PhD Dissertation

# Epidemiology of Representations: An Empirical Approach

—original title may change—

Sébastien Lerique[1]

Supervisor: Jean-Pierre Nadal[2]
Co-supervisor: Camille Roth[3]

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

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

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

# Contents

# 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. Further modeling the process, we showed that utterance transformations can be clearly described as chunks of word insertions and deletions which loosely correspond to each other, interspersed with individual word replacements. We determined the prevalence of each of these operations, described their dependencies both between each other and with respect to the utterance they appear in, and generally argued that such an approach can serve as a middle-ground description between low-level accounts of lexical features and high-level accounts of the aggregated evolution of content along chains. However, none of these approaches has brought 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, as they are both crucial for further progress and closely linked to the ongoing debate about the nature of cognition and representations. To do so we will take a closer look at the problems involved in tackling the semantics, and more broadly the meaning, of utterances. 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. This limitation finds its root in what is known in philosophy of mind as the "hard problem of content", which we then expose more precisely. We continue to discuss two important approaches to studying the meaning of utterances:

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. We further show that the debate between these two approaches is closely linked to the overall construal of cultural evolution, as critiques of the cultural attraction framework have shown. Finally, we present possibilities for refining and advancing the debate through empirical investigation.

## 4.2 Empirical epidemiology of linguistic representations

### 4.2.1 Synthesis of 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 another), 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. The transformation process was shown to have several regularities: operations strongly depend on each other (in particular, insertions appear to make up for some of the deletions, while still introducing substantial change), and also depend on the length of, and their position in, the utterance. 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, and the overall evolution of the lexical makeup of utterances reflected those biases by drifting in a specific direction on each lexical feature (corresponding to better recall).

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 made of the context surrounding the quotations. The transmission chain experiments were led with an extremely (though intentionally, as a first step) simple read-and-rewrite task, which also did not open the analysis to the role of

context in transformations and 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 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.

what limits progress, from examples blending into theory

- examples of meaning change in Gistr, with interpretations depending on context
- example on twitter
- under-specified convergence problem
- divides into: content problem, context problem, dynamics problem

## 4.3 No theory of content

The problems of meaning: hard problem of content

- no clear definition of meaning (beyond word meaning)
- varies with history, situation, attention
- many levels
- obviously comes from continuous interactions, but bears value, and the value can also be recognised and repeated

### 4.3.1 Approaches to utterance meaning

expose the theoretical problem, and contenders: RT/Millikan and enactivism

- inside the representational approach, the problem is 1) the content of representations (Millikan), 2) how communication infers the content (RT)
- inside enactivism, you start with some content, but you need a way to arrive to structure communication and fleshing out the details

### 4.3.2 Outside the linguistic domain

show it's not limited to linguistics (ingold critique)

- three layers
- three degrees of critique, relating to NCT/DST

## 4.4 Baby and bathwater: From theory 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

# Chapter 6

# References