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. 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[1] 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 is complex to implement and must face some version of the hard problem of content in order to provide a complete account of meaning. The second starts from a simpler endogenous notion of meaning

[1] In what follows, we will always assume that meaning is meaning *to someone*. In other words, meaning to a listener is the listener's interpretation, and meaning to a speaker is the meaning intended to be communicated.

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.[2]

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)

[2]We 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.

The second part of this sentence is slightly misleading, and could be seen as a mild case of garden path sentence:[3] to "tell about" can be either transitive or intransitive, and while the final "that party" 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" $\rightarrow$ "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" $\rightarrow$ "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)

[3] A 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 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".

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").

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 then 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 questionning 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 abund, 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 construc-

tion depending on context, past history, and interaction dynamics.[4] It starts with the following tweet:

"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.[5] 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 humourous 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

[4] The 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".

[5] A 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.

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 attractors in the evolution of meaningful utterances is, at least in our current formulation, underspecified. 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 state 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 thus broaden the scope of empirical studies of CAT to all interactions (face-to-face or digitally mediated, but not necessarily linguistic, and in any case beyond interaction-less transmission chains), as will eventually be necessary for strong pragmatics, 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.

In this section, we present two prominent approaches to meaning and pragmatics, both of which can prove useful for further exploration of cultural evolution. The first, Relevance Theory, fleshes out the idea (first introduced by Grice 1989) that human communication is ostensive communication, based on the recognition of relevant communicative intentions. The second, the Enactive approach, starts from a more bare-bones level of description and proposes an understanding of how meaning emerges from the interaction of agents seen as dynamically coupled organisms. As we will see, both these theories provide (part of) an answer to how agents select, infer or construct subtly varied meanings in the course of an interaction, but they do so by starting from opposite ends. The first builds on a propositional notion of representations that are processed and combined in inference

processes, while the second starts from a representation-less description of organisms whose interaction and coupling with the environment endogenously generate (non-representational) meaning. The notions of meaning to which they arrive are quite disjoint, and have historically been considered in contradiction; indeed, we will then show how these differences can be grounds for a critique of CAT and other Darwin-inspired cultural evolution approaches. In spite of this, we will argue that both approaches to meaning could be productive guides for generating empirical questions and experiments regarding cultural evolution.

4.3.1 Relevance theory

Principles

As a general theory of communication, Relevance Theory has a very broad scope and relates to many areas of cognitive science. Sperber and Wilson (1995) and Wilson and Sperber (2002) provide detailed presentations of the full theory, and many publications in between and since then have fleshed out its relations to a number of neighbouring questions. Wilson and Sperber (2012), in particular, provides a thorough discussion of several linguistic phenomena based on Relevance Theory, as well as openings towards experimental and cultural evolution-related approaches to the question of meaning and relevance (for language evolution see in particular Sperber and Origgi 2012). Here we will restrict ourselves to a sorely abridged presentation of the already summarising Wilson and Sperber (2004), in the hopes that it will be enough for an approximate understanding of the principles underlying the theory and the explanations it provides.[6]

Relevance Theory (RT) opposes itself to the code model of linguistic communication, according to which a speaker's meaning is encoded in an utterance, passed on to the listener for instance by means of sound (the channel, or conduit), and then decoded by the listener to obtain the communicated meaning. By contrast, RT adopts an inferential model according to which an utterance does not encode a meaning per se, as the semantics of utterances provide only under-determined information about the speaker's meaning (as illustrated by the examples discussed above); instead, the inferential model considers that utterances provide evidence (and exactly the right amount of evidence) for the intended meaning to be inferred given the situational context. This model of communication was first elaborated by Grice, building on the fact that people who are communicating usually assume that what the other person is saying is meant to be understood given the context at hand; in other words, people take their interlocutors to be neither stupid nor adversarial, and assume (consciously or not) that what a speaker says is a signal for a meaning that the listener should be able to understand, through inference. Grice thus identified four general rules (maxims) that listeners generally assume their interlocutor will follow, and on which they rely to infer meaning: Quality (truthfulness), Quantity (informativeness), Relation (relevance), and Manner (clarity). RT agrees with the intuition behind Grice's observations (although it differs on exactly which listener expectations should be necessary), and fleshes out this general inferential model of communication in cognitively plausible terms.

RT proposes that inferential communication is based on a cost-reward comparison of possible conclusions that derive from a speaker's utterance. Indeed, a given utterance (or non-linguistic communicative act) in a given context can lead a listener to any number of conclusions about the world.

[6] Our presentation focuses on the founding principles of Relevance Theory. The remainder of Wilson and Sperber (2004) fleshes out the way the inferential procedure is applied to linguistic utterances, how the theory explains typical phenomena such as loose uses of language (e.g. the meaning of 'square' in expressions such as 'square face' or 'square mind'), irony, or poetry, and how it fits with the massive modularity of mind approach introduced in Sperber (1996), along with many detailed examples.

Each of those conclusions about the world can matter more or less to the listener (RT formulates this as the strength of the contextual effects created by the conclusion), and is also more or less costly to derive from the speaker's utterance and its context (processing cost in RT terminology). A conclusion that matters more to the listener achieves higher relevance, and conversely a higher processing cost will lower the relevance realised by a conclusion. These two dimensions let listeners order the conclusions that can be derived from a speaker's utterance based on their (context- and listener-dependent) relevance. For instance, hearing that Sperber or Wilson's train to work is one minute late is less relevant to them (because it matters less) than hearing that their train is late by a half hour. Conversely, a public announcement stating that their train is late provides a more relevant conclusion (because easier to derive) than the same conclusion derived through more deductive effort from bits of a conversation overheard between the people sitting next to them. A central claim of RT is that evolution has shaped human cognition in such a way that people automatically and easily perform this derivation and comparison process on all the stimuli they perceive, picking out those among the myriad available which maximise relevance. The Cognitive Principle of Relevance expresses this claim: "Human cognition tends to be geared to the maximization of relevance" (Wilson and Sperber 2004, 610).

Wilson and Sperber (2004) then define inferential communication as consisting of two elements. An *informative intention*, that is the intention of a speaker to inform an audience of something (more precisely, to make certain assumptions more, or less, manifest to the listener), and a *communicative intention*, that is the speaker's intention to inform the audience of their informative intention. In other words, inferential communication happens whenever the speaker says (or does) something in order to make her audience recognise that she wants to convey X. The meaning is successfully understood when the audience recognises the speaker's informative intention, that is when the audience recognises that the speaker wants to convey X (note that X itself might not be conveyed if the audience does not trust the speaker – the communication event is nonetheless successful, since the intention to convey X was recognised). Most often, the speaker does this by making an ostensive communication act (e.g. pointing, staring, or saying something that attracts the audience's attention) which signals to the audience that there is something worth processing to attend to. Indeed, ostensive stimuli create in the audience an automatic expectation for relevance, as the audience looks for the reason for which the speaker is attracting their attention. More precisely, RT posits that the audience automatically expects the stimulus to be *optimally* relevant; in the theory's terminology, this is formulated as the Communicative Principle of Relevance: "Every ostensive stimulus conveys a presumption of its own optimal relevance." This principle is the basis on which the audience's inferential process works: the speaker's ostensive stimulus signals something worth processing to reach a relevant conclusion (since she attracted their attention to process it), and it is also the stimulus that makes that relevant conclusion the easiest one to reach.

The authors discuss an example to illustrate this: we are at a table and my glass is empty, a fact that you might notice. If you do (without me communicating anything), one conclusion you could reach is that I might like a drink. If, however, I wave my glass at you, or say "My glass is empty" (ostensive stimulus attracting your attention to my empty glass), a relevant conclusion you would reach is that I want to communicate that I want a drink (and, if you trust me, conclude that I want a drink, although that is not necessary for the communication to succeed). RT thus proposes a procedure that can account for the way utterances are understood: when perceiving a stimulus (possibly ostensive), and given a certain context, compute the relevance of its conclusions (i.e. the strength of contextual effects pitted against the processing costs) following a path of least effort first (since the stimulus is expected to be optimal), and stop whenever you have reached your expected level of relevance. In other words, test hypotheses about the speaker's utterance such as possible disambiguations, resolution of entities and implicatures, and stop whenever the conclusions you

have reached seem relevant enough to you. The conclusion you have then obtained will be your assumption of the speaker's meaning for the context chosen at the beginning. Finally, note that this inference procedure can be operated in different possible contexts, as long as they are available given the memory constraints of the listener. The procedure thus also optimises on the context in which conclusions are drawn, and selects the context for which the final conclusion is most relevant.

Application

The framework provided by Relevance Theory is extremely rich, and has been the subject of extensive experimental exploration and validation (see Noveck and Sperber 2012; and Henst and Sperber 2012 for reviews). In particular, it brings direct insight to some of the limitations concerning meaning in the experimental approach of the previous chapter. The transmission chains we set up are a clear case of ostensive communication: we ostentatively ask the subjects to direct their attention to the utterances they are asked to memorise and rewrite. However, the utterances are presented with no context other than the task itself, which frames the experiment as a memory exercise. There is no background information against which the subjects can evaluate the relevance of conclusions derived from the utterances, nor are the subjects involved in an activity that would make the conclusions matter in one way or another. Without an ecological activity to which the utterances can become relevant, the experiment has no control over the meanings that subjects will infer from the utterances, leaving the matter entirely under-specified. It is also easier to understand why subjects spontaneously wrote sentences directed to the experimenter such as "I can't remember": if the task does not create an ecological communicative activity that relates subjects to each other, and is instead (correctly) understood to be a memory task, the only valid interlocutor is the experimenter evaluating the accuracy. As a consequence, asking the subjects to keep in mind that their productions were sent to other subjects (as we did) was likely to create a slight discrepancy.

The relevance-theoretic approach also opens the door for the analysis of interaction, context, or past history in the evolution of meanings. Using carefully constructed contexts to orient what is available to the inferential process, it should be possible to greatly reduce the possible meanings interpreted by subjects, and thus explore the way interpretations evolve through chains of contextually-augmented utterances (or chains of constrained interactions). In practive however, such an implementation is likely to be extremely challenging (much more so than the procedure developed in the previous chapter): whatever the theory, it is still necessary to extract the basic propositions semantically encoded in the utterances typed in by subjects, determine the way subjects select contexts, and automatically or manually derive the possible conclusions that can be reached for a given utterance in a given context. The three tasks are far from trivial. Second, efficiently constraining the context in which an utterance is interpreted will likely be much more difficult than it sounds, as real life interpretation involves our own personal history, memory, preoccupations and any other pregnant contexts we can recruit during the process (see Sperber and Wilson 1995, secs 3.3–4, which discusses the ways contexts are chosen in the inferential process). It is doubtful that simply adding a few sentences around the target utterance would suffice. Instead, it could be necessary to create more encompassing situations such as a controlled video game where much larger parts of the many contexts available to subjects can be experimentally manipulated.

Meaning as indexed on knowledge optimisation

We have just seen that the notion of meaning provided by Relevance Theory is based on a maximisation of the relevance of conclusions derived by the listener. This account ultimately rests on the

three following cognitive mechanisms:

- Reconstruction of the logical form of an utterance, in order to start the inferential process (see Sperber and Wilson 1995, sec. 4.3, for details),
- Creation or selection of contexts inside which the inferential process operates (see Sperber and Wilson 1995, secs 3.3–4, for details),
- The inferential process itself, operated by what is hypothesised to be a special-purpose deductive device (see Sperber and Wilson 1995, secs 2.4–5, for details).

This account can also be formulated in terms of a system which optimises its representation of the world (without access to the truthfulness of what it perceives), and for which relevance indicates the path of strongest optimisation growth. To see this we must briefly return to the exact definition of relevance in the theory. Relevance in Sperber and Wilson (1995) is defined in the following three broad steps:

1. Define the *deductive device* that is used for inference: it is a mechanism for deriving conclusions from a set of premises P (usually coming from an utterance from a speaker) in a set of contextual assumptions C (Sperber and Wilson 1995, secs 2.4–5); in this device, all assumptions, premises and conclusions have a certain strength, corresponding to their level of accessibility (this is not a logical measure of confidence that can be quantified, but rather an ordinal property that can be used to compare assumptions between each other).
2. Define the notion of *contextual effect*: a set of premises P (coming from an utterance) in a set of contextual assumptions C generates a contextual effect if and only if the deductive device can derive conclusions from the combination of P and C that it cannot derive from P or C alone. Such conclusions can be new to C, can strengthen existing assumptions in C, or can weaken and even erase assumptions in C (Sperber and Wilson 1995, secs 2.6–7). The notion of contextual effect thus provides a indication of the strength of the relationship between an utterance and a set of contextual assumptions C.
3. Define the degree of *relevance* realised by conclusions derived by the deductive device from P and C: conclusions are more relevant if they have stronger contextual effects, and less relevant if they have higher processing costs (in terms of deductive steps involved).

With relevance thus defined, Relevance Theory's procedure for interpreting utterances and other stimuli can be seen as a tentative procedure to always improve the system's representation of the world: if the processing system has no access to the truthfulness of the stimuli and utterances it perceives, its best option is to process the incoming information in a way that maximises what it can infer about it, given its level of trust in the speaker. In other words, it will look for the set of assumptions that can most benefit from the new information. If the system functions with a deductive device such as the one defined above, this corresponds to finding the set of contextual assumptions C on which the new information (premises P) has the strongest contextual effect, with given processing cost constraints and for a given level of trust in the speaker (i.e., strength of the premises P); given the above definition of relevance, that is precisely the procedure to reach the most relevant conclusions. Under this description, higher relevance indicates a stronger update in contextual assumptions, thus a stronger update to the system's overall representation of the world. Thus Relevance Theory can be seen as indexing meaning on the increase in reliability of a listener's representation of the world. The Communicative Principle of Relevance then simply states that speakers know that listeners work by maximising inferred relevance, and will behave accordingly in order to be understood (i.e. by saying things which they know will trigger the right inferences for the listeners).

Whichever the formulation we choose, RT crucially relies on a deductive device which can represent

propositions extracted from utterances and process such propositions in order to derive new conclusions. Once the inference process is completed however, the meaning defined by RT is partially propositional, as it is made of a set of new assumptions that are made more less manifest (through a change of strength). Such meanings can thus be at least partially defined and used without reference to the situation they were deduced from, making them (partially) comparable to other meanings from other situations. The approach to pragmatics developed by RT is thus quite amenable to the study of cultural evolution (which is no surprise, given the authors).

One disadvantage of the relevance-theoretic approach is the high level at which it starts its description, making concrete implementations more challenging. As we indicated earlier, it relies on the reconstruction of the logical form of utterances, on a mechanism for the creation of contexts, and on the inferential process itself. Second, a theoretically important, but experimentally less important consequence of the theory's reliance on propositions and mental representations is that a full account of meaning will eventually require an account of the content of such representations (a problem known in philosophy of mind as the hard problem of content, Hutto and Myin 2013). We will now turn to a second approach to meaning, one that does not rely on propositions or even representations, and starts from a much lower level of description. Our hope is to convince the reader that this second approach, although less connected to the standard body of linguistic research, should also be a fruitful avenue to explore the effects of interaction in relation to meaning in cultural evolution.

4.3.2 The enactive approach

A non-computational metaphor of cognition

The enactive approach proposes a different foundational metaphor for the study of cognition. Indeed, Relevance Theory and Cultural Attraction Theory, along with our own experimental approach to their questions, mostly rely on a computational metaphor for the description of cognition: the mind is taken to be like a computer, that is an information-processing device which continuously receives stimuli, updates an internal representation of the world based on what it perceives, and acts based on its current representation and predictions given current stimuli. The human brain is our implementation of such an information-processing device. The approach we are interested in here is part of a range of approaches that question the utility of conceiving the mind as such an information-procssing system,[7] and explore the extent to which parts of (or all) the metaphor can be relaxed or replaced by other paradigms (see Chemero and Silberstein 2008 for a review of the options available in the debate, and the fields corresponding to each choice). In this area, the enactive approach has the advantage of being extremely consistent in its rejection of the computational paradigm and of the idea that cognitive systems represent their environment, and is to our knowledge the only contender that has started developing a non-computational approach to language itself. As we will see, instead of a computational paradigm the enactive approach proposes to base cognition on the dynamical coupling of organisms with their environment and with each other. Compared to Relevance Theory it can be located at the opposite end of the computational spectrum and starts from a different initial level of description, but is nonetheless concerned with questions common with RT. In particular, both aim to reach complete and plausible explanations of language and meaning. As Chemero and Silberstein (2008) argue however, the two are not necessarily opposed, and could be usefully combined to form complementary explanations. Our goal here is to present the basic tenets of the enactive approach and show how, by starting with a different metaphor, it faces an

[7] These approaches are sometimes collectively termed the "E turn", in reference to the many titles starting with the letter "e" (in particular, enactive, embedded, embodied and extended approaches to cognition).

orthogonal set of problems compared to RT. In particular, the notion of meaning it develops seems more endogenous than that of RT (among other things, by being non-representational, it does not face the hard problem of content), but is currently much more low-level and not yet usable to fully comprehend actual linguistic interactions. What the theory currently provides can be seen as the explanation of preliminary steps common to language and less structured interaction, eventually to grow into a full theory of linguistic interactions (or, in enactive terms, "enlanguaged" interactions).

The first concrete articulation of this approach in cognitive science is usually attributed to F. Varela, Thompson, and Rosch (1991) who develop a view of cognition based on Merleau-Ponty's phenomenology. They propose to look at mind, cognition and meaning as fundamentally embodied and situated processes in which self-organisation plays a central role. Of course, nobody basing themselves on the computational paradigm would deny that what they talk about is ultimately grounded in physical embodied things; however the specificity of the enactive approach (and of other non-representational approaches with it) is that its explanations draw deeply on the embodiedness and situatedness of the processes, in that they create notions of cognition and meaning defined in terms of the coupling of physical systems, rather than in terms of symbolic processing. The initial formulation by F. Varela, Thompson, and Rosch (1991) led to many developments. We present here the main theory going under the name "enactive approach", and do so in four important conceptual stages which we believe roughly (though drastically) summarise what has been developed by Torrance (2006), Thompson (2007), De Jaegher and Di Paolo (2007), and Cuffari, Di Paolo, and De Jaegher (2015). While this will by no means do justice to the complete approach, we hope these stages will provide a clear-enough stetch of the dynamical and embodied account of cognition that the enactive approach develops and proposes to use instead of the computational metaphor of mind.

Sensorimotor contingencies

The first stage is a reconceptualisation of the way an organism perceives its environment. This conceptualisation, known as the sensorimotor approach to perception (and thoroughly developed for vision by O'Regan and Noë 2001), essentially takes perception to be an exploratory activity based on a continuous perception-action loop. The default approach to perception is to construe it as an inference problem: through its senses, an organism receives information about the world and attempts to reconstruct an internal representation of it, which is challenging because the information is degraded in a number of ways. Instead, the sensorimotor approach construes perception as the exploration of the regularities in the way stimulations change when the organism moves around or acts on its environment (or on an object). Thus, rather than inferring and internally representing the properties and shape of an object that is being perceived (for instance), the sensorimotor approach construes an organism as exploring the changes it generates in the sensory stimulations when moving. thus making perception and action two parts of a common loop. As O'Regan and Noë (2001) put it: "seeing constitutes the ability to actively modify sensory impressions in certain law-obeying ways." An extreme example of such actively perceived properties is the softness of a sponge, which is felt by prodding and sequeezing it but not through static contact (Myin 2003).

One of the strong motivations for this approach is that it provides an endogenous account of the feel of a perceptual modality (i.e. its perceptual consciousness), a longstanding problem in inferential approaches to perception. According to the sensorimotor approach, seeing and hearing feel differently (i.e. one can easily differentiate visual from auditory consciousness) not because they are processed by different parts of the brain, but because of the specific regularities with which stimulations are deformed in each sensory modality when the organism moves. Turning our head, for instance, generates a certain change of stimulation in vision, and a different change in hearing. The way each

modality sees its stimulation change with movement is referred to as its *sensorimotor contingencies*, and is directly tied to the type of perceptual consciousness the modality creates.

An interesting confirmation of this approach is found is experiments using "Tactile Visual Substitution systems", where blind people are equipped with a device that reproduces on their skin (through an array of stimulators) the luminance patterns captured by a camera. The subjects are then tested on their ability to recognise objects using this cutaneous stimulation, and are only able to do so if they actively control the movements of the camera itself (O'Regan and Noë 2001, 958). Furthermore, once they do control the camera, their sensations seem relatively close to actually seeing, because the sensorimotor contingencies are so similar: they begin to perceive objects as not on their skin but in front of them (in particular, they can be frightened by a zooming effect in the stimulations, which corresponds to an object approaching very fast), and the location of the stimulator array on the body becomes unimportant (subjects can easily transfer from stimulation on the back to on the forehead). Such experiments have contributed to showing that perception and action are two sides of the same dynamical interaction loop with the environment, and by generalising to other modalities, they suggest that sensorimotor contingencies provide an endogenous account of perceptual consciousness.

Sense-making

The second stage extends this approach to life itself (here we follow De Jaegher and Di Paolo 2007; and Thompson 2007). In a nutshell, it can be seen as taking the reconceptualisation operated by the sensorimotor approach, which goes from a notion of perceptual consciousness based on inference to a notion made of sensorimotor contingencies arising in perception-action loops, and applying it to meaning in cognition: instead of being seen as the result of an inferential process, meaning will be seen as a property (or a regularity) of the dynamical interaction of an organism with its environment.

Let us make this step more precise. Inspired by the notion of autopoiesis developed by Maturana and Varela (1980), the enactive approach considers a living organism as an *autonomous system*, that is a network of processes with the following properties:

1. The system is self-produced and self-maintained. As a consequence the processes depend on each other for continued operation, that is, every process in the network is conditioned on the activity of one or several other processes of the network (a property called *operational closure*). As a consequence, the network of processes acquires an identity (defined by its operational closure).
2. The system continually produces a boundary that distinguishes it from the environment (this need not be a physical boundary).
3. The system actively regulates its interaction with the environment in order to maintain its identity.

Crucially, the identity generated by operational closure is precarious: it disappears if some or all of the processes that make up the system cease. The system is thus in a permanent tension to regenerate the conditions for the continuation of its identity, and any interaction with the environment thus acquires an inherent value to the system since it can have positive or negative consequences on the continuity of the system's identity and autonomy. Since interactions with the environment are necessary for the network of processes to keep self-generating, the system is continuously regulating the strength of its coupling with the environment in order to maintain its identity. Interaction with the environment then becomes inherently meaningful to the system, such that enactive approach calls it "sense-making".

In this framework, cognition *is* precisely the sense-making activity, that is a system's actively regulating its coupling to the environment in order to maintain its identity. Notice how the enactive notion of meaning is defined in a parallel manner to the sensorimotor account of perceptual consciousness: instead of being inferred and represented, it is a property of the dynamics of the system's interaction with its environment.

Participatory sense-making

The third stage extends the theory to interaction between two autonomous systems, and introduces the notion of an autonomy of the interaction itself. De Jaegher and Di Paolo (2007) develop this in two steps. First, they show that some interactions can only be explained at the level of the interaction itself, rather than at the level of participants. An interesting point in that direction has been made in an experiment by Auvray, Lenay, and Stewart (2009). The experiment involves two subjects who share a virtual line on which they each have a cursor. The line and cursors are all invisible, but the subjects receive haptic feedback whenever their cursor is overlapping with the other's cursor. Aside from the subject's cursors, two fixed obstacles are placed on the line (each is perceivable by one subject and not the other), and the cursor of each subject has a shadow that follows it at a fixed distance: when subject A's cursor touches the shadow of subject B's cursor, subject A receives haptic feedback but subject B does not (and vice-versa). The subjects are told about obstacles but not about shadows, and are tasked with clicking as much as possible on each other's cursors. Interestingly, they succeed in doing so, but not because they are able to distinguish between real cursor and shadow. The experiment shows instead that they are not able to make the distinction individually, but solve the task because the interaction of real cursors is more stable (and thus more frequent) than the an interaction with a shadow: subjects individually fail the task while succeeding collectively, in a way that can only by understood because of the inherent (and unnoticed) stability of their interaction. The principle highlighted by this experiment is that of the stability of *perceptual crossings*: two organisms can have a dynamically stable interaction because they each look for a behaviour that they themselves create, without necessarily being aware of that fact (for instance mutual gaze of an infant and his mother, where the infant may not be aware that his mother maintains the gaze because he does too).

Second, De Jaegher and Di Paolo (2007) argue that such stable interactions can acquire an autonomy of their own. An example that most people have experienced in everyday life usefully illustrates their point: when trying to cross someone else in the corridor of a train, and moving to the side to avoid them, at times the other person spontaneously moves to the same side you did; when this happens, you and the other person enter an interaction which both are trying to break from the start: each one moves to one side, and the other does the same, until your movements desynchronise and the interaction breaks down. During the time it persisted however, the interaction acquired its own autonomy which constrained both you and the other person, as neither could break free from it.

This autonomy serves as the basis for defining sense-making at the level of the interaction itself: when two organisms interact while at the same time regulating their coupling to their environment (and respecting each other's autonomy), the interaction itself can spontaneously self-organise and become self-sustaining. Similarly to organisms, then, it acquires an identity of its own, and an interest in maintaining that identity: in that case, since the couplings of each organism to their environment and with each other have an impact on the continuation of the interaction, they become meaningful *to the interaction* which can then partly regulate them. A new sense-making activity thus appears at the level of the interaction itself, a level that neither of the participants fully control, and which has the potential to create constraints on them. This notion is termed *participatory sense-making* (De Jaegher and Di Paolo 2007).

Languaging

The fourth and final stage brings us to language. Relying on the concepts defined above, Cuffari, Di Paolo, and De Jaegher (2015) propose to see language as a specially structured pattern of participatory sense-making, governed by several levels of conventions interlocked with one another.

- C&S 2008: "the idea that thinking is computation allows one to see haw abstractions (numbers, meanings) can be encodoed in a mechanical system."
- 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
- explore how much more context constrains the evolution of utterances

Chapter 5

Conclusion

References

Auvray, Malika, Charles Lenay, and John Stewart. 2009. 'Perceptual Interactions in a Minimalist Virtual Environment'. *New Ideas in Psychology* 27 (1): 32–47. doi:10.1016/j.newideapsych.2007.12.002.

Chemero, Tony, and Michael Silberstein. 2008. 'After the Philosophy of Mind: Replacing Scholasticism with Science'. *Philosophy of Science* 75 (1): 1–27.

Cuffari, Elena Clare, Ezequiel Di Paolo, and Hanne De Jaegher. 2015. 'From Participatory Sense-Making to Language: There and Back Again'. *Phenomenology and the Cognitive Sciences* 14 (4): 1089–1125. doi:10.1007/s11097-014-9404-9.

De Jaegher, Hanne, and Ezequiel Di Paolo. 2007. 'Participatory Sense-Making'. *Phenomenology and the Cognitive Sciences* 6 (4): 485–507. doi:10.1007/s11097-007-9076-9.

Grice, Herbert Paul. 1989. *Studies in the Way of Words*. Cambridge, MA, USA: Harvard University Press.

Henst, Jean-Baptiste van der, and Dan Sperber. 2012. 'Testing the Cognitive and Communicative Principles of Relevance'. In *Meaning and Relevance*, edited by Deirdre Wilson and Dan Sperber, 279–306. Cambridge, UK: Cambridge University Press.

Hutto, Daniel D, and Erik Myin. 2013. *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, Mass.: MIT Press. http://public.eblib.com/choice/publicfullrecord.aspx?p=3339551.

Maturana, Humberto R, and Francisco J Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht, Holland; Boston: D. Reidel Pub. Co.

Myin, Erik. 2003. 'An Account of Color Without a Subject?' *Behavioral and Brain Sciences* 26 (1): 42–43.

Noveck, Ira, and Dan Sperber. 2012. 'The Why and How of Experimental Pragmatics: The Case of "Scalar Inferences"'. In *Meaning and Relevance*, edited by Deirdre Wilson and Dan Sperber, 307–30. Cambridge, UK: Cambridge University Press.

O'Regan, J. Kevin, and Alva Noë. 2001. 'A Sensorimotor Account of Vision and Visual Consciousness'. *Behavioral and Brain Sciences* 24 (5): 939–73. doi:10.1017/S0140525X01000115.

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.

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

Sperber, Dan, and Gloria Origgi. 2012. 'A Pragmatic Perspective on the Evolution of Language'. In

Meaning and Relevance, edited by Deirdre Wilson and Dan Sperber, 331–38. Cambridge, UK: Cambridge University Press.

Sperber, Dan, and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Oxford UK & Cambridge USA: Blackwell.

Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Torrance, Steve. 2006. 'In Search of the Enactive: Introduction to Special Issue on Enactive Experience'. *Phenomenology and the Cognitive Sciences* 4 (4): 357–68. doi:10.1007/s11097-005-9004-9.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Wilson, Deirdre, and Dan Sperber. 2002. 'Truthfulness and Relevance'. *Mind* 111 (443): 583–632. doi:10.1093/mind/111.443.583.

———. 2004. 'Relevance Theory'. In *The Handbook of Pragmatics*, edited by Laurence R Horn and Gregory Ward, 607–32. Oxford, UK: Blackwell.

———. 2012. *Meaning and Relevance*. Cambridge, UK: Cambridge University Press.