., & Roth, C. (2009). Socio-semantic dynamics in a blog network. In *Proc. IEEE 4th Intl. Conf. Social Computing* (pp. 114–121).

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Dawkins, R. (1976). Memes: The New Replicator. In *The Selfish Gene* (pp. 189–201). Oxford University Press.

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.

Ehrlich, P. R., & Levin, S. A. (2005). The evolution of norms. *PLoS Biol.*, *3*(6), e194.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*, 35–41.

Gregg, V. (1976). Word frequency, recognition and recall.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: predicting fluency with PageRank. *Psychol. Sci.*, *18*(12), 1069-76.

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information Diffusion Through Blogspace. In *Proceedings of the 13th International World Wide Web Conference (WWW'04)* (pp. 491–501).

Kuper, A. (2000). If memes are the answer, what is the question? In R. Aunger (Ed.), *Darwinizing Culture: The Status of Memetics as a Science* (pp. 180–193). Oxford University Press.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*.

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009a). *MemeTracker: tracking news phrase over the web.* `http://memetracker.org/`. (Retrieved on August 19, 2012)

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009b). Meme-tracking and the dynamics of the news cycle. In *Proc. ACM SIGKDD'09 15th Intl. Conf. on Knowledge Discovery and Data Mining* (pp. 497–506).

Liben-Nowell, D., & Kleinberg, J. (2008). Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences*, *105*(12), 4633-4638.

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

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.

Nelson, D., McEvoy, C., & Schreiber, T. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods*, *36*(3), 402–407.

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & cognition*, *41*(6), 797–819.

Nickels, L., & Howard, D. (2004). 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.

Omodei, E., Poibeau, T., & Cointet, J.-P. (2012). Multi-level modeling of quotation families morphogenesis. In *Proc. ASE/IEEE 4th Intl. Conf. on Social Computing "SocialCom 2012".*

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999, November). *The PageRank Citation Ranking: Bringing Order to the Web.* (Tech. Rep. No. 1999-66).

Rey, A., Jacobs, A., Schmidt-Weigand, F., & Ziegler, J. (1998). A phoneme effect in visual word recognition. *Cognition*, *68*(3), B71–B80.

Simmons, M. P., Adamic, L. A., & Adar, E. (2011). Memes online: Extracted, substracted, injected and recollected. In *Proc. 5th ICWSM - AAAI Intl Conf Weblogs & Social Media* (pp. 353–360).

Sperber, D. (1996). *Explaining culture: A naturalistic approach.* Oxford: Blackwell Publishers.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.

Weide, R. (1998). *The CMU Pronouncing Dictionary, release 0.6.* Carnegie Mellon University.

WordNet. (2010). *Princeton University "About WordNet.".* `http://wordnet.princeton.edu`. (Retrieved on August 19, 2012)

Yonelinas, A. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517.

Zevin, J., & Seidenberg, M. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*(1), 1–29.

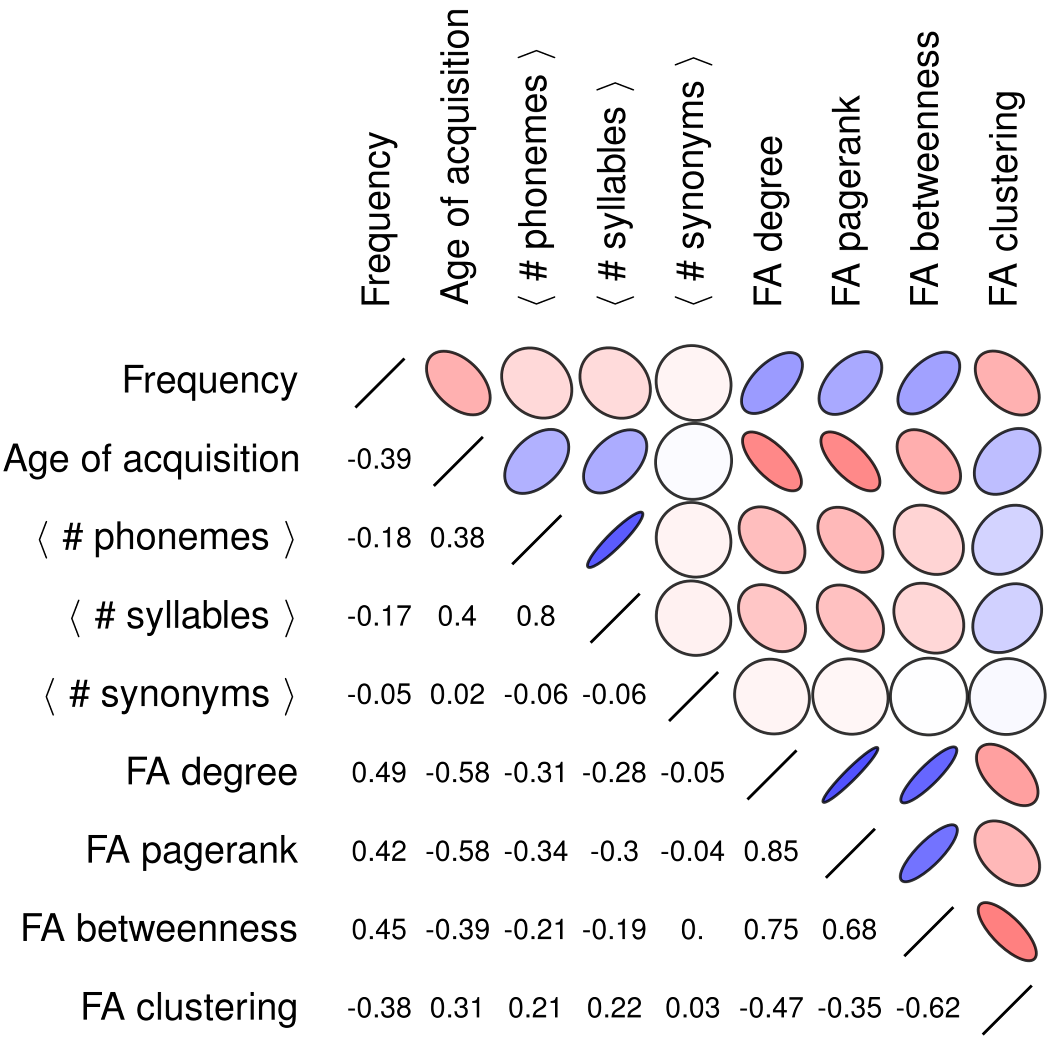


Figure 1. Spearman correlations in the initial set of features

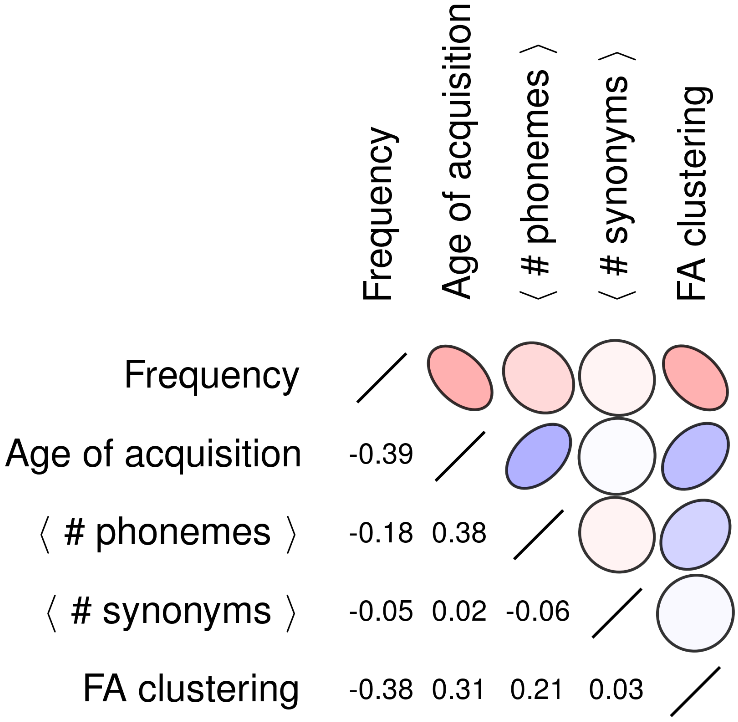
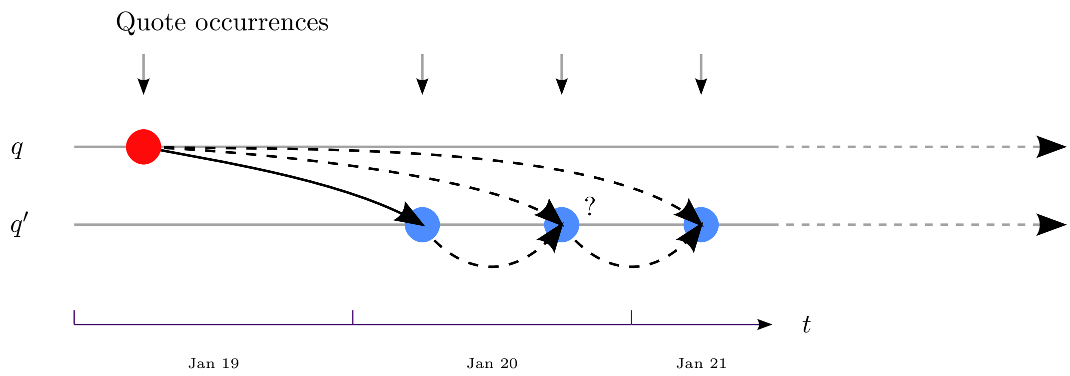
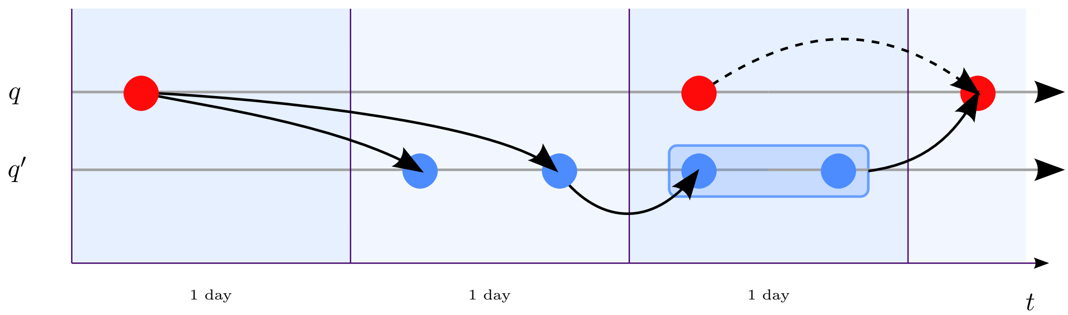


Figure 2. Spearman correlations in the filtered set of features



(a) Possible paths from occurrence to occurrence



(b) Binned quotation family with majority rule

Figure 3. **Temporal binning of quotation families.** q and q' are two versions of a quotation belonging to the same cluster. In the bottom panel (b), q' holds the majority in the 3rd bin and is considered the unique basis for the last occurrence of q (in the 4th bin). This is despite the fact that q also appears in bin 3 alongside q', and despite it having appeared earlier at the very beginning of the quotation family (indeed in the situation shown in Fig. 3b, this seems to be the most likely scenario). Conversely, if q had been the most frequent quote in bin 3, the last occurrence of q in bin 4 would have been considered a faithful copy of the occurrence of q in bin 3.

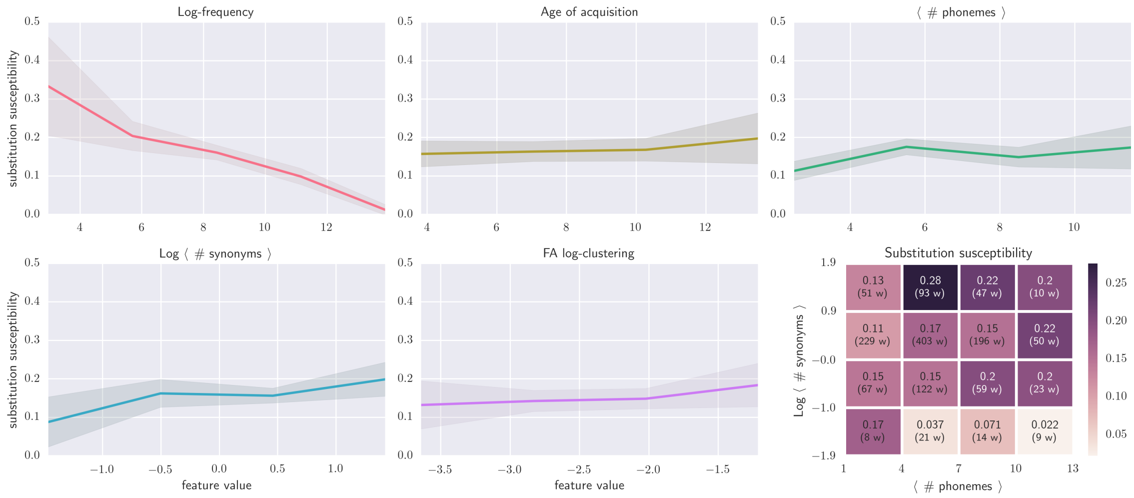


Figure 4. **Substitution susceptibility:** average susceptibility to substitution versus average feature value of a candidate word for substitution, with 95% asymptotic confidence intervals. The heatmap on the lower-right shows the joint effect of Number of synonyms and Number of phonemes on susceptibility, averaged over the respective single-variable ranges, with sample size (word numbers) in parentheses.

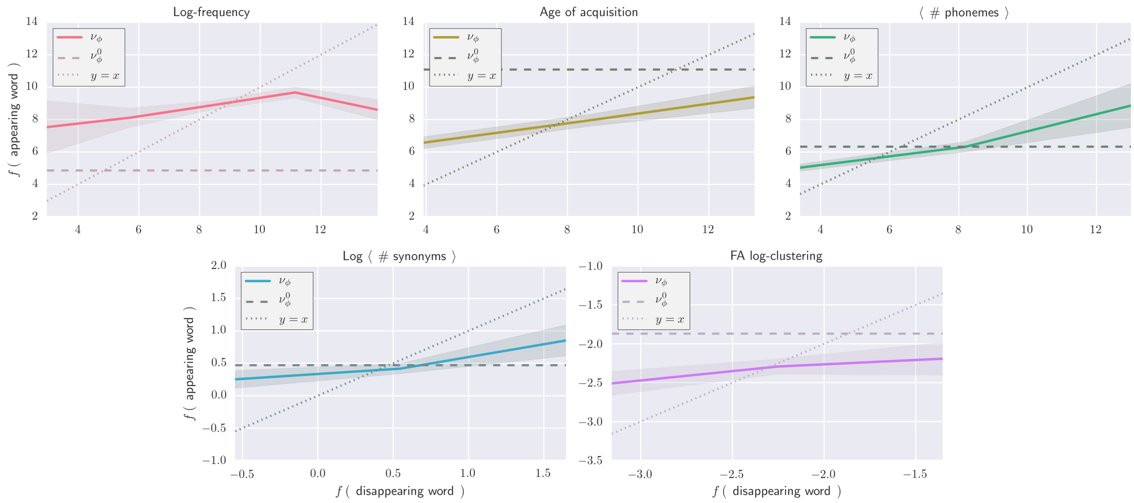


Figure 5. **Feature variation upon substitution:** ν_ϕ, average feature value of the appearing word as a function of the feature value of the disappearing word in a substitution, with 95% asymptotic confidence intervals. The overall position of the curve with respect to the dashed line representing $\mathcal{H}_0$ (constant ν_ϕ^0) indicates the direction of the cognitive bias. The intersection with $y = x$ marks the attractor value.

Pour une étude du contexte d'interprétation

Sébastien Lerique*

Introduction

La mémétique a connu un renouveau récent au travers du programme d'épidémiologie des représentations proposé par Dan Sperber dans le courant des années quatre-vingt-dix (Sperber, 1996). Si son ambition initiale a été relativement délaissée, ses principes restent utilisés aujourd'hui dans l'étude des mèmes internet sur les réseaux sociaux numériques et dans le courant de recherche épidémiologique initié par Sperber.

Ces deux champs participent à un débat actif (bien que parfois peu explicite) à propos des approches permettant de lier les sciences humaines et la psychologie au sens large, et en particulier les sciences cognitives. La position qu'elles ont sur cette question est clairement arrêtée et partagée bien au-delà de leurs communautés respectives. Je voudrais montrer en quoi cette discussion, et la position portée par la mémétique, sont pertinentes au moment d'envisager une utilisation de cette théorie en linguistique.

Je commence par rappeler les principes de l'épidémiologie des représentations, qui s'est constituée comme une critique de la mémétique et connaît actuellement un succès réel. Puis j'examine en détail la façon dont Tim Ingold, principal critique explicite de cette approche aujourd'hui, discute ces deux théories depuis l'anthropologie sociale.

Je tente ensuite de montrer comment cette discussion apparaît concrètement dans une étude de cas inspirée de l'épidémiologie des représentations, et comment les questions de fond du débat sont pertinentes pour (et connues de) la linguistique au moment de s'inspirer de la mémétique.

Je conclus en défendant une utilisation critique de cette théorie, c'est-à-dire en explicitant que les situations qu'elle peut traiter ne peuvent que difficilement impliquer la sémantique, et en attirant l'attention sur les développements récents des sciences cognitives énactives qui pourraient permettre de dépasser ces limites de fond.

1 Les principes de la mémétique

1.1 L'épidémiologie des représentations

Dans le milieu des années 1990 Dan Sperber propose une nouvelle synthèse entre sciences sociales et psychologie, inspirée par les développements contemporains des sciences cognitives (Sperber, 1996). Sa théorie, l'épidémiologie des représentations, reconnaît l'analogie de principe entre évolution biologique et évolution culturelle, et conçoit également la culture comme un ensemble d'atomes de culture qu'il appelle représentations. Mais à l'opposé de la mémétique, les représentations subissent des mutations de façon quasi systématique lors de leur passage par les appareils cognitifs successifs des individus qui

*Centre d'Analyse et de Mathématique Sociales (UMR 8557, EHESS, Paris), et Centre Marc Bloch (Berlin). Mél : sebastien.lerique@normalesup.org.

les perçoivent et les produisent de nouveau. En effet une représentation peut être *mentale* – lorsqu'elle existe dans l'esprit de quelqu'un – ou *publique* – lorsque cette personne la produit sous une forme physiquement perceptible par d'autres individus. Par exemple : j'ai une mélodie en tête (représentation mentale), que je siffle (représentation publique) ; une autre personne la perçoit et en fait sa propre représentation mentale. La nouvelle représentation mentale dans l'esprit de cette personne est dans la plupart des cas sensiblement différente de la mienne (Sperber, 1996 : 25-26).

La notion de représentation mentale que Sperber utilise est celle des sciences cognitives classiques, ce qui lui permet de baser sa théorie sur ce courant : la façon dont chaque individu perçoit une représentation dépend des modules cognitifs (ceux introduits par Fodor, 1983) mis en jeu. En effet chaque type de représentation correspond à un certain nombre de modules chez l'individu qui la perçoit, ce qui définit la partie du système cognitif utilisée pour la perception. Ces modules ont évolué par sélection naturelle il y a plusieurs centaines de milliers d'années et sont donc adaptés à l'environnement pour lequel ils ont évolué, à savoir celui de chasseurs-cueilleurs de cette époque. Le décalage entre les tâches pour lesquelles les modules ont évolués et celles pour lesquels ils sont utilisés aujourd'hui est la principale source de transformations des représentations au cours de leur diffusion (Sperber, 1996 : 138-139).

Sperber combine donc l'idée d'une modularité de l'esprit – fait de modules évolués dans un environnement bien différent de celui d'aujourd'hui – à la description de la culture à partir de parties élémentaires – les représentations – perçues, transformées, et reproduites de façon continue. Sur cette base, Sperber fait l'hypothèse que l'accumulation des transformations opérées par les modules cognitifs fait converger des sous-ensembles de représentations d'une société vers des attracteurs culturels. L'étude de l'existence de ces attracteurs ainsi que leur caractérisation permet alors de rendre compte de l'évolution ou du maintien d'une culture (le maintien s'opérant lorsque des représentations ont convergé vers des formes stables à force de transformations successives) ; c'est une des questions centrales de l'épidémiologie des représentations (Sperber, 1996 : 106 ; Miton et al., 2015) .

Comme tout principe théorique, cette proposition permet de rendre des phénomènes plus intelligibles au prix de certaines simplifications. Ce sont certaines de ces simplifications, présentes également dans la mémétique (même réduite aux mèmes internet), que je voudrais ici discuter.

1.2 Des fondements discutés

Dans une série d'articles publiés à partir de la fin des années 1990 (Ingold, 1997, 1999, 2004, 2007), Tim Ingold identifie les fondamentaux qui sous-tendent les modèles mémétique et épidémiologique de la société et développe une critique exhaustive qui éclaire utilement les limites de fond de cette famille d'approches. L'alternative qu'il propose s'appuie sur le changement de paradigme initié au tournant du siècle en sciences cognitives et qui, en permettant une plus profonde intégration entre les sciences cognitives et les sciences sociales, laisse entrevoir la résolution de certaines des dichotomies qui divisent ces deux approches de l'humain.

Je voudrais ici rapporter les points principaux de cette discussion et montrer sa similarité avec des problématiques bien connues de la linguistique. L'objectif est donc d'identifier et de placer les positions et les dichotomies dont on hérite lorsqu'on utilise une approche du type mémétique. Commençons par détailler la discussion elle-même.

1.2.1 La thèse de la complémentarité

D'après Ingold, pour décrire l'ensemble de l'approche mémétique il faut voir que cette combinaison entre sciences cognitives classiques et conception de la culture comme assemblage de représentations, telle que la propose Sperber,[1] est basée sur la synthèse néo-Darwinienne de l'évolution : celle-ci fournit le substrat sur lequel la combinaison des deux disciplines précédentes est pensée.

Cette synthèse entend expliquer l'évolution en considérant les gènes comme unique lieu du changement biologique : dans cette approche un ensemble de gènes permet la spécification directe d'un organisme, et l'objectif général est de décrire l'évolution de cette spécification au cours du temps. Le code génétique est donc le principal objet d'étude pertinent : les organismes individuels se développent suivant leur spécification génétique et, bien que ce processus se fasse en interaction avec l'environnement, les variations de développement qui peuvent apparaître sont circonscrites à la vie des organismes en question ; elles n'impactent pas l'évolution de l'espèce elle-même au-delà de la variation du nombre de descendants qu'ils peuvent avoir, quantité qui est ensuite moyennée sur tous les individus (Fisher and Bennett, 1999). Ainsi, connaître l'évolution de la spécification génétique revient à décrire l'essentiel de l'évolution.[2] C'est dans ce processus que prend place l'évolution des modules cognitifs des sciences cognitives ; et c'est sur la base de cette théorie que l'épidémiologie des représentations peut proposer une évolution *culturelle* parallèle à l'évolution *biologique*.

La synthèse néo-Darwinienne forme donc, en association avec les sciences cognitives classiques et la théorie de la culture comme assemblage de représentations, une proposition théorique cohérente pour rendre compte du vivant aux trois échelles principales : biologique, cognitif, et culturel (ou social). Ingold appelle cette association la « thèse de la complémentarité » en référence à la façon dont le corps, l'esprit, et la culture y sont complémentaires l'un de l'autre, chacun s'arrêtant là où le suivant commence (Ingold, 1999). Il décrit alors comment chacune de ces trois disciplines fonctionne sur une séparation entre fond et forme – ou spécification et réalisation – qui a été perdue de vue, et a de ce fait été abusée.

1.2.2 Niveau biologique

Au niveau biologique, si les gènes sont une partie indispensable du processus de synthèse des protéines, c'est précisément parce que la machinerie cellulaire dont ils font partie fonctionne de cette façon : au lieu de considérer que la cellule sait *lire l'information codée* par les gènes pour synthétiser des protéines, il est plus correct de considérer que c'est la façon dont tous les éléments d'une cellule interagissent qui donne aux gènes le rôle qu'ils ont dans la synthèse des protéines, rôle qui est loin d'être celui d'une spécification univoque (Ingold, 1999).

Bien plus qu'une question de formulation, cette critique est au fondement de la biologie évolutive du développement, ou « évo-dévo », qui montre comment des processus développementaux différents peuvent produire des organismes différents à partir des mêmes gènes (West-Eberhard, 2003) : si la machinerie cellulaire ou l'environnement dans lequel se développe l'organisme changent, l'organisme qui se développe n'est plus le même (pour plus de détails sur ces phénomènes, on peut s'intéresser par exemple à l'évolution des ailes des oiseaux ; Prum and Brush, 2014). Les gènes ne sont donc pas une information en soi, mais une composante utilisée d'une certaine façon dans un processus de développement, donc à laquelle un sens est attribué par l'intermédiaire de ses interactions. Plutôt qu'une spécification qui se réalise dans un environnement à des fluctuations près, on est face à un processus de développement fait d'interactions entre de nombreuses parties qu'il faut considérer dans son ensemble (Ingold, 1999).

Il n'est pas question d'un curseur qu'on déplace entre inné et acquis, ni même d'une nuance consistant à dire que l'inné s'instancie avec des variations individuelles dues à l'environnement. Il s'agit bien de dire que les notions d'inné et d'acquis sont un modèle qui déforme la réalité : ce qui serait l'inné n'a de

[1] Dans les termes d'Ingold, « un corpus de connaissances ou d'information qui peut être transmis à travers les générations indépendamment des situations d'application concrètes dans le monde » (Ingold, 1999).

[2] Cette description paraît souvent évidente, car c'est ce paradigme qui domine aujourd'hui notre sens commun de l'évolution. Nous verrons dans la suite qu'il s'agit d'une simplification abusive de la réalité, qui n'est plus du tout consensuelle dans la biologie contemporaine.

rôle et de sens que dans la mesure où l'organisme et l'environnement l'utilisent de cette façon, et n'a donc pas un statut différent du reste de l'organisme. Notamment, sa description ne suffit pas à rendre compte de l'évolution. Ce qui serait l'acquis est la façon dont l'organisme se développe, et il n'y pas de sens à le comparer à une hypothétique référence (qui serait la façon standard de se développer à partir de l'inné), puisque c'est ce même processus de développement qui définit la lecture et l'utilisation de l'inné.

1.2.3 Niveau cognitif

On retrouve ce principe de séparation entre spécification et réalisation en sciences cognitives classiques et en psychologie évolutionniste, particulièrement visible dans le débat autour du niveau de modularité de l'esprit. D'après ces approches, les modules cognitifs définissent l'architecture d'un esprit vierge d'éducation et d'expériences, et sont le substrat sur lequel prend place l'apprentissage de capacités telles que la marche ou le langage. Ils sont le fruit de l'évolution de l'espèce humaine, et sont donc transmis de façon héréditaire dans le matériel génétique de l'espèce. Nos gènes encodent donc des instincts d'apprentissage qui agissent lors du développement (Cosmides and Tooby, 1997) pour rendre naturel l'apprentissage d'une capacité (la notion d'instinct d'apprentissage, plutôt qu'une innéité directe des capacités, permet d'accommoder en partie la diversité des façons dont chaque individu se développe).

Cette vision, explique Ingold, postule une séparation entre spécification et réalisation analogue à celle du niveau biologique, et s'y ajoute (Ingold, 1997) : on trouve d'un côté les instincts d'apprentissage, c'est-à-dire la spécification indépendante du contexte (qui semble être elle-même tirée directement de la spécification génétique, sans effet de contexte), et de l'autre le processus d'apprentissage culturel, c'est-à-dire la réalisation variable dans un contexte particulier. Par exemple dans le cas du langage : le dispositif d'acquisition du langage encodé en tant que module dans le matériel génétique serait mis en place et prêt à apprendre dès la naissance, et grâce à lui le bébé apprendrait la langue de son environnement (Cosmides and Tooby, 1997).

Comme au niveau biologique, la critique d'Ingold ne porte pas sur l'existence d'un rôle génétique dans le développement, mais sur l'exclusivité de ce rôle. Il est tout à fait entendu qu'un chat, tel que nous connaissons son espèce aujourd'hui, ne pourra pas apprendre à parler comme un humain au cours de sa vie. Mais cela ne veut pas dire pour autant que les humains portent en eux les gènes d'un dispositif d'acquisition du langage qui rendrait compte de l'ensemble des capacités langagières de l'humain. La seule conclusion qu'on peut en tirer est que le processus de développement dans sa totalité (gènes, cellules, organisme, environnement et interactions langagiers) permet le développement du langage, celui-ci n'ayant pas nécessairement d'existence ontologiquement séparée des réalisations concrètes (Ingold, 1999).[3]

1.2.4 Niveau culturel

Enfin, la même dichotomie est présente dans la vision de la culture défendue par la mémétique et l'épidémiologie des représentations. Dans cette conception le biologique, le cognitif, et le culturel sont des couches parallèles les unes aux autres où chacune sert de substrat à la couche supérieure : le culturel est un monde de significations en soi qui se développe et évolue sur le substrat cognitif qu'est l'esprit, lui-même existant sur le substrat biologique qu'est le corps.

[3]Un argument de type « pauvreté du stimulus » (Chomsky, 2005) n'est pas non plus opposable ici : le développement d'un organisme peut être contraint (comme l'est celui du chat par rapport à celui de l'humain, ou celui de l'humain par rapport à l'apprentissage du langage, étant supposé qu'il apprend plus que ce que l'environnement fournit) sans que cette contrainte soit due uniquement à une différence génétique. On ne peut donc pas attribuer la contrainte à une supposée information innée.

On retrouve alors une séparation entre d'un côté une spécification hors contexte d'un contenu défini à la couche culturelle (une représentation ou un mème), et de l'autre la réalisation concrète dans une situation d'interaction impliquant le cognitif (une chanson sifflée, un tweet écrit, sont des réalisations de représentations). De façon analogue au niveau biologique, où le code génétique est séparé conceptuellement de l'organisme qui se développe et dont on abstrait les fluctuations, on sépare conceptuellement la représentation de ses réalisations concrètes dans des contextes d'interaction : on pense ainsi l'évolution culturelle au niveau de la spécification, c'est-à-dire en étudiant des ensembles de représentations qui évoluent parallèlement à l'évolution cognitive et biologique. On parle alors de coévolution gène-culture (Sperber, 1996 : 114).

Pour illustrer la limite de cette approche, Ingold reprend l'exemple de la transmission d'une recette de cuisine discuté par Sperber (1996 : 61) : une recette (par exemple chez Sperber, la sauce Mornay) est une représentation qu'on peut trouver sous forme publique dans un livre de recettes. Une personne devrait donc pouvoir lire cette recette et s'en faire une représentation mentale, « dont il peut se souvenir, qu'il peut oublier, ou transformer, ou qu'il peut aussi suivre – c'est-à-dire en faire un comportement physique » (Sperber, 1996 : 61). Pour Ingold au contraire, la connaissance qu'un cuisinier a n'est pas la mémoire du livre de recettes appris par cœur. En effet il ne suffit pas de savoir lire la recette de la sauce Mornay et en retenir les mots pour la réaliser : il faut pouvoir la mettre en pratique, c'est-à-dire attribuer un sens à chacune de ces instructions. Or savoir interpréter une instruction telle que « faire fondre le beurre et y incorporer la farine » relève d'un savoir pratique de la cuisine, au sens de Bourdieu (1980) : il faut avoir pratiqué et appris par le corps, en général aidé d'un instructeur, de nombreuses actions en cuisine. Il faut avoir appris, par essais et erreurs répétés, à reconnaître quand le beurre va brûler, ou à identifier la quantité de farine à ajouter au bon moment ; il faut une pratique des mouvements pour incorporer cette dernière au beurre. Pour Ingold, la connaissance qu'un cuisinier a est le savoir pratique de toutes ces tâches, qui est ce qui permet de donner un sens aux instructions d'une recette (Ingold, 1997).

Pour clarifier la distinction entre les deux approches, Ingold fait une analogie entre la réalisation d'une recette de cuisine et une randonnée en campagne : les indications de la recette sont comme les indications peintes au sol sur le chemin de randonnée, et le savoir pratique du cuisinier est le chemin lui-même. Chaque indication au sol (instruction de la recette) est stratégiquement placée à un endroit où le chemin n'est pas clair. Mais grâce au chemin (le savoir pratique), il est facile d'avancer jusqu'à l'indication suivante, même si le tracé est sinueux (tant qu'il n'y a pas d'ambiguïté). Effacez le chemin (un cuisinier sans savoir pratique), et les indications n'ont plus aucun sens. C'est donc par l'existence du chemin que la signalisation peut indiquer la route à prendre. La signalisation ne porte aucune information en elle-même : elle est un outil pour la combinaison de savoirs pratiques.

Ingold décrit ainsi la cuisine comme une navigation dans un paysage de tâches apprises au travers de nos interactions et de nos pratiques. Ainsi, défend-il, bien plus que la transmission de représentations, ce que les générations précédentes lèguent aux générations suivantes est une éducation de l'attention : savoir naviguer et s'orienter dans le paysage de tâches consiste à savoir regarder au bon endroit au bon moment. Savoir faire une sauce Mornay n'est pas avoir la connaissance du texte de la recette mais bien avoir le savoir pratique qui permet de discerner à quel moment ajouter chaque ingrédient, savoir réagir si la sauce est en train de brûler, savoir couper le fromage à la bonne taille en fonction de sa dureté. Si tel est le cas, la culture n'est en rien un niveau parallèle à la biologie et au cognitif, mais les trois s'interpénètrent et se développent ensemble (Ingold, 1997, 2004).

La ligne d'Ingold rapportée ici s'appuie sur l'anthropologie du corps de Mauss (2013), l'approche écologique de la psychologie de Gibson (2014), l'écologie de l'esprit de Bateson (1972), et bien sûr sur le savoir pratique de Bourdieu (1980). Mais il prend également racine dans le courant énactiviste des sciences cognitives.[4] Celui-ci, en changeant le paradigme de fond de la cognition, a été extrêmement

[4]On parle plus largement des « quatre E » pour décrire les différents aspects du changement de paradigme : la cognition est vue comme énactive, « embodied » (corporisée), « embedded » (située), et étendue.

productif notamment dans le domaine de la perception (O'Regan and Noë, 2001), et développe depuis peu une approche non-représentationnelle du langage (Cuffari et al., 2014).

Ces deux approches apparaissent au fond comme des théories de la relation entre le cognitif et le social. Elles diffèrent dans leur façon de rendre compte du réel, principalement au niveau des présupposés de simplification : la première théorie fonctionne sur un principe d'abstraction de l'environnement qui permet de définir des contenus hors-contexte, pour ensuite en étudier la structure et les interactions. Comme j'espère avoir expliqué clairement, la deuxième montre comment la première est un cas particulier qu'il n'est pas possible de généraliser : il y a de nombreux cas où la description en termes de représentations n'est pas adéquate et déforme trop la réalité. Je voudrais maintenant montrer comment cette limitation apparaît concrètement dans la réalisation d'expériences autour de la mémétique.

2 L'expérience à l'épreuve de la mémétique

2.1 Une problématique linguistique

La problématique détaillée rappelle des questions bien connues de la linguistique. On retrouve un écho au débat entre représentationnalisme et énactivisme dans la séparation entre sémantique d'un côté et pragmatique de l'autre, ou encore entre signifiant et signifié. Mais c'est surtout la vision du second Wittgenstein (2010) qui apparaît chez Ingold et dans l'approche énactiviste (Cuffari et al., 2014) : le langage et le sens y sont décrits comme tramage et interaction avec l'environnement et les autres organismes qui le constituent.

Cette problématique de la signification a habité la linguistique et l'anthropologie depuis plus d'un siècle, et il n'est pas surprenant que la question réapparaisse dans des expériences concrètes autour de la notion de mème internet. Venant tout droit des sciences dures, la mémétique n'a pas bénéficié de l'intense débat ontologique qui anime les sciences humaines depuis leurs débuts, et a tout simplement ignoré la question plutôt que de la résoudre.

2.2 L'expérience face à ses présupposés

Dans le cadre de ma thèse sur l'épidémiologie des représentations, j'étudie quantitativement la transformation de courtes citations lorsqu'elles sont propagées sur internet de blog en blog et dans des situations d'expérimentation plus contrôlées. Les billets de blogs (de la blogosphère anglophone dans le cas présent) incluent en effet de nombreuses citations reproduites sous forme de discours direct (Leskovec et al., 2009), mais ces reproductions sont également régulièrement transformées par rapport à l'original, au-delà du simple rognage (Simmons et al., 2011) : un mot disparaît, une contraction apparaît, une expression est légèrement déformée (pour une typologie complète des transformations qu'on rencontre voir Lauf et al., 2013). Ces courtes citations sont donc un bon exemple de représentation publique ou de mème qui – malgré la règle implicite de non-modification d'un discours rapporté – est régulièrement transformé au cours de sa propagation sur internet par ce qui semble être des biais cognitifs : un cas d'étude empirique idéal pour l'épidémiologie des représentations. Grâce à un corpus conséquent rendant l'analyse quantitative possible (Leskovec et al., 2009), on peut poser la question des types de transformations observés, de leur origine cognitive, de l'effet de leur accumulation à long terme (par exemple l'apparition d'attracteurs culturels), et de l'interaction de ces transformations avec des processus de diffusion.

Mais le problème de présupposés détaillé dans la première partie est bien une réalité pour ce type d'étude. Au-delà des difficultés techniques qui ne permettent pas d'avoir accès aux contextes réels de lecture et de production d'une phrase, l'aspect énactif de ces représentations devient évident au moment de réaliser l'analyse : le sens est une construction émanant de la situation concrète du lecteur et de sa relation au texte lu. Prenons l'exemple du tweet suivant :

[1] « On est tous le beau et le moche de quelqu'un »[5]

Cet énoncé paraît neutre a priori, et est suffisamment classique et consensuel pour être « aimé »,[6] repris et publié à nouveau régulièrement.[7] Mais comme le montre la conversation qui suit, on ne peut pas connaître le *sens* de l'échange. Une réponse au message initial est d'abord faite sur le ton de l'humour :

[2] « mais être moche c'est quand même la base ahah »[8]

Puis deux échanges plus tard la conversation se termine :

[3] « [mort de rire,] pour certaines filles surtout, je pense »[9]

Même après cinq répliques, on ne sait toujours pas s'il s'agit de sexisme et de rejet ou d'une plaisanterie sans conséquence. Sans plus d'informations sur la relation entre les interlocuteurs, leurs interactions, ou l'histoire commune qu'ils peuvent avoir, cette conversation pourtant publique ne nous permet pas de savoir ce qui a été échangé sur le fond, ni même ce que signifie le message initial pour l'un ou l'autre des participants.

Revenons alors à la question d'origine : la mutation de telles représentations publiques, ou mèmes. À la lumière de l'exemple ci-dessus et de son explication dans la première partie, il devient clair qu'un projet visant à étudier les mutations *sémantiques* sera confronté à une indétermination de fond pour saisir le sens des phrases en question. Étant donné que les informations de contexte nécessaires sont hors de portée (sans même parler de l'appareillage pratique et théorique pour les traiter), il semble indispensable de forcer une détermination du sens en créant des situations artificiellement plus contraintes, c'est-à-dire en essayant de décider de toutes les ramifications du contexte à l'avance. Mais alors l'essence du modèle mémétique, qui consiste à abstraire le contexte dans l'étude de la diffusion et de la mutation, est perdue.

Les alternatives sont donc :

- l'étude de la diffusion sans mutation des représentations,
- l'étude des régularités qui apparaissent dans les transformations lors de la propagation de ces représentations, et l'effet de leur accumulation au cours de la propagation.

Je développe ce dernier angle d'étude dans ma thèse, en m'intéressant aux niveaux lexical (évolution des variables correspondantes à chaque mot, comme par exemple la fréquence d'utilisation ou l'âge moyen d'acquisition, et influence de ces variables sur la probabilité d'être transformé), grammatical (relation entre type de structure et degré de transformation), et informationnel (perte d'information dans les phrases lors de la propagation). Mais quelle que soit la combinaison de niveaux envisagée, le passage à la sémantique s'avère extrêmement difficile : il n'existe pas d'abstraction formelle de la signification qui permette d'agréger quantitativement un grand nombre de transformations pour en rendre compte de façon synthétique et systématique. Le fait qu'une méthode qualitative et interprétative soit bien mieux adaptée à un tel objectif ne fait que renforcer l'observation : il est quelque chose de la sémantique que la science interprétative (à savoir dans notre cas, plutôt énactive) saisit parfaitement, et que la science quantitative ou « formelle » (plutôt représentationnelle) ne peut ni saisir ni circonscrire.

[5] https://twitter.com/LFCYlies/status/585127416621883393

[6] C'est-à-dire marqué comme « favori ».

[7] Une recherche simple sur Twitter montre que l'énoncé apparaît en moyenne une fois par mois, et la plupart des instances sont republiées plusieurs fois.

[8] https://twitter.com/rwlcff/status/585130263606390785

[9] https://twitter.com/LFCYlies/status/585131310366269440

L'étude des réseaux sociaux numériques semble refléter ce conflit : la discipline, qui a en grande partie utilisé le modèle mémétique pour l'étude de la diffusion (Adamic et al., 2014), voit se développer un courant s'intéressant à la « contagion complexe », c'est-à-dire l'étude de systèmes dont le processus de base n'est plus la transmission d'un objet atomique (un hashtag, une vidéo, une URL), mais la *contagion*, avec possible mutation, d'un objet plus compliqué doté d'une structure interne (par exemple une phrase ou un discours, Moussaïd et al., 2015). À ce jour, les principaux résultats portent sur des mesures non sémantiques telles que la transmissibilité (Miton et al., 2015) ou la mémorabilité (Danescu-Niculescu-Mizil et al., 2011), et tous sont obtenus à l'aide de méthodes alliant qualitatif et quantitatif.

La mémétique ne résout donc pas ce problème, ce qui l'a contraint à réduire son champ d'application et de questionnement en se limitant à des situations où il est su à l'avance que le modèle explique la plus grande partie du phénomène observé ; mais on remarquera que ce champ n'est pas clairement délimité et ne correspond pas à celui des réseaux sociaux numériques. Dépasser cette limite nécessite d'écarter le postulat de base d'abstraction du contexte de la mémétique ; il faut au contraire penser le contexte, comme la linguistique et l'anthropologie le font depuis longtemps et comme l'approche énactive des sciences cognitives est en train de développer.

2.3 Penser le contexte

Que ce soit par la pragmatique, les notions de situation de communication et situation de locution, le genre ou la scénographie (Maingueneau, 2004), l'analyse des différentes composantes du contexte et la façon dont elles participent à la signification ne sont pas des questions inédites en linguistique. La difficulté apparaît lorsqu'il s'agit d'utiliser ces notions dans une analyse à sensibilité quantitative (comme la mémétique), précisément parce que le passage est conçu comme une abstraction progressive du contexte, qui entraîne une perte de précision significative (Becker, 1996).

Je ne prétends bien sûr pas résoudre cette contradiction qui sous-tend les notions mêmes de qualitatif et de quantitatif, ni même esquisser une proposition qui pourrait s'y atteler ; le programme est trop vaste. Néanmoins le virage énactif des sciences cognitives apporte des éléments de réflexion : celui-ci consiste en effet à passer d'une science des contenus (définis en délimitant des frontières entre intérieur et extérieur, puis en s'attachant aux propriétés de l'intérieur) à une science des liens et de l'attention. Au niveau cognitif, au lieu d'étudier ce qu'un organisme peut faire en toute généralité, il s'agit de se concentrer sur ce qu'une situation donnée permet comme interactions, et étudier ce à quoi l'organisme porte attention dans sa situation. L'approche coupe ainsi transversalement la dichotomie séparant intérieur et extérieur, ou organisme et environnement, en considérant la relation organisme-environnement comme l'objet central de toute analyse. L'application de cette approche à la linguistique est en plein développement, et les récents travaux de Cuffari et al. (2014) promettent un éclairage extrêmement riche pour ces questions.

Conclusion

La mémétique et l'épidémiologie des représentations proposent des descriptions unifiées de la société. Dans leurs versions restreintes, ces propositions se traduisent en modèles applicables à l'analyse linguistique de la diffusion et des mutations dans les réseaux sociaux numériques. Si ces modèles apportent de l'intelligibilité aux réseaux en ligne, il faut garder à l'esprit que leur champ d'application est en réalité limité aux phénomènes de diffusion stricte ou aux mutations non sémantiques.

Tenter de généraliser ces approches, que ce soit au niveau théorique ou empirique, fait apparaître les dichotomies de fond qui les limitent. Plutôt que de chercher l'éclairage que la mémétique apporte à la linguistique, on peut alors se demander ce que la linguistique elle-même apporte à la mémétique.

L'approche non représentationnelle du langage, en particulier, semble pouvoir éclairer utilement la question de la sémantique dans son contexte ; une fois pleinement développée, il sera certainement intéressant de voir ce qu'elle peut apporter à l'étude des réseaux sociaux numériques.

Bibliographie

Lada A. Adamic, Thomas M. Lento, Eytan Adar, and Pauline C. Ng, February 2014. Information Evolution in Social Networks. *arXiv :1402.6792 [physics]*. URL http://arxiv.org/abs/1402.6792. arXiv : 1402.6792.

Gregory Bateson, 1972. *Steps to an Ecology of Mind : Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. University of Chicago Press. ISBN 978-0-226-03905-3.

Howard Saul Becker, 1996. *The epistemology of qualitative research*. Howard S. Becker, [S.l.]. URL http://www.soc.ucsb.edu/faculty/hbecker/qa.html.

Pierre Bourdieu, 1980. *Le sens pratique*. Editions de Minuit, Paris. ISBN 978-2-7073-0298-4.

Noam Chomsky, 2005. *Rules and Representations*. Columbia University Press. ISBN 978-0-231-13270-1.

Leda Cosmides and John Tooby, 1997. Evolutionary Psychology : A Primer. URL http://www.psych.ucsb.edu/research/cep/primer.html.

Elena Clare Cuffari, Ezequiel Di Paolo, and Hanne De Jaegher, 2014. From participatory sense-making to language : there and back again - Springer. doi : 10.1007/s11097-014-9404-9. URL http://link.springer.com/article/10.1007/s11097-014-9404-9/fulltext.html.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg, December 2011. Echoes of power : Language effects and power differences in social interaction. *arXiv :1112.3670 [physics]*. URL http://arxiv.org/abs/1112.3670.

Ronald A Fisher and J. H Bennett, 1999. *The genetical theory of natural selection : a complete variorum edition*. Oxford University Press, Oxford. ISBN 978-0-19-850440-5.

Jerry A Fodor, 1983. *The modularity of mind : an essay on faculty psychology*. MIT Press, Cambridge, Mass. ISBN 978-0-262-06084-4 978-0-262-56025-2.

James J. Gibson, November 2014. *The Ecological Approach to Visual Perception : Classic Edition*. Psychology Press. ISBN 978-1-317-57938-0.

Tim Ingold, January 1997. From the Transmission of Representations to the Education of Attention. URL http://lchc.ucsd.edu/MCA/Paper/ingold/ingold1.htm.

Tim Ingold, April 1999. Three in one : on dissolving the distinctions between body, mind and culture. URL http://lchc.ucsd.edu/MCA/Paper/ingold/ingold2.htm.

Tim Ingold, 2004. Beyond biology and culture. The meaning of evolution in a relational world. *Social Anthropology*, 12(2) :209–221. ISSN 1469-8676. doi : 10.1111/j.1469-8676.2004.tb00102.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8676.2004.tb00102.x/abstract.

Tim Ingold, 2007. The trouble with 'evolutionary biology'. *Anthropology Today*, 23(2) :13–17. ISSN 1467-8322. doi : 10.1111/j.1467-8322.2007.00497.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8322.2007.00497.x/abstract.

Aurelien Lauf, Mathieu Valette, and Leila Khouas, 2013. Analyzing Variation Patterns In Quotes Over Time. *Research in Computing Science*, 70 :223–232. ISSN 1870-4069. URL http://www.micai.org/rcs/2013_70/Analyzing%20Variation%20Patterns%20In%20Quotes%20Over%20Time.html.

Jure Leskovec, Lars Backstrom, and Jon Kleinberg, 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*, KDD '09, pages 497–506, New York, NY, USA. ACM. ISBN 978-1-60558-495-9. doi : 10.1145/1557019.1557077. URL http://doi.acm.org/10.1145/1557019.1557077.

Dominique Maingueneau, 2004. La situation d'énonciation, entre langue et discours. In *Dix ans de S.D.U.*, pages 197–210. Editura Universitaria Craiova, Craiova, Roumanie. URL http://dominique.maingueneau.pagesperso-orange.fr/pdf/Scene-d-enonciation.pdf.

Marcel Mauss, 2013. *Sociologie et anthropologie*. PUF, Paris. ISBN 978-2-13-060880-6.

Helena Miton, Nicolas Claidière, and Hugo Mercier, February 2015. Universal Cognitive Mechanisms Explain the Cultural Success of Bloodletting. SSRN Scholarly Paper ID 2560786, Social Science Research Network, Rochester, NY. URL http://papers.ssrn.com/abstract=2560786.

Mehdi Moussaïd, Henry Brighton, and Wolfgang Gaissmaier, May 2015. The amplification of risk in experimental diffusion chains. *Proceedings of the National Academy of Sciences*, 112(18) :5631–5636. ISSN 0027-8424, 1091-6490. doi : 10.1073/pnas.1421883112. URL http://www.pnas.org/content/112/18/5631.

J. Kevin O'Regan and Alva Noë, October 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(05) :939–973. ISSN 1469-1825. doi : 10.1017/S0140525X01000115. URL http://journals.cambridge.org/article_S0140525X01000115.

Richard O. Prum and Alan H. Brush, May 2014. Which Came First, the Feather or the Bird ? *Scientific American*, 23 :76–85. ISSN 0036-8733. doi : 10.1038/scientificamericandinosaurs0514-76. URL http://www.nature.com.gate3.inist.fr/scientificamerican/journal/v23/n2s/full/scientificamericandinosaurs0514-76.html.

Matthew P. Simmons, Lada A. Adamic, and Eytan Adar, July 2011. Memes Online : Extracted, Subtracted, Injected, and Recollected. In *Fifth International AAAI Conference on Weblogs and Social Media*. URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2836.

Dan Sperber, 1996. *Explaining culture : a naturalistic approach*. Blackwell, Oxford, UK ; Cambridge, Mass. ISBN 978-0-631-20045-1 978-0-631-20044-4.

Mary Jane West-Eberhard, 2003. *Developmental plasticity and evolution*. Oxford University Press, Oxford ; New York. ISBN 978-0-19-512234-3 978-0-19-512235-0.

Ludwig Wittgenstein, March 2010. *Philosophical Investigations*. John Wiley & Sons. ISBN 978-1-4443-0797-9.

The Gistr Platform

Sébastien Lerique*

Introduction

Studies of online interactions that are part of the "born-digital" sub-field of Digital Humanities usually focus on activities for which online records are readily available (mostly, blogs and large online social networks with varying degrees of access to the data). In stark contrast to "in vitro" laboratory experiments, this "in vivo" approach studies human interactions in an ecological environment by analysing their digital traces.[1]

Yet despite the large amounts of data and computing power that this quantitative observation approach takes advantage of, it remains very difficult to ask questions linking cognitive and social levels of human interaction, and the revolution that was expected to happen in social science has yet to materialise. More precisely: the question, both theoretical and experimental, of how to study and link together cognition and what social science defines as *social* proper, is still very much open.

Since better tools afford better questions and better theory, we focus here on the experimental side of this problem which is in fact quite simple: online interaction studies are based on digital traces, but most researchers have no control over what data is recorded in those traces, or over the way the interaction is framed and defined by the technical system that mediates it. As a result, only a very specific class of questions can be asked by online interaction studies: those that use only the available data and focus on the existing interaction framing, excluding any question for which some information is missing and not inferable or available elsewhere, as well as questions that need to tinker with the interaction framing. The latter is a very high barrier to asking questions that link the cognitive and social levels of human interaction.

The purpose of this document is to briefly introduce the approach of Web and Smartphone experiments – a promising method for human interaction experiments – and the trade-offs it makes, and Gistr, the Web experiment I am currently developing as part of my PhD thesis to overcome the shortcomings of our previous study on sentence transformation in blogspace.

1 Web and Smartphone experiments

Complementary to studies using existing digital traces, Web and Smartphone experiments strike a different balance in the trade-offs of experimental work and seem very promising in addressing the problems outlined above (Miller 2012).

*Centre d'Analyse et de Mathématique Sociales (UMR 8557, EHESS, Paris), and Centre Marc Bloch (Berlin). Email: sebastien.lerique@normalesup.org.

[1] I use the terms *in vivo* and *in vitro* in analogy to their meaning in biology. As explained in the main text, *in vivo* refers to the study of human interactions in their ecological environment, be it through direct observation or through digital traces of those interactions. *In vitro* refers to the study in conditions defined by the researcher, usually in a laboratory experiment.

Indeed, smartphones and browsers (both on desktop and mobile) have evolved into powerful, ubiquitous application environments for which one can relatively easily develop any kind of experiment involving text, graphics, and human interactions. At the cost of higher engineering and recruitment efforts, this approach gives the experiment designer full control over what data is collected and the way interactions are framed. Given the omnipresent nature of smartphones, such experiments can also be deeply embedded in everyday life which, as I explain below, opens even more possibilities for questions on social interaction.

1.1 The need for embedding

Any quantitative study relies on abstracting out details of particular cases by reducing (most often averaging) values in each dimension to a few indicators. Being able to render a precise view of the studied phenomenon then depends on being able to determine which are the right dimensions to describe it, and having access to them (Becker 1996).

Embedding experiments in the everyday life of subjects gives access to dimensions that can be otherwise unavailable: through the use of smartphones, an experiment designer can trigger interactions with subjects (for instance asking questions) at any moment of the day, or have measures running while subjects are offline (using the ever-increasing number of sensors present on the devices), both of which are impossible with digital traces. Above all, embedding an experiment means getting greater access to context, which opens the possibility of understanding phenomena the way they are meant in the lives of subjects, and not only in the way they are construed by the experiment designer.

Daydreaming is an example experiment developed as a smartphone application with Vincent Adam, Mikaël Bastian, Jérôme Sackur, and Gislain Delaire, that took advantage of this embedding. The experiment, focused on our awareness of daily mind-wandering, would probe subjects during a month at random moments of the day to ask them if they were mind-wandering (and, if so, what were the qualities of their thoughts).[2] While our team spent over a year developing the application, it allowed asking questions related to ecological situations which cannot exist in laboratory or passive collection studies. Figure 1 shows a sample question asked to the subjects, and Figure 2 shows an example of the results produced at the end of the experiment (this particular screen shows the results for one subject; seeing their own results was part of the reward for subjects participating in the study).

1.2 Pros and Cons

Experiments in the browser and on smartphones make specific trade-offs which differ from most other methods. The most important points are as follows.

Pros

- *Control*: similar to laboratory experiments, complex designs are possible, and all the interactions the subjects can be involved in are defined by the experiment designer. This includes for instance the way in which the experiment is framed (e.g. as a game or a self-improvement aid, aside from being a scientific study) and, more importantly, the ways in which subjects can interact with each other through the experiment.
- *Embedding*: as explained above, smartphone-based experiments allow for real-life embedding: the experiment designer can choose when and how interactions with the experiment and between subjects take place, and measure any number of variables the device gives them access to (geolocation, time, phone agitation through its accelerometers, general noise level, etc.), virtually at any moment.

[2] See http://daydreaming-the-app.net/ for more details.

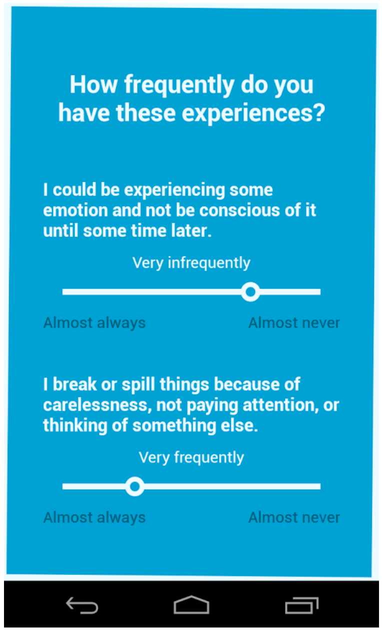


Figure 1: Example questionnaire in the Daydreaming app

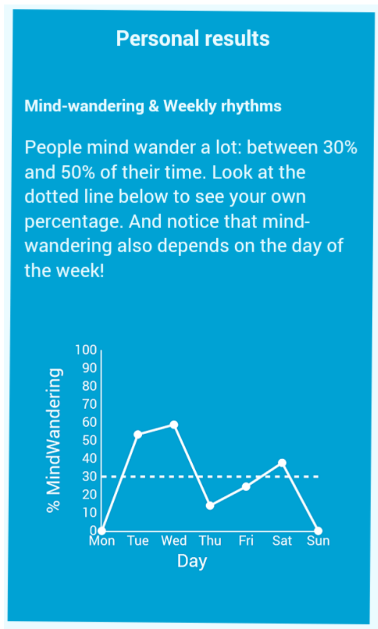


Figure 2: Results on weekly mind-wandering rhythms from the Daydreaming app

- *Scale*: if needed, the technical platform can relatively easily scale the number of subjects to several tens of thousands. This also applies to interactions between subjects, which can directly involve many people, be it at the same time or (for instance) in chains.
- *Flexible recruitment*: subject recruitment, while also a challenge (see *cons* below), is very flexible: services like Prolific Academic[3] let experiment designers recruit at reasonable costs in pools of tens of thousands of subjects with fine-grained demographic filters. Aiming for wider audiences can be done by offering rewards, framing the experiment as a self-improvement application, or turning it into a game.

Cons

- *Technical challenge*: developing such an experiment involves a substantial amount of engineering and makes use of several technologies most researchers, even technical, are not familiar with. While some all-in-one kits exist,[4] creating an experiment that meets one's research often requires learning average skills in most of the various technologies at play: a native or cross-platform smartphone environment, Web application development, backend server programming, and some server administration skills. Most importantly, the paradigms and problems encountered are often new to researchers: most programming is asynchronous due to network communication or user interface, and technicalities such as user management or email validation can grow into real engineering challenges.
- *Spam-control*: subjects are not constrained or encouraged by the face-to-face interaction of a laboratory experiment, neither are they (in most experiments) in the course of an interaction with friends where is it natural to them to not write spam, as can be the case in digital traces. Participants must have an incentive to perform the experiment's tasks well. If spam is naturally isolated in the experiment's design, one can for instance filter it once the data is collected and make payment depend on it prevalence. But if spam propagates in the experiment, real anti-spam pressures and motivations need to be factored into the whole design.
- *Recruitment cost*: while recruiting a few dozen or even a few hundred subjects is generally cheaper (not counting the cost of developing the experiment) than the equivalent for a laboratory experiment,[5] and can be very easy for fast prototyping and pilots, recruitment cost rises linearly with the number of subjects and the time they spend on the experiment, unless a different strategy is used. Turning an experiment into a playful application or an application useful to the user involves its own set of skills, can prove challenging, and must be factored into the development cost.

2 The Gistr Platform

2.1 Rationale

As part of my PhD thesis we are studying the transformation of short sentences – such as quotations from politicians or spokespeople – as they are propagated through various media. Our first study focused on the evolution of such short quotations as they are copied from blog to media outlet to blog (Lerique and Roth 2015). Indeed, authors often transform quotations when publishing them online despite the implicit and common-sense injunction to quote people verbatim: a few words disappear, a contraction appears, the quote is cropped, and so on and so forth (Simmons, Adamic, and Adar 2011; Lauf, Valette, and Khouas 2013). Given this observation, the data collected by Leskovec, Backstrom, and Kleinberg (2009) is at first sight a very good candidate to study the evolution of online content

[3] https://www.prolific.ac/
[4] See e.g. http://funf.org/ and http://www.epicollect.net/.
[5] Global competition on online platforms like Prolific Academic drive subject payments down.

as it is transformed by users. But the actual analysis proved itself much more challenging, for two fundamental reasons:

- *Missing information*: most blog and media outlet authors do not quote their source when they publish a quote online (it's often not relevant to the article), meaning there are no source-destination links in the data collected; this information must instead be inferred anew to study the evolution of content. There is also no access to author information (gender or age, experience in writing, but also psychological factors like memory span), ruling out any study of individual author effects in transforming the content.
- *Missing context*: the lack of access to the context of production and reception of quotes makes it impossible to interpret what a quotation means to its author or its reader (Wittgenstein 2010; Briggs 1992; Cuffari, Di Paolo, and De Jaegher 2014). Analysing any kind of semantic evolution is therefore out of reach for this kind of passively collected online data (Lerique 2016, to be published).

The Gistr platform[6] emerged from a concern to address these two problems by taking advantage of the possibilities offered by Web experiments. The general aim for this project is the study of interpretation and sense-making of short sentences in particular contexts, and the question of how interpretation and sense-making have global scale effects when accumulated and iterated.

2.2 State of the Art

This experiment aims to shed some light on the *cultural attractors* hypothesis presented by Dan Sperber in his work "Explaining culture: a naturalistic approach" (Sperber 1996). Up to now Epidemiology of Representations, the theory behind the idea of cultural attractors, has mainly focused on the evolution of cultural bodily practices with long intergenerational lifecycles like religion (Boyer 2001), smoking (Claidière and Sperber 2007), the way portraits are made (Morin 2013), and the practice of bloodletting (Miton, Claidière, and Mercier 2015). The field has also started studies of diachronic evolution of language (Claidière et al. 2014).

Practices with short intragenerational lifecycles that have less to do with changes in bodily practice and more with interpretation have also been recently studied, like for instance music (MacCallum et al. 2012) or risk perception (Moussaïd, Brighton, and Gaissmaier 2015).

With the development of the Gistr platform, we aim to bring a new case in this area of short lifecycle opinion dynamics by studying the semantic evolution of short sentences and short stories in interpretation chains. What change takes place here is mainly due to interpretation and the reconstructional component of memory which involves many levels and is influenced by many factors. Therefore, after starting at the macro scale where individual variation and context details are abstracted out, we also aim to gradually move towards the mesoscopic scale, integrating more contextual and personal details and factors as the experiments are iterated.

2.3 General approach

Interpretation is currently at a theoretical crossroads between fodorian (Fodor 1983) and formal semantic analyses on one side, and enactive accounts of languaging on the other side (Cuffari, Di Paolo, and De Jaegher 2014). Epidemiology of Representations rests on the former: in this account of cognition and interpretation, the brain is a storehouse of representations and most of what interests us happens in that storehouse (perception, processing and transformation of representations).

[6] As explained further down, we aim for a series of experiments based on the same paradigm and gradually introducing complexity to the problem. Hence the term *platform*.

During the 2000's however, a profound shift in philosophy of mind departed from this position and developed a theory positing the central role of the environment. In this theory, cognition is seen as a way of exploring and relating to the environment (vs. storing representations of that environment), and as such cannot be considered out of or separated from it (some philosophers go so far as to consider that the environment is part of the cognitive device). This line of thought has greatly improved the way cognitive science analyses contextual and situational information and interaction, has led to many experimental breakthroughs (mostly in perception; of particular interest are Sensory Substitution Devices (O'Regan and Noë 2001)), and has a number of implications for how perception and affect should be tackled (see for instance (Bower and Gallagher 2013)).

This approach, usually labelled "enactive",[7] construes interpretation as a kind of attention-perception-action loop that constitutes exploration of a meaningful environment through "educated attention" (Ingold 1997; Cuffari, Di Paolo, and De Jaegher 2014). While the enactive approach does bring an extremely promising conceptualisation of interpretation and how it can be studied, we chose to design the experiment following the paradigm of transmission chains which is a clear application of the more fodorian ideas behind Epidemiology of Representations. This paradigm is also very analogous to real life situations of interest (such as the propagation of short sentences in blogspace) and, as we are currently seeing with the first results, is useful in pinpointing the shortcomings of the fodorian approach and highlighting the areas in particular need of an enactive approach to interpretation.

Other disciplines, such as Social Anthropology, oppose valuable criticisms to this design (see in particular Briggs 1992; Ingold 1997; Ingold 1998; Ingold 2004). However as explained above, the development will first focus on Epidemiology of Representations itself and, if time allows, gradually incorporate criticism as it serves the purpose of explaining the collected data and refining the conditions. (The enactive critique will, for instance, prove useful in trying to develop conditions taking context and interaction into account, which in turn will likely allow us to explain some noise.)

2.4 Breakout and development

Let us now present the current state of the experiment itself, and the directions we will develop it in.

2.4.1 Breakout

In the first iteration we aimed to explore some interpretation effects, at the single and cumulative levels, in tasks involving sentence and story rewriting or reformulation. Since interpretation involves so many levels of complexity, we started with the simplest possible condition (although it is probably underspecified), and will add new measures and conditions through further iterations, guided by the questions and the noise of the previous iteration.

To do this we built a Web experiment aimed at generating trees of short sentences that have been repeatedly memorised and rewritten (i.e. interpreted and reconstructed) by a large number of subjects. Recruited participants went through a series of steps:

- sign up (Figure 3),
- setting their mothertongue and answering a preliminary questionnaire (Figure 5),
- they then started training for the main task, consisting in repeatedly memorising and rewriting short sentences as accurately as possible. As the instructions show in Figure 6, a sentence is presented to the subject, and after a short pause, the subject must rewrite the sentence as they remember it. The whole process loops until the experiment is finished. The real trials started after five training trials.

[7] The family of approaches developed following the shift in philosophy is also known as the "4Es": embodied, embedded, extended, and enactive cognition.

Once the subjects completed the experiment, the application switched to a game mode where each subject could suggest new sentences (depending on the number of sentences they have already transformed) that get fed to other game-mode subjects (thus keeping experiment and game sentences isolated). Subjects could also explore the interpretation trees generated by the experiment and see how content was transformed along transmission chains (Figure 7).

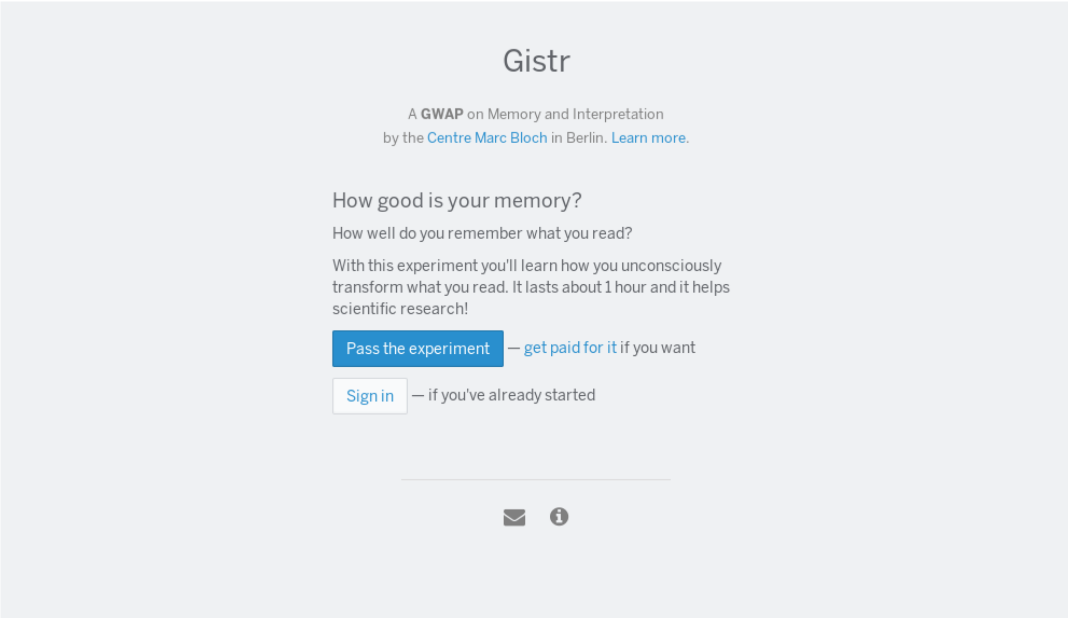


Figure 3: Gistr welcome screen

The initial sentences we selected consisted in quotations from the MemeTracker dataset (Leskovec, Backstrom, and Kleinberg 2009), famous quotes obtained on WikiSource, and a few quotations from the 1957 film "12 Angry Men". Each subject was presented either with original sentences, or with sentences further down a transmission chain, that is resulting from the iterated interpretation of other subjects.

Measures

In each iteration of the experiment, we will measure the following (among several other exploratory measures):

- Possible correlation between the transformation rate of sentences[8] and age and gender of the authors,
- Cumulative transformation rate of sentences as a function of their depth in the tree,
- Transmissibility of sentences as defined by Claidière et al. (2014).

[8]Sentence *transformation rate* is measured by (1) extracting content words from the sentence by removing stopwords and lemmatising, (2) computing the content-word-based edit distance between the sentence and its parent, normalized to the maximum number of content words between the sentence and its parent.

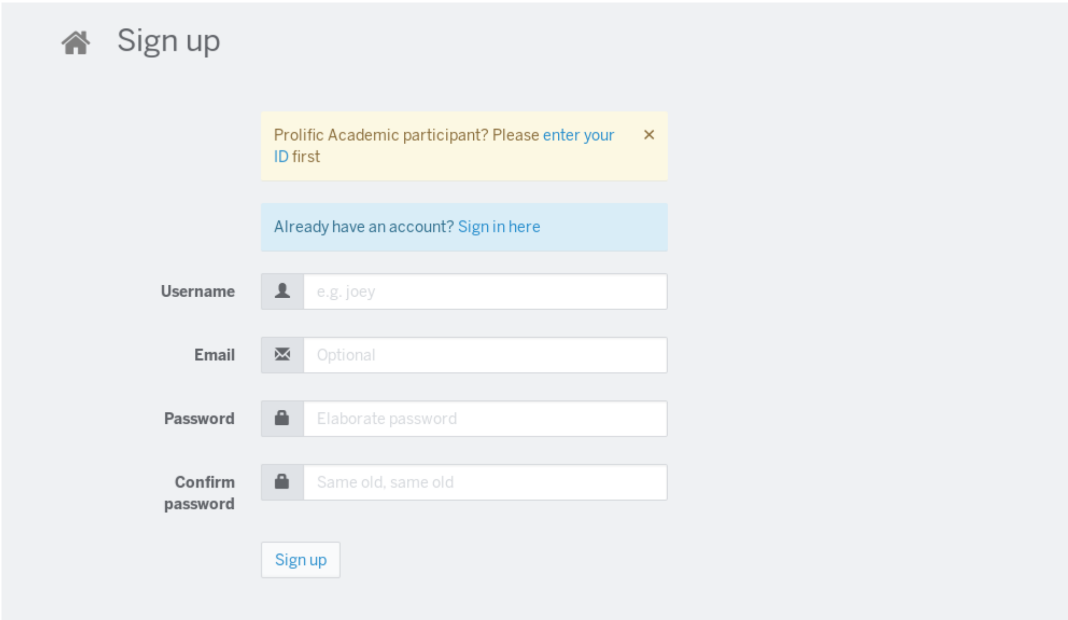


Figure 4: Gistr signup screen

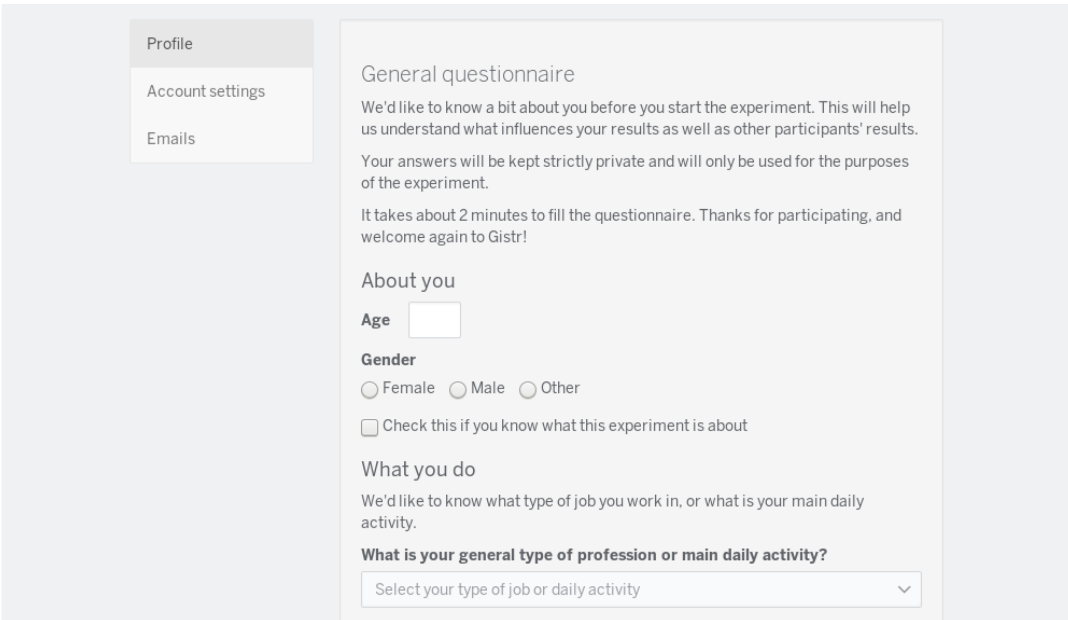


Figure 5: Initial questionnaire

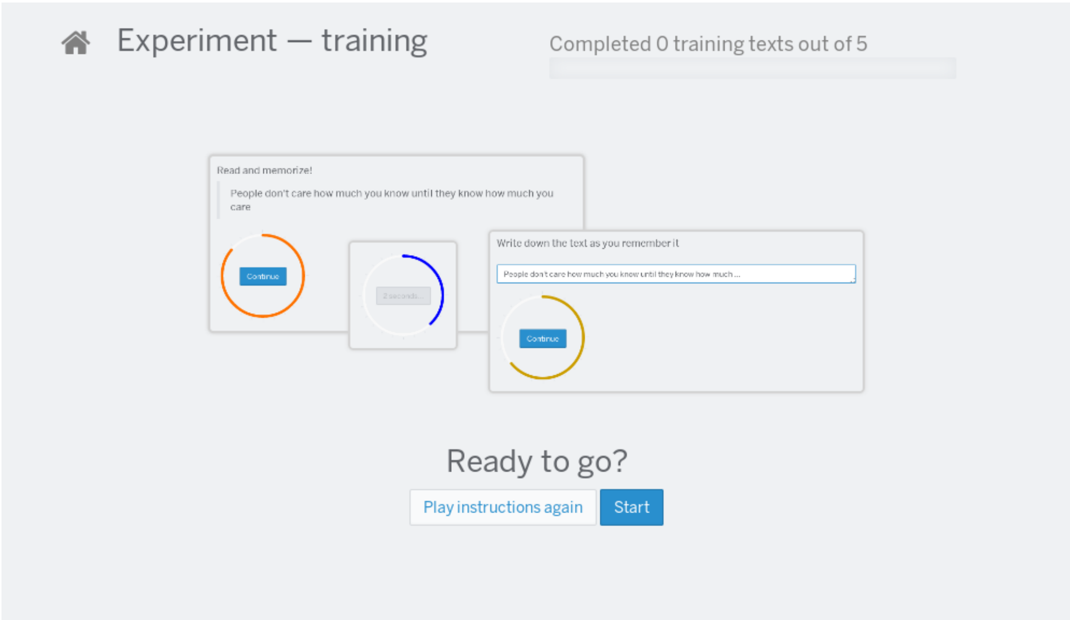


Figure 6: Experiment instructions and demo

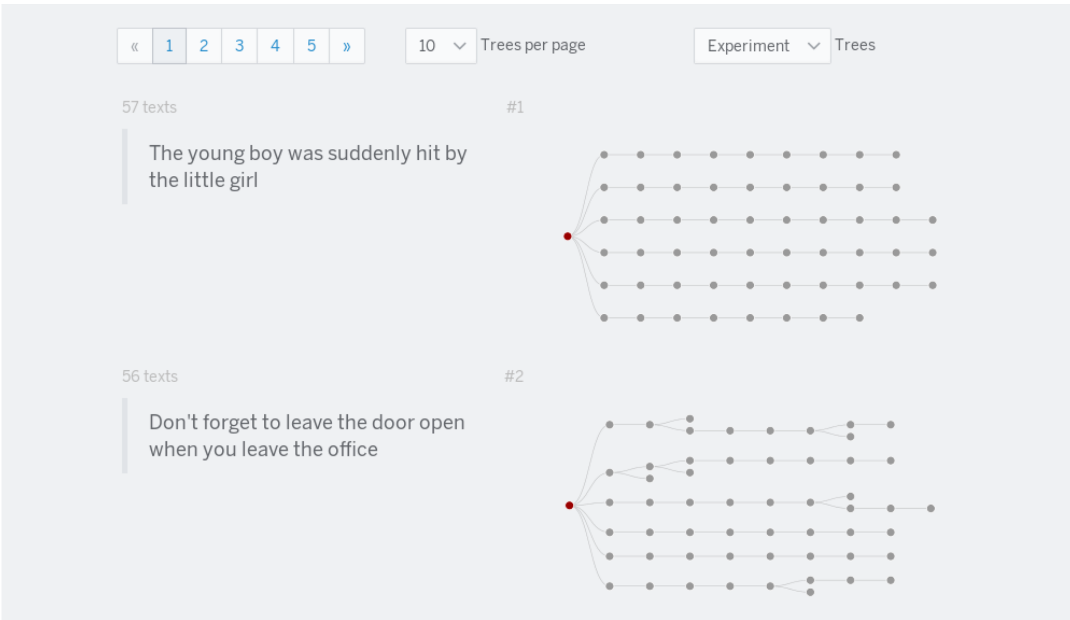


Figure 7: Interpretation trees resulting from the experiment

2.4.2 Development

The experiment, freely accessible on https://gistr.io/, lets us start experimenting on paid (Prolific Academic) or volunteer (Crowd Crafting[9]) platforms, and later advertise for platform-free participation.

The technical development itself happens on the project's repositories, https://github.com/interpretation-experiment/gistr-app/wiki, and the scientific design and exploration is documented on the project's Open Science Framework repository, https://osf.io/k7d38/. All the code written for this experiment (browser application, server backend, data analysis) is released as free software.

References

Becker, Howard S. 1996. 'The Epistemology of Qualitative Research'. In *Ethnography and Human Development: Context and Meaning in Social Inquiry*, edited by R. Jessor, A. Colby, and R. A. Shweder, 53–71. The John D. and Catherine T. MacArthur Foundation Series on Mental Health and Development. Chicago, IL, US: University of Chicago Press.

Bower, Matthew, and Shaun Gallagher. 2013. 'Bodily Affects as Prenoetic Elements in Enactive Perception'. *Faculty of Law, Humanities and the Arts - Papers*, January, 78–93. http://ro.uow.edu.au/lhapapers/1279.

Boyer, Pascal. 2001. *Religion Explained: The Evolutionary Origins of Religious Thought*. Basic Books.

Briggs, Jean L. 1992. 'Mazes of Meaning: How a Child and a Culture Create Each Other'. *New Directions for Child and Adolescent Development* 1992 (58): 25–49. doi:10.1002/cd.23219925804.

Claidière, Nicolas, and Dan Sperber. 2007. 'The Role of Attraction in Cultural Evolution'. *Journal of Cognition and Culture* 7 (1): 89–111. doi:10.1163/156853707X171829.

Claidière, Nicolas, Kenny Smith, Simon Kirby, and Joël Fagot. 2014. 'Cultural Evolution of Systematically Structured Behaviour in a Non-Human Primate'. *Proceedings of the Royal Society of London B: Biological Sciences* 281 (1797): 20141541. doi:10.1098/rspb.2014.1541.

Cuffari, Elena Clare, Ezequiel Di Paolo, and Hanne De Jaegher. 2014. 'From Participatory Sense-Making to Language: There and Back Again - Springer'. doi:10.1007/s11097-014-9404-9.

Fodor, Jerry A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.

Ingold, Tim. 1997. 'From the Transmission of Representations to the Education of Attention'. http://lchc.ucsd.edu/MCA/Paper/ingold/ingold1.htm.

———. 1998. 'From Complementarity to Obviation: On Dissolving the Boundaries Between Social and Biological Anthropology, Archaeology and Psychology'. *Zeitschrift Für Ethnologie* 123 (1): 21–52. http://www.jstor.org/stable/25842543.

———. 2004. 'Beyond Biology and Culture. The Meaning of Evolution in a Relational World'. *Social Anthropology* 12 (2): 209–21. doi:10.1111/j.1469-8676.2004.tb00102.x.

Lauf, Aurelien, Mathieu Valette, and Leila Khouas. 2013. 'Analyzing Variation Patterns In Quotes Over Time'. *Research in Computing Science* 70: 223–32. http://www.micai.org/rcs/2013_70/Analyzing%

[9] http://crowdcrafting.org/

20Variation%20Patterns%20In%20Quotes%20Over%20Time.html.

Lerique, Sébastien. 2016. 'Pour Une étude Du Contexte d'interprétation'. *To Be Published.*

Lerique, Sébastien, and Camille Roth. 2015. 'How Do We Copy and Paste? The Semantic Drift of Quotations in Blogspace'.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. 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*, 497–506. KDD '09. New York, NY, USA: ACM. doi:10.1145/1557019.1557077.

MacCallum, Robert M., Matthias Mauch, Austin Burt, and Armand M. Leroi. 2012. 'Evolution of Music by Public Choice'. *Proceedings of the National Academy of Sciences* 109 (30): 12081–6. doi:10.1073/pnas.1203182109.

Miller, Geoffrey. 2012. 'The Smartphone Psychology Manifesto'. *Perspectives on Psychological Science* 7 (3): 221–37. doi:10.1177/1745691612441215.

Miton, Helena, Nicolas Claidière, and Hugo Mercier. 2015. 'Universal Cognitive Mechanisms Explain the Cultural Success of Bloodletting'. SSRN Scholarly Paper ID 2560786. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2560786.

Morin, Olivier. 2013. 'How Portraits Turned Their Eyes Upon Us: Visual Preferences and Demographic Change in Cultural Evolution'. *Evolution and Human Behavior* 34 (3): 222–29. doi:10.1016/j.evolhumbehav.2013.01.004.

Moussaïd, Mehdi, Henry Brighton, and Wolfgang Gaissmaier. 2015. 'The Amplification of Risk in Experimental Diffusion Chains'. *Proceedings of the National Academy of Sciences* 112 (18): 5631–6. doi:10.1073/pnas.1421883112.

O'Regan, J. Kevin, and Alva Noë. 2001. 'A Sensorimotor Account of Vision and Visual Consciousness'. *Behavioral and Brain Sciences* 24 (05): 939–73. doi:10.1017/S0140525X01000115.

Simmons, Matthew P., Lada A. Adamic, and Eytan Adar. 2011. 'Memes Online: Extracted, Subtracted, Injected, and Recollected'. In *Fifth International AAAI Conference on Weblogs and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2836.

Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach.* Oxford, UK; Cambridge, Mass.: Blackwell.

Wittgenstein, Ludwig. 2010. *Philosophical Investigations.* John Wiley & Sons.