

PhD Dissertation

Epidemiology of Representations: An Empirical Approach

—original title may change—

Sébastien Lérique¹

Supervisor: Jean-Pierre Nadal²
Co-supervisor: Camille Roth³

¹Centre d'Analyse et de Mathématique Sociales (CAMS, UMR 8557, CNRS-EHESS, Paris). Email: sebastien.lerique@normalesup.org.

²CAMS, and Laboratoire de Physique Statistique (LPS, UMR 8550, CNRS-ENS-UPMC-Univ. Paris Diderot, Paris). Email: nadal@lps.ens.fr

³CAMS, Centre Marc Bloch (CMB, UMIFRE 14, CNRS-MAEE-HU, Berlin), and Sciences Po, médialab (Paris). Email: camille.roth@sciencespo.fr

Contents

1	Introduction	5
2	Brains Copy Paste	7
3	Gistr	9
4	Discussion	11
4.1	Introduction	11
4.2	Empirical epidemiology of linguistic representations	12
4.2.1	Relevant results	12
4.2.2	Challenges	13
4.3	Approaches to meaning	17
4.3.1	Relevance theory	17
4.3.2	The enactive approach	18
4.3.3	Outside the linguistic domain	18
4.4	Down to empirical study	18
5	Conclusion	19
	References	21

Chapter 1

Introduction

Chapter 2

Brains Copy Paste

Chapter 3

Gistr

Chapter 4

Discussion

4.1 Introduction

In this chapter, we aim to take a broader view on what would be necessary to achieve a fuller understanding of the processes at work in cultural change at the linguistic level. So far we have adopted wholesale the paradigm put forward by Cultural Attraction Theory, by seeking to identify and elucidate situations where linguistic representations are transformed as they are transmitted, and assessing, on one side, the extent to which the empirical evolution of content agrees with what is expected under CAT, and on the other side, the extent to which CAT provides productive guiding questions in understanding what is at work in the situations studied. This has led us to identify a number of behaviours which are consistent with Cultural Attraction Theory: studying word substitutions in online quotations first, and more general transformations in controlled transmission chains of short utterances second, we showed that the low-level lexical features of words evolve in a systematic manner to make utterances easier to produce, and that the direction of the evolution is consistent with the attraction pattern that can be observed in the individual step of word replacements. However, these approaches did not bring us any closer to understanding the semantic changes that utterances undergo when they are transformed, be it online or in controlled transmission chains.

We now wish to discuss the reasons for this limitation. Our purpose is first to convince the reader that meaning is a crucial aspect in the evolution of content which must eventually be analysed in order to fully understand the way representations circulate and change. Manual exploration of the changes in transmission chains in particular show that the surface measures that we used in quantitative analysis have no handle on the evolution of meaning. Indeed, meaning will appear as a deeply context- and interaction-dependent property, which cannot be understood by simply focusing on the utterances themselves. Second, we aim to show how this challenge can be traced to what is known in philosophy of mind as the “hard problem of content”, and to how approaches to pragmatics deal with it. We will thus discuss two important approaches to studying the meaning of utterances in relation to the context and interaction in which they appear: Relevance Theory and the Enactive approach. The first is better developed and integrated with linguistics, but must face some version of the hard problem of content in order to provide an operational account of meaning. The second starts from a simpler endogenous notion of meaning which avoids the problem of content, but has yet to prove its viability and usefulness for the study of language. Favouring one or the other, or possibly combining parts of the two, is further related to the overall construal of cultural evolution

and to the importance of representations in a theory of cultural change, as critiques of the cultural attraction framework have shown. Finally we believe that the question of meaning in cultural change, and the debates it relates to, can be moved forward through informed empirical investigation. After having laid out the alternatives, our final goal is therefore to put forward the approaches we believe are most useful to turn this problem into an empirical question.

We begin by discussing detailed examples of the role of semantics in our transmission chain experiment, to show how the lack of an account of utterance meaning renders the empirical question of attractors in this case under-specified. Next, we present in more detail two possible approaches to pragmatics and meaning, and discuss their relationship to an overall view of cultural change. Finally, we present possibilities for refining and advancing the debate through empirical investigation.

4.2 Empirical epidemiology of linguistic representations

4.2.1 Relevant results

The path we took so far has consisted in entirely adopting the cultural attraction paradigm and developing experiments to evaluate one of its strong hypotheses, namely the existence of attractors in the evolution of representations. Indeed, cultural attractors are in many ways a cornerstone for the theory, as they reflect its explanation of the stability of culture in spite of strong micro-level transformations (they are the product of ecological and psychological factors interacting with each other), and they provide intelligibility into the complexity of cultural change as a whole. Linguistic utterances appeared as a good proxy to study representations that are part of everyday life and for which large corpora are readily available. Language is also one of the most versatile means by which representations circulate, making linguistic utterances an important study case for the theory.

Our initial high-level question was thus whether attraction could be observed in the evolution of linguistic utterances as they are interpreted and produced anew by successive people. The first case-study we developed relied on online quotations, a type of representation for which an implicit rule mandates perfect copy, yet which often changes as it propagates across blogs and news outlets. Our investigation of single-word replacements showed that, when transformed, words are reliably replaced by words easier to produce. Evaluated on standard lexical features, individual word replacements showed an attraction pattern specific to each feature and consistent with the hypothesis of an attractor at the lexical level, which could be due to cognitive biases in word production. Our second case-study explored utterance transformations in a more controlled situation, by setting up artificial transmission chains of short utterances on an online platform. Here, the analysis first focused on developing a descriptive model that would provide an overview of transformations decomposed into more basic operations. We showed that transformations can be reliably described as made of chunks of word insertions and deletions interspersed with individual word replacements (as well as chunk exchanges, which were set aside for the analysis). This level of description showed that the transformation process has several regularities: operations strongly depend on each other (in particular, insertions appear to make up in size for some of the deletions, while still introducing substantial change), and their prevalence also depends on the length of, and their position in, the utterance (longer utterances receive more operations; replacements preferentially target the interior of utterances, and insertions and deletions the second half of utterances). The behaviour of insertion and deletion chunks, as well as replacements, was shown to be consistent with the biases identified in individual replacements in online quotations: the susceptibilities for being targeted by deletion or replacement, and appearing by insertion or replacement, closely complemented each other, in

accordance with the hypothesis of an attractor at the lexical level; the overall evolution of the lexical makeup of utterances also reflected those biases by drifting in a specific direction on each lexical feature (corresponding to easier recall). More generally, we argued that the modeling approach provides crucial detail about the transformations, achieving a middle-ground between the focus on lexical features in individual word replacements and the wide-angle view of contrasts in the aggregated evolution of content along chains.

These analyses were made at the cost of several trade-offs. Transformations in the online data set were restricted to single-word replacements so that we could infer missing source-destination links between quotations, and lack of data meant that no analysis could be made of the context surrounding the quotations. The transmission chain experiments were led with an extremely simple read-and-rewrite task (though this was an intentional first step), which also did not open the analysis to the role of context in transformations and in the overall evolution of content. Nonetheless, these studies demonstrate that it is possible to decompose the transformations of utterances into combinations of smaller operations, and fully connect the behaviour of those operations with known effects in psycholinguistics, be it online (with a partial view of the process) or in controlled transmission chains (with a full view of the transformations). They further suggest that, due to cognitive biases in the way utterances and words are recalled, the evolution of short utterances like quotations could be subject to an attractor at the lexical level, making the words of utterances gradually easier to recall, on top of other changes in the actual content conveyed.

4.2.2 Challenges

However, these studies do not tell us the way utterances evolve semantically. Indeed, apart from the vector-based comparison of individual words for scoring matched and mismatched pairs in utterance alignments (an arguably simplistic approach to word comparison), none of the analyses we put forward have a grip on the meaning of the utterances, and much less on the change in meaning upon transformation. While it is noteworthy that it was still possible to extract reliable decompositions of the transformations without such information (as the manual evaluation of alignments attests), these analyses are blind to changes in the content circulated by the utterances.¹

Let us show a few examples of the types of meaning change that were observed in the transmission chain experiments of the previous chapter.

Minor operations can change the function of a part of the utterance

Consider the following root utterance from in Experiment 2:

“Can you think of anything else, Barbara, they might have told me about that party?” (4.1)

The second part of this sentence is slightly misleading, and could be seen as a mild case of garden path sentence:² to “tell about” can be either transitive or intransitive, and while the final “that party”

¹We also explored the semantic distance traveled by words upon replacement, and the possible hyponym-hypernym relationships between replaced and replacing words, but did not present the analyses in the previous chapter as they provided no additional insight about the process.

²A garden path sentence is a sentence that misleads the reader into parsing its syntax one way, but necessitates a different structure to be understood once all the words have been read. A classic example is the sentence “The horse raced past the

determines it as a transitive verb (for which it is the object), several subjects in the experiment rewrote the following sentence:

"Can you think of anything else, Barbara, they might have told me about **at** that party?" (4.2)

The added "at" turns "tell about" into an intransitive verb, and turns the final part of the sentence into an adverbial phrase of time, thus changing its function in the sentence. More importantly, the sentence in its new form implies that the speaker was at the party, whereas the original sentence implies the contrary (although one could imagine the speaker was present but does not remember the details of the party). There is therefore a substantial change in the high-level meaning of this utterance, through the addition of a single word which changes the function of part of the utterance.

Once this change has occurred, regularisations often happen in the rest of the branch, for instance removing "about" to turn the ending of the sentence into "told me at the / that party". Looking at the leaves of the seven branches this tree contains, 4 of them maintain the implication that neither the speaker nor Barbara were at the party, 2 imply that the speaker (and not Barbara) was at the party, and one implies that Barbara was at the party (and the speaker was not).

Minor operations can create an ambiguity triggering larger changes

As we discussed at the end of the previous chapter, minor changes can also lead to larger downstream consequences in surface representation (as well as in meaning). A second example of typographical error can be seen with the following sub-chain in Experiment 3 (putting aside the UK/US spelling change, "canceled" → "cancelled"):

"The charge of embezzlement against the artillery has been canceled." (4.3)

"The charge of embezzlement **again** the artillery has been cancelled." (4.4)

"The charge of embezzlement again, the charge has gone." (4.5)

In this case, the "against" → "again" replacement operated in the first transformation leads the following subject to interpret the sentence quite differently, making an larger change. This behaviour is far from systematic, as many times such small errors are corrected by later subjects. Consider for instance the following error made by a subject in Experiment 2:

"At least when they say they're going to have a war, they keep their word." (4.6)

"At least when they say they are going to have a war, they keep **there** word." (4.7)

The "their" → "there" replacement is maintained by the next subject, but then reverted by the one after that, thus coming back to the original sentence (aside from the change in contraction, "they're" → "they are").

barn fell", which misleads the reader into interpreting "The horse" as the subject of "raced"; the correct parse corresponds to the meaning of "The horse that was raced past the barn fell", where "The horse" is the object of "raced". The difficulty comes from the fact that the search for the correct parse becomes necessary only once the reader has seen the final word, "fell".

Weak and strong pragmatics

The examples above, and the variability they illustrate, testify to the fact that different subjects can interpret the same utterance in strongly divergent ways. Subjects do not only differ on their performance in accurately reproducing the utterances presented, their productions also signal that different meanings can be constructed from the same root utterance (such differences accumulate, as was illustrated by the divergence of branches observed in the previous chapter). Part of this observation is commonplace, as the meaning of an utterance depends in obvious ways on the context in which it is produced or read, such as when deictics are used (words such as “today” or “here”, which are context-bound by nature). However, most isolated utterances are under-specified in a way that makes them much more dependent on the context and on the interaction they appear in than what deictics suggest.

Consider once again utterance 4.1. With no further context, it is not clear what party the speaker is referring to, who were the participants, or why the speaker is asking about it. As the interpretations made by subjects in Experiment 2 illustrate, one could imagine that the speaker was at the party but does not recall its events, or that Barbara witnessed someone telling the speaker about the party, a telling that the speaker would not recall, and so on and so forth with other hypotheses. The sentence is originally extracted from a movie script, and in this case the sentence immediately preceding it in the script is enough to drastically reduce the possible interpretations:

“I’ve spoken to the other children who were there that day. Can you think of anything else, Barbara, they might have told me about that party?” (4.8)

With this minimal context, it is now clear that the speaker was not at the party, but is asking Barbara to tell him or her something he or she already knows. To fully understand the utterances however, much more information on the interaction is needed: one must know that the utterances, extracted from the 1997 movie “The Devil’s Advocate”, are pronounced by a lawyer defending his client, a sexual abuser, while accusingly questioning Barbara, a victim of the abuser and witness in his trial.

This example illustrates what Scott-Phillips (2017) calls *strong* pragmatics. Contrary to *weak* pragmatics, which construes the context-dependence of meaning as a layer to be added on top of semantics, syntax, morphology, phonology and phonetics, strong pragmatics refers to the fact that all communication fundamentally depends on social cognition, which cuts through the other layers of linguistic analysis such as semantics and syntax. Indeed, what is communicated through the utterances discussed above, which can be rephrased as “tell me this thing I already know but that the audience does not”, could have been conveyed through an entirely different set of sentences (that is different semantics, syntax, morphology, and so on) because it depends above all on the social cognition situation that participants find themselves in.

Examples of this phenomenon abound, and are not restricted to face-to-face interactions as depicted by films: no matter the type of mediation, any interaction is likely to feature strong pragmatics. Twitter conversations are a good case in point for online platforms. The short conversation reproduced below, for instance, illustrates the fact that the meaning as understood by participants is a construction depending on context, past history, and interaction dynamics.³ It starts with the following tweet:

³The conversation is originally in French, and reads as follows: “On est tous le beau et le moche de quelqu’un” / “mais être moche c’est quand même la base ahah” / “[mort de rire,] pour certaines filles surtout, je pense”.

"We are all good-looking and ugly to someone else's eyes" (4.9)

This utterance seems a priori neutral, and is commonplace and consensual enough for it to be marked as favorite, retweeted and published anew regularly.⁴ But as illustrated by the answers following it, the actual meaning exchanged in the conversation is not available to the non-interacting reader. A first answer is made in a humorous tone:

"but we're still ugly in the first place haha" (4.10)

Then, two replies later, the conversation ends:

"[laughing out loud,] true for some girls especially, I would say" (4.11)

Even after five replies, we cannot determine whether the meaning exchanged is about sexism and rejection, or simply a flimsy joke without consequence; yet when taken as cultural tokens, these two representations are diametrically opposed to each other. With no further information about the relationship between the participants, their past interactions or common history, and in spite of the conversation being entirely public, we cannot determine what the exchange is fundamentally about, or even decide what the initial tweet means to one participant or the other.

Summary of problems

Let us now return to our initial question, namely the identification of attractors in the evolution of linguistic content. As might be clear by now, this goal is challenging in at least three new and related ways. First, the importance of strong pragmatics renders it much more difficult to collect all the necessary data to understand the meaning that a subject attributes to an utterance, or what is exchanged in an interaction. Indeed, it is often necessary to rely on detailed information about the interactive situation to understand that meaning. Leaving aside the question of the theoretical and technical apparatus that would be required to quantitatively analyse such data, the situations in which an experimenter can have access to the whole interactive situation, and thus have access to meanings exchanged (i.e. determine the content of the representations that circulate), are extremely rare. In most cases, an experiment only gives access to artefacts that are part of a broader cycle of meaning creation.

Second, even when the interactive situation is available to observation, the meaning of an utterance is not reducible to a simple object, and remains a multi-scale (and inside each scale, multi-dimensional) target. Coming back once again to utterance 4.1 with its surrounding context, what aspect of the meaning should one focus on when examining its evolution to identify attractors? The presence of a request to publicise private information, the implication that Barbara is lying or holding back such information, the lower-level structure of the question? In other words, the goal of identifying

⁴A simple search on Twitter using the original text in French shows that the utterance appears about once a month, with most instances retweeted several times.

attractors in the evolution of meaningful utterances is, at least in our current formulation, under-specified. This problem is not new, and might even not be a theoretical problem for Cultural Attraction Theory: behind the multi-dimensionality of meaning is the fact that culture itself is a multi-scale phenomenon (and multi-dimensional at each scale), difficult to characterise in simple mathematical terms. CAT works around this problem, as Sperber insists that it should not aim for a “grand unitary theory”, and should rather generate useful domain-specific questions that depend on the matter at hand (Sperber 1996, 61, 83). Thus the empirical decision of which meaning level to focus on must be resolved by appealing to the importance of each level as individually observed.

Finally, a more important theoretical challenge comes from the fact that strong pragmatics puts an important part of the responsibility for meaning, that is for the content of a linguistic representation, in the interactive situation itself. If the meaning of an utterance is determined in great part by the interaction it features in, and if that meaning corresponds to the content of the linguistic representations whose epidemiology we wish to study, then how is it possible to identify two representations from different situations as being the same (or being close to each other)? To make progress in the epidemiology of meaning-bearing utterances, an approach to strong pragmatics must thus be able to relate meanings that come from different interactive situations, to some extent at least. Indeed, evaluating the evolution of representations requires us to be able to identify, if not the path taken by specific strands of representations which inherit from each other, at least the overall trajectory of a population of representations in a common space. As a consequence, an approach to pragmatics useful to CAT must provide a way to declare meanings different, or identical, or evaluate the extent to which they differ, across situations (without which evolution can only be observed inside fixed interactive situations).

4.3 Approaches to meaning

If we broaden the scope to all interactions, whether face-to-face or digitally mediated, we are faced with the concrete question of how to understand the way an agent (participant, subject, person, or non-human organism) extracts or constructs meaning in such an interaction. That is, which of the infinite possible meanings the agent selects (or constructs), and how that selection (or construction) operates. As we just saw, such meaning is highly dependent on the context and interaction the agent finds itself in, such that viable approaches to meaning will necessarily be coping with the complexity of possible interactive situations. This makes the picture considerably more complicated than when dealing with simple context-free utterances.

4.3.1 Relevance theory

- fleshes out gricean communication with propositional meaning, and defines how meaning is selected
- relies on a “cognitive effect” which grounds relevance
- in theory this opens the door for interaction, although in practice there is no concrete way of doing it, as you need to infer the propositions, and context and interaction come back to kick you in the back
- it also poses a philosophical problem, which is the content of representations; that’s a huge debate and we’re not going there in detail

4.3.2 The enactive approach

(from article)

- this other approach breaks a bit more with representations, so doesn't have the philosophical problem, and also starts much lower-level: it's enactivism.
- C&S 2008: "the idea that thinking is computation allows one to see how abstractions (numbers, meanings) can be encoded in a mechanical system."
- comes in several flavours, with SM, 4E, REC
- this flavour gives you a different perspective/framework than representations, which gives a notion of relevance (that is, value to participants) straight up, but not how value can be transposed from another situation (that is, repetition/recognition, work done by the notion of representation)
- it's applicable at the very low-level, but needs some work to structure communication
- the two are not far from one another, as they provide a foundation and use of relevance to an organism (as CS2008 say, "the very same systems can profitably be explained dynamically and mechanically")

4.3.3 Outside the linguistic domain

(from article)

- going without representations is not limited to the linguistic domain, and CAT's reliance on it is grounds for critique by Ingold
- three layers of description in CAT
- three degrees of critique, relating to NCT/DST
- however, the question comes back as to whether you can compare different values emerging in particularity (as that work is done by representations)

4.4 Down to empirical study

ways to move forward

- further determine if they compete for the exact same space
- it's a productive contradiction to build from (without falling into scholasticism), which can inspire experiments

Chapter 5

Conclusion

References

- Scott-Phillips, Thomas C. 2017. 'Pragmatics and the Aims of Language Evolution'. *Psychonomic Bulletin & Review* 24 (1): 186–89. doi:[10.3758/s13423-016-1061-2](https://doi.org/10.3758/s13423-016-1061-2).
- Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford, UK; Cambridge, Mass.: Blackwell.