READING MINDS

A COGNITIVE-DIGITAL APPROACH TO THE STUDY OF LITERATURE

A Dissertation
Submitted in Partial Fulfilment of the Requirements for the Degree of

Doktor der Philosophie

to the Fachbereich Philosophie und Geisteswissenschaften
of Freie Universität Berlin

by

Martín Azar

Berlin, 2019

Supervisor: Prof. Dr. Joachim Küpper

Second examiner: Prof. Dr. Dr. Friedemann Pulvermüller

Date of defense: 28th Feb 2019

STATETEMENT OF AUTHORSHIP / SELBSTSTÄNDIGKEITSERKLÄRUNG

§ 7 Abs. 4 Promotionsordnung / Doctorate Regulations

Hiermit versichere ich, alle Hilfsmittel und Hilfen angegeben und die Arbeit auf dieser Grundlage selbständig verfasst zu haben. Die Arbeit ist nicht schon einmal in einem früheren Promotionsverfahren angenommen oder abgelehnt worden.

I hereby confirm that I have indicated all resources and aids and that I have independently written the thesis on this basis. The dissertation has not been accepted or rejected in an earlier doctoral procedure.

(Datum/Date) (Unterschrift/Signature)

SUMMARY / ZUSAMMENFASSUNG

ENGLISH VERSION

This dissertation represents my doctoral research at the Friedrich Schlegel Graduate School of Literary Studies (FU Berlin), between the years 2015 and 2018. Prof. Dr. Joachim Küpper and Prof. Dr. Dr. Friedemann Pulvermüller were the supervisors of this project. The main thesis of this work consists in presenting a theoretical framework combining cognitive science and data science that for studying literature in a naturalistic and empirical (evidence-based) way.

The book is structured in three parts. The first part describes the mental faculties that allow humans to process literature, as well as many of the cognitive biases that intervene in this process. The second part of this book is devoted to the exploration of different digital tools for analyzing language and literature in quantitative ways. The third one is a case study meant to illustrate the potential of integrating cognitive science and data science as a framework for studying literature. This case study consists of three steps: (1) a text-mining analysis of a digital register of reader-responses to a video that included the text, (2) a formulation of informed hypotheses of likely reader-responses, (3) a verification of the predictive accuracy of these hypotheses through an open survey.

The general goal of introducing this framework is to collaborate with the project of bridging the sciences and the humanities in a productive and mutually enlightening dialogue.

DEUTSCHE VERSION

Diese Dissertation ist das Ergebnis der wissenschaftlichen Studien meiner Doktorarbeit, die von 2015 bis 2018 an der Friedrich Schlegel Graduiertenschule für literaturwissenschaftliche Studien (FU Berlin) durchgeführt wurden. Prof. Dr. Joachim Küpper und Prof. Dr. Dr. Friedemann Pulvermüller waren die Betreuer dieses Projekts. Die Hauptthese dieser Arbeit besteht aus einer empirischen Methode.

Die Thesis ist in drei Teile gegliedert. Der erste Teil beschäftigt sich mit den geistigen Fähigkeiten, die es dem Menschen erlauben, Literatur zu verarbeiten, sowie mit den kognitiven Verzerrungen, die in diesen Prozess eingreifen. Der zweite Teil widmet sich verschiedenen digitalen Werkzeugen für die Analyse von Sprache und Literatur in quantitativer Hinsicht. Der Dritte ist eine Fallstudie, die das Potenzial der Integration von Kognitionswissenschaft und Datenwissenschaft als Rahmenwerk für Studienliteratur aufzeigt. Diese Fallstudie besteht aus drei Schritten: (1) eine Text-Mining-Analyse eines digitalen Registers der Leserantworten auf ein Video, das den Text enthält; (2) eine Formulierung informierter Hypothesen über wahrscheinliche Leserreaktionen; (3) ein Verifizierung der Vorhersagegenauigkeit dieser Hypothesen durch eine offene Umfrage.

Das übergeordnete Ziel dieses Rahmenwerks ist, interdisziplinär Natur- und Geisteswissenschaften in einen produktiven und gegenseitig erhellenden Dialog zusammenzubringen.

TABLE OF CONTENTS

Statement of authorship 3

Summary 5

Prologue 9

Part I

COGNITIVE SCIENCE
The predictable biases of the reading mind

1 **LANGUAGE AS A NATURAL PHENOMENON** 19
From atoms to poetry

2 **WORDS** 38
How we encode and decode meaning

3 **RHETORIC** 57
How phrasing affects our interpretations and behavior

4 **STORY** 74
How we process narrative

5 **CHARACTERS** 90
How we judge others through language

6 **EMOTIONS** 105
How to move people with words

Part II

DATA SCIENCE
Your computer knows you better than you know yourself

7 **BIG DATA AND PSYCHOMETRICS** 123
Digital lens for reading minds at large scales

8 **QUANTITATIVE LITERARY STUDIES** 129
How to read a thousand books in a second

9 **TEXT MINING** 136
The power of computers for finding patterns

10 **ROBOTS THAT READ AND WRITE** 150
Frontiers of language technology

Part III

EVERY BREATH YOU TAKE

Cognitive-digital literary studies in practice

11 **CASE STUDY – FIRST PART** 173
Measurements and hypotheses

12 **CASE STUDY – SECOND PART** 184
Verification and discussion

Afterword 205

Acknowledgements 209

Bibliography 213

Appendix 1 - Survey Questions 229

Appendix 2 - Survey Results 235

PROLOGUE

MINDS THAT READ

Are we really aware of how language affects us? Let me illustrate this problem with a thought experiment. Which of the following dishes do you think would taste better?

> Option A. *A plate of baked potatoes with salt and pepper.*
>
> Option B. *A flavorful plate of rustic and earthy finely-baked crispy potatoes, harvested by self-managed workers, splashed with sparkles of Himalayan pink salt and fresh aromatic ground peppercorn.*

We tend to think that what ultimately determines our taste experience is not the description of food but the food in itself. So, even though those descriptions differ, if they would refer to identical dishes, then the dishes would still *taste* the same. Right? Not really.

Many psychological experiments have proven that, when people are offered the exact same dishes but described in different ways, they tend to chose more often the dishes presented with longer and more lyrical descriptions (Wansink et al., 2001), and they also find them actually *tastier* (Ibid., 2005). Adjectives, adverbs, and other modifiers can really affect our very perceptions, judgments, and experiences. And they do not do it randomly, but in systematic ways that –to a certain extent– can be measured, predicted, and explained, by considering *how our minds work* –paraphrasing Steven Pinker (1997).

In the case of the potato dishes, the evidence suggests that most people would prefer and enjoy Option B the most. Here are some of the reasons: Reading language that describes sensorial experiences (*flavorful*, *crispy*, *fresh*, *aromatic*) activates the sensory cortex of our brains, and makes us mentally simulate those sensations in our minds and expect them (Lakoff&Johnson, 1999) –which is why our mouths water when reading menus (Spence, 2011). Using seemingly technical terms to describe taste, such as *rustic* and *earthy*, gives us the idea of a complex expertise behind the elaboration of the dish, which predisposes us in advance to pay more attention to nuances in the flavors and to enjoy them more –as it has been extensively shown in wine-tasting experiments, for instance (Goldstein et al., 2008; Kringelbach, 2005). Linking a product with an exotic place (*Himalaya salt*) makes us feel it as more singular, and people consistently show readiness to pay more for so-described products (Ariely&Kreisler, 2017). Even mentioning generic agents involved in the chain of production of the food seems to make a difference: talking about *self-managed workers* suggests human stories behind the dish. Framing an object within a story makes us feel it as more meaningful, valuable, and even moral (s. Walker, 2009; Haidt, 2012; Ariely&Kresiler, 2017). In short, our brains start to taste and appreciate the food already when *reading* about it.

These and many other effects of language have been repeatedly verified in controlled psychological experiments. Nevertheless, when people are later asked in these experiments to explain why they enjoyed so much the chosen product (in this hypothetical case, the potatoes), they never say: "Because they were seasoned with the perfect words." Instead, they answer things like: "I liked these ones more because they were very tasty and well cooked" –even if both dishes were identical (s. Mladinow, 2012). This evidences that the ways in which words affect us are not transparent to ourselves by simple introspection. We often ignore the causes of our feelings, interpretations, judgments, and decisions (Ariely, 2008).

Understanding the biological systems, senses, instincts, and biases that make us process language and respond to it in certain ways is not only pertinent for making dishes more appealing, but also for every cultural field where language is used: from politics, economy, and law, to journalism, religion, education, and, of course, literature. The title of this book (*Reading Minds)* can be interpreted in two ways, because it is about two things: it is a book about minds that read but also about how to read them.

TRYING TO READ MINDS THAT READ

We, humans, are quite good at interpreting language. The universally extended popularity of literature is an evidence of this: there is no culture without stories, idioms, poetry, and jokes. And all of us are –to a higher or lower degree– extensively trained for making sense of these verbal creations, for interpreting them, for extracting meaning from them. But what about a slightly different task: How good are we at predicting how another people might interpret a given text? This is called mindreading.

Cognitive scientists have discovered that we are all intuitive mindreaders: we all have an instinct to imagine the mental states of others. Different names have been given to this faculty: *mindreading* (Nichols&Stich, 2003), *theory of mind* (Frith&Frith, 2005), or *mentalization* (Allen&Fonagy, 2006). But they all agree that it is evolutionary acquired –other primates have less developed forms of this faculty as well (Tomasello, 2010). And even the neural circuitry that performs this cognitive task has been localized in our brains (concerning notably the medial prefrontal cortex –Mahy et al., 2014). We use this faculty every time we make guesses and assumptions about the *intentions* of others (Call&Tomasello, 1998), their *beliefs* (Wimmer&Perner, 1983), and even when we perform *joint attention* (looking together at the same point) (Baron-Cohen, 1991). And this is crucial for getting along in our social realities, in every interaction: It is what allows us to engage in conversations, in collective games, dancing together, and making any kind of teamwork (Tomasello, 2010). It is, ultimately, what has allowed humans to unite the forces of plural individuals under a *shared intentionality* and, thereafter, prosper collectively as a species (Searle, 1983). Every shared-intentionality activity requires minds that are sensible to the states of other minds, minds capable of making guesses about others, correcting these guesses, and responding accordingly (Howhy, 2013). Also reading is an activity of this kind (Zunshine, 2006).

However, as we experience every day, our intuitive mindreading faculty –like any of our mental faculties– is not flawless. I have studied literature and philosophy for a long time. And, unlike many of my colleagues, I was not only interested in elaborating original interpretations of literary works (that is, in *criticism*), but also in imagining and guessing what interpretations others would elaborate and how they would be affected by the texts in question. For instance, if my colleagues were studying *Moby-Dick* (Melville, 1851) for a class, I would try to guess in advance their readings and their reactions: "Gustavo is a Medievalist, he will probably do a theological reading, where the whale represents God or the Devil –he won't care which. Flor is a cinephile, she will probably compare the novel with the films that were based on it and, as always, feel the coincidences as merits and the differences as defects. Jérôme is French, he looks like he likes Lacan, so I guess he will interpret the white whale as incarnating the *super-ego* and the Captain Ahab as the *id*. Laura hates long adventure novels, she probably found it so boring that did not even finish it. And the professor sees everything in terms of power relations and identity politics, so I am sure she interpreted some underlying message in relation to the color of the whale and the lack of ethnic and sexual diversity of the crew." These guesses were a measure of my understanding of myself, of others, and of the cultural space that surrounded me. But, very often, these guesses were wrong. People's responses to language can be very personal and difficult to predict. I remember, for instance, a professor that built an all-encompassing Freudian reading of a poem based on the fact that the poem began with a word whose initial letter was "m" ("…Like in *mother*," she argued). Sometimes people's interpretations seem completely random and arbitrary. But the point is that this mental game of trying to predict others' mental responses to literary texts brought me to realize both the importance and the limits of our mindreading instinct.

We are also deficient at predicting how others will respond to *our own* messages. This is evident not only in the field of literature, but in every field of communication. Consider the following examples: In 2017, Coca-Cola tried to link Diet Coke with the feeling of getting psyched up before performances, and came up with the slogan "You're on – Coke." After this, the company ended up being accused of promoting drugs. For the 2012 Olympic Games, Nike tried to foster competitiveness with the slogan "Gold digging." The expression backfired by offending a significant part of the feminine public. And 2012 Pepsi's slogan "Live for now" was the opposite of inspiring for many that could not help but read it as "Live, *for now*…"

Our inaccuracy in predicting how others respond to language is often amusingly evident, but it is also a serious aspect to inquire, because it is the root of every miscommunication: not only of bad slogans and unfunny jokes, but also of confusing contracts, unexciting speeches, interpersonal misunderstandings, unmemorable articles, incomprehensible textbooks, and boring novels. In limit cases, deficient mindreading skills can even turn out to be lethal. In *The Sense of Style,* Steven Pinker provided an example on this regard (2014). Portable generators and combustion heaters used to carry warning messages that looked like this:

> Mild exposure to CO can result in accumulated damage over time; extreme exposure to CO may rapidly be fatal without producing significant warning symptoms.

Despite the accurate warning, several hundreds of people a year asphyxiated themselves and their families by running heaters and generators indoors. This only changed when somebody –with a more precise idea of how people respond to language– decided to write instead:

> Using a generator indoors can kill you in minutes.

Only with this rhetoric improvement, the frequency of accidents decreased immediately. Our mindreading skills can effectively be a matter of life and death, literally –the rhetoric evolution of health warnings in tobacco packages and other harmful products provides clear evidence of this, as well.

But what makes people respond to language one way or another and how can we predict these things with more accuracy than allowed by our direct senses and intuitions? We have known for a long time that our naked senses and intuitions are inaccurate and unreliable: our visual sense can be tricked with optical illusions, and our sense of time with entertainment, we cannot perceive certain kinds of lights (such as ultraviolet, infrared, or x-rays), we cannot guess the *exact* temperature of an object only with our touch, nor the *exact* frequency of a sound only with our hearing sense. Evolution has not designed our bodies and minds to be *accurate* but to be *efficient* in coping with our most directly pertinent environment. For this reason, modern science was only possible –among other reasons– thanks to the invention of technologies and techniques with greater measuring precision than that of our senses, techniques that allowed us to extend many of our cognitive capacities beyond their natural limitations. A simple ruler, trivial as it is, already permits us to outperform the precision of our naked vision for measuring distance; telescopes, microscopes, and x-ray glasses allow our vision to reach levels that the humanity in the past would have considered supernatural; and calculators allow us to make algebraic computations faster than any human brain. We have extended many of our faculties through technology, and that has increased enormously our understanding of the world and our agency. Nevertheless, it seems much more difficult to conceive how to extend the mindreading faculty that we use to guide our behavior in things such as choosing a political candidate, taking financial decisions, handling relationships, or writing a novel. By means of which measuring methods and instrument could we become better mindreaders, better predictors of people's responses when reading stories, poems, or even words? And how would a discipline devoted to this endeavor even look like?

As a literary scholar, I was surprised to find some of the most interesting answers to these questions not exactly in the tradition of the humanities, but in the natural sciences and in computer-based statistics. Indeed, the novelty that motivated this research is that, in the last decades, the world of knowledge has produced two revolutions that have changed the ways in which we approach the study of human subjectivity, of our mental states and responses: The Cognitive Revolution and the Digital Revolution.

MINDREADING SCIENCES: OBJECTIVE STUDIES OF SUBJECTIVE FACTS

How can scientific disciplines produce any kind of knowledge about subjective facts, such as how we respond to language and literature? Isn't there a contradiction in that project, since cultural phenomena are *subjective* whereas scientific measurements are *objective*? This epistemological riddle has a long philosophical tradition (Searle, 1996). The disagreement about it is one of the causes of the historical indifference (and some times even hostility) between significant parts of the scientific and the humanistic communities (Snow, 1959) – which lasts until today (Jardine, 2010). Nevertheless, the riddle is no longer such: Today it is clear –because it is routinely done– in which particular way it is *de facto* possible to produce objective knowledge about subjective facts (Searle, 2015). Perhaps the simplest way to illustrate the way in which the natural sciences do this is with the case of pain.

Just as meanings, happiness, the value of money, or the perception of beauty, pain is, ontologically, a subjective fact: That is, it is observer relative, it is a fact that only exists insofar as it is experienced as such by somebody. In consequence, it is not possible to study pain *in itself* –because it does not exist *in itself*. However, it is perfectly possible to measure the signs that show whether the subjective experience of pain is really occurring *in an individual*. Functional magnetic resonance imaging (fMRI) is a technology that was created in the 90s for measuring brain activity by detecting changes associated with blood flow –which are coupled with neural activations. Since it was verified (by systematic trial and error) that pain is produced by the activation of the anterior insula and the anterior cingulate cortex of the brain, any neurologist can today scan a patient's brain so as to *objectively assess* whether that person is *subjectively experiencing* pain. Just like beauty, pain is in the eye of the beholder; the news is that now we have instruments and techniques that allow us to behold the minds and brains of the very beholders. And there is no contradiction here: What you are subjectively thinking might not be objectively true (you could be thinking unicorns are real, for instance), but it is objectively true the fact that you are having a subjective thought (about unicorns). Of course, it can be very difficult to discover the means to objectively measure subjective facts (such as *being in pain* or *thinking of unicorns*). But that is a technological problem, not an epistemological one.

In this book, I will be using the term *cognitive science* as an umbrella-concept to refer to three interrelated disciplines that approach the problem of the human mind in this naturalistic manner. Each of these disciplines is based on a core idea about human nature.[1] The first one is cognitive psychology, whose core idea is that the functioning of the human mind can be accounted for in computational terms (as if it was a sort of biological software) (Putnam, 1961). The second one is neurocognitive science, whose core idea is that this mental software is an emergent phenomenon generated by our neurons by forming patterns of activation (s. Calvin, 1996). The third one is evolutionary psychology, whose core idea is that our brains and the fundamental structures of our minds were formed by the same process of natural selection that formed all other organs of our bodies (Barkow et al., 1992).

On the basis of these ideas, the naturalistic studies of the mind have advanced and gotten to address subjective phenomena that are much more complex and abstract than pain: things such as language, perception, memory, moral sense, aesthetic sense, our ways of categorizing, our emotions, and even consciousness are being now regularly modelled in computational terms and empirically tested in labs (s. Pinker, 1997). The progress of these disciplines is allowing us to understand each time better the complex connections between our natural endowments and our cultural developments (Barkow et al., 1992). And, in doing this, these disciplines have shown us that the human soul, with all its complexity, is not as inscrutable as we thought; in fact, some of our innermost subjective states and responses have proved to be, in many aspects, as readable as a book –when read with the adequate lenses.

Data science, on its behalf, offers us a somewhat more indirect way of measuring and predicting subjective responses. The approach of data science is purely quantitative. Data science is the discipline that designs digital tools for making statistical analysis of data. We have known for a long time that computers do statistics better than humans. But the data produced by human behavior has grown so much in the last decades, that these tools have revealed a new and unexpected importance.

We register every day hundreds of digital footprints in the cloud just by being online and using social media, search engines, digital maps, and other apps (UNECE, 2016). On this basis, digital tools have turn out to be like telescopes and microscopes measuring the cultural space at scales beyond the capacities of any human reader. We make use of these studies –and are object of them– constantly in our everyday lives. Online video engines predict which video we would like to watch next. Social networks predict who would we be interested in meeting. Automatic translators predict what we might have meant in different languages. Even our email account makes predictions of our subjective responses –v.g. which emails we will not find interesting and should then be discarded as *spam*– and it is accurate more often than not. Our current data about human behavior is so huge and growing at such a speed that, with enough data-processing power, virtually any subjective response that is not completely random, anything that we do in any patterned way, could become predictable (Floridi, 2016).

These mindreading sciences are effectively making a huge progress in increasing our understanding of the human mind, and this progress is empirically verified by their increasing accuracy for predicting our subjective responses (Bermúdez, 2010). However, if we can use this knowledge to better *read* minds, it means that we can also use it to *manipulate* minds. And this entails a great risk that is proportional to the predictive power of these disciplines. Indeed, the influencing strategies based on our understanding of the human mind can be used to help people avoid smoking, to foster them to eat healthier, to recycle, or to read more

[1] Indeed, cognitive linguist Steven Pinker has often referred to this set of disciplines as the *sciences of human nature* (s. Pinker, 2007).

books, but they can also be used for making people eat junk food, vote against their convenience, or support war. Recent political events have actually revealed a sample of the enormous potential of impact of the mindreading sciences (s. Grassegger&Krogerus, 2017; Chapter 7). For this reason, it is indispensable to develop strategies and policies for fostering the advantages of the mindreading sciences and preventing the perils of their misuse. Indeed, this is currently an intense topic of debate world-wide (Broeders et al., 2017).

In any case, we must first understand how these disciplines work, what kinds of knowledge they produce, and what they can tell us about our minds and our cultures. This is one of the main goals of this book.

WHAT CAN SCIENCE SAY ABOUT LITERATURE?

In what sense can these mindreading sciences be concretely useful for thinking about language and literature? For authors, understanding how language and literature are cognized and increasing their abilities to predict how different readers might respond to them can certainly be of great utility in a very direct sense: for better calculating the effects of their words. Influencing an audience is not only something that ill-intentioned institutions might leverage. Influencing strategies are also the techniques we all use everyday so as to make a text or speech more moving, memorable, or even clear, for example. We can see this in the teacher that uses a story or a metaphor to make a concept more appealing and easier to grasp, the stand up comedian that manages to make everybody laugh with a word game, the social communicator that brings a community together by framing a problem in a way that conciliates diverse views, or the filmmaker that makes us empathize with a character and feel moved by his or her passions. All of them are skilled mindreaders making a positive use of their abilities to influence their audiences, as well. Understanding the triggers of the human mind and learning how to use them is not in itself pernicious, but it is what makes communication efficient. In short, the better we are at mindreading, the better communicators we will be, because reading minds entails better understanding ourselves, others, and the cultural space in which we interact.

For literary scholars this is also crucial. Many of the fundamental questions posed by literary studies ultimately roll back to overt attempts of reading minds. Rhetoricians ask things like: Which of these expressions is more persuasive? Historians of literature ask: How were Shakespeare's poems interpreted in Elizabethan times? Scholars of comparative literature ask: To what extent did Dostoevsky's *Crime and Punishment* (1866) influenced Kafka's *The Trial* (1959)? Formalists ask: Is the ellipsis of the main event of the plot what gives Hemingway's short-story "The Killers" (1927) its suspense? Asking what makes people be persuaded, what meaning they might have attributed to a text, how people influence each other, or what makes them feel a particular effect is to inquire about their subjective responses, about their mental states. The school of Reader-Response criticism emerged in the 60s precisely out of understanding that literature is a subjective phenomenon, that it occurs in our minds, not in the books (Freund, 1987). However, until recently, it was not possible to really address these questions empirically –due to the lack of a systematic theory of the mind and the technical means to test it. What I will argue along this book is that cognitive science and data science are providing us with the theoretical models and tools necessary to start doing it. And, this way, these mindreading sciences are inviting us to explore a new chapter of literary studies.

Signs of this approach to cultural research are already visible. In the last decades, we have seen the emergence of two new fields: cognitive cultural studies (Zunshine, 2010) and digital humanities (Burdick et al., 2016). These fields were born in the United States, but they soon spread to Europe and then to the rest of the world. Each of these fields has brought ´a novelty to the humanities and, in particular, to literary studies. The so-called cognitive literary studies have integrated into literary research the insights developed in cognitive psychology, neurosciences, and evolutionary psychology about how our minds work when creating and consuming literature (Hogan, 2003; Jaén&Simon, 2013). And digital humanities have made of the methods developed in data science tools for analyzing unprecedentedly large corpora of texts and registers of reader-responses (Jockers, 2013).

However, until now, these emerging trends of cognitive and digital approaches to culture have lead programs of research quite independent from each other. They have not been yet considered in terms of the insight they can give us about our cultural realities when working in tandem.[2] The main thesis that I will put forward through this book is that these two fields are complementary –their compatibility lays in the fact both approach our subjective realities from an objectivist perspective. And they can be therefore combined in a coherent framework with a great potential for studying cultural phenomena empirically. Each of these disciplines would contribute to this framework by assuming a particular role: data science as a provider of tools for collecting and analyzing representative samples, and cognitive science as a provider of models for accounting for the patterns of data found.

Along this book, we will analyze many of the insights that this cognitive-digital framework can give us of language and literature. The book consists of three parts.

The first one describes the mental faculties that allow humans to process literature, as well as many of the cognitive biases that intervene in this process. The presentation of these faculties is articulated in relation to different formal levels of literary texts: how our minds manage to attribute meaning to sounds and letters (Words), how our interpretations are affected by the choice and order of of words (Rhetoric), how we process narrative information (Story), how we judge others through language (Characters), and how literature makes us feel (Emotions).

The second part of this book is devoted to the exploration of different digital tools for analyzing language and literature in quantitative ways. We will consider the kind of information that these tools can discover about things such as the identity of the author of a novel, about his or her style and influences, about the genre of the text, the time, or even about what random readers might think the text means and how much they might like it. The last chapter of this second part presents a panorama of current development in verbal artificial intelligence –which is a field that effectively combines cognitive and digital perspectives–, as well as an analysis of the interest of these inventions for understanding some aspects of human language and literary practices.

The third part of this book consists of a case study: a cognitive-digital study of the lyrics of a song (Sting, "Every Breath You Take," 1983). The aim of this case study was to illustrate the potential of integrating cognitive science and data science as a framework for studying literature. The study consisted of three steps: (1) a text-mining analysis of a digital register of reader-responses to a video that included the text, (2) a formulation of informed hypotheses of likely reader-responses, (3) a verification of the predictive accuracy of these hypotheses through an open survey. As for the results, the most pertinent was the following one: the cognitive-digital model predicted people's interpretations 36% more accurately than humans do. This means that, by using this model, we can become *that much better* at mindreading

[2] As a notable exception, I would like to mention the activity of the Digital Humanities & Literary Cognition Lab (Michigan State University).

others, i.e. *that much better* than by the use of our “naked” mindreading sense. These results are taken as a measure of the validity and potential of the framework proposed in this book.

Exploring why we respond to language and literature in the ways in which we do has led me to understand that many ancestral questions about our nature, our culture, and our behavior are finally finding answers in the scientific study of the mind, of our brains, of the evolutionary history of our species, and also in the footprints that all of us produce day by day in the cloud, which is the largest collective register that humans have ever created, the highest-resolution mirror of humankind.

This research has helped me to better understand the complex and dynamic relationship between nature and culture, between our biology and our poetry. Literature, arts, and other cultural practices are a manifestation of fundamental parts of human nature. In this sense, this book intends to further introduce science into the humanities as much as the humanities into the sciences. This book is intended as a contribution to the project of bridging these two spheres of knowledge into a collaborative and mutually enriching dialogue. I hope these pages transmit to the reader the same excitement I found in exploring these subjects and writing about them –I can only *hope*: after all, one can only mindread so much.

Part I

COGNITIVE SCIENCE

THE PREDICTABLE BIASES OF THE READING MIND

CHAPTER 1

LANGUAGE AS A NATURAL PHENOMENON

FROM ATOMS TO POETRY

As you read these words, your mind is identifying units of meaning, making sense of this discourse, interpreting, understanding, agreeing or disagreeing. But how does all this happen? Our neurons could be checked one by one, and we would find that none of them is in itself actually *thinking*. Yet, as a group, our neurons certainly are thinking. The gap between these two levels of phenomena seems at first sight unsolvable: How is it that mindless neurons can produce mindful brains? In philosophy of mind, this question is traditionally known as the Mind-Body Problem (Lakoff&Johnson, 1999; Searle, 2005; Dennett, 1992 & 2017). This chapter provides a general view of cognitive science by paying attention to how this problem has been approached. And this will be done by focusing on one of our most important mental faculties: language processing –which will be the subject of this first part of the book.

Philosophically speaking, the perspective of cognitive science is naturalistic. This means that cognitive science considers the mind not as a supernatural miracle (such as the *soul* of religious traditions) nor as a completely separate and independent dimension of reality (such as Descartes' *res cogitans*) but, instead, as a phenomenon of nature, a fact that occurs within the natural world we inhabit and that must therefore be explained in terms of natural principles. Cognitive science actually encompasses several disciplines and it interacts with many others. We will use the term *cognitive science*, in this book, to refer mainly to three disciplinary field that are deeply intertwined. Each of these fields accounts for a level of the relationship between or biology and our minds. In words of cognitive linguist Steven Pinker, each of them enlightens "a bridge between nature and culture" (S. Pinker, 2007). The main interest that I see in these disciplines is that, together, they can give us a coherent integrative model that can solve the gap that bring us from the utmost physical reality (*from atoms…*) to the most abstract cultural reality *(…to poetry*)[3]. Each of these disciplines is grounded on a fundamental thesis:

- *Cognitive psychology*: Its main thesis is that the mind can be described as a software that provides us with particular capacities and constraints and –within those parameters– is then further informed, developed, and attuned by our experiences in the world. (Putnam, 1961)
- *Neurocognitive science*: The abovementioned mental software is considered to run on the hardware of the brain tissue by forming patterns of neural activations. This thesis, in combination with the previous one, constitute what is today called the Computational Theory of Mind. (Calvin, 1996)
- *Evolutionary psychology*: To the extent that our mental and neuronal endowments are shared by the species, we must provide an evolutionary account of how they came to be as they are.

[3] The formulation of this idea is inspired in Daniel Dennett's book *From Bacteria to Bach: Evlutionary Mind*, 2017.

> This means that the adaptive rationale and evolutionary history of our minds and brains must be inquired. (Barkow et al., 1992)

In this first part of the book, I will give an overview of these disciplines and of the ways in which they approach the problem of language –our does our capacity to cognize language work, how is it produced by our brains, and how it might have evolved. These disciplines will constitute the theoretical framework that we will use for making sense of literary phenomena in further chapters of this first part of the book.

COGNITIVE PSYCHOLOGY: THE MIND AS A SOFTWARE

Until the 20th Century, there was no precise scientific model for studying the mind. Most discussions about the subject laid on untested (and often contradictory) assumptions, typically inherited from religion, popular knowledge, or speculative philosophies. In this context, behaviorists psychologists assumed that it was ultimately impossible to formulate theories about the human mind without falling into superstitions or paradoxes. They came, thus, to consider that the only thing that could be scientifically studied was external overt behavior, avoiding hypotheses about the possible subjective states to which behavior could be related. But this program –in its pure form– could not last very long, because by leaving the mind out of the equation, behaviorists were missing the fundamental causal grounds of behavior. In concrete, the problem is that similarly-looking external behavior can respond to different mental states, and vice versa, so we cannot properly account for one dimension by discarding the other. As an answer to this, cognitive science proposed to bring back the mind as a legitimate and necessary object of study. And it did it by proposing the first properly scientific conceptualization of the mind: considering it as a form of software –this was the beginning of the Computational Theory of Mind (s. Turing, 1950; Putnam, 1961; Fodor, 1978). A clear example of how this theory increased our explanatory power can be found in the case of language.

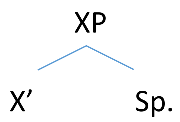
A child says “Hi, Paco!” and a parrot says “Hi, Paco!” These are two very similar-looking kinds of behavior. But where is the difference between human and parrot speech? If we analyze more extensive samples of utterances produced by humans and parrots –as a proper behaviorist would also do– we can observe a measurable difference in the kind of patterns humans and parrots produce. In particular, parrots produce a finite (and usually quite limited) number of utterances that they repeat with scarce variation; whereas humans can elaborate seemingly infinite new arrangements of words. But how can we explain what causes these differences? A behaviorist cannot answer this question, because it requires a theory of the mind: a hypothesis of the inner mechanisms that make one organism behave in a way and the other in another way. This is the gap that cognitive science aimed to solve by proposing that our minds are a kind of software, constituted by systems that process information and translate it into bodily and mental experiences.

In computational terms, the patterns of utterances of parrots can be described as responding to a program that would produce statistical assimilations of environmental sounds and a probabilistic assembling: i.e. if one says “Hi, Paco” frequently enough, one’s parrot will eventually catch it and repeat exactly that. The case of human speech is certainly more complex. What kind of computational model could allow the production of infinite utterances? Doesn’t the variety of human speech suggest that our utterances are purely random and arbitrary? Or couldn’t it be that our minds work just like the parrots’, simply registering and repeating the sounds that we hear in our surroundings by sheer frequency?

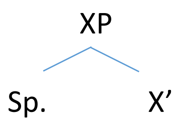
Until the 1950s, a computer meant "a person that performs computation." Computers were salaried employees that computed figures. But Alan Turing (the father of computation as we understand it today) realized that these information-processes carried out by the minds human-computers could be reproduced by a system that would automatically follow certain algorithms (Turing, 1950). And he made a mathematical model to prove it. The system postulated by this model is today known as Turing Machine. This made the first computers possible, and it also provided evidence for the plausibility of the theory to the human mind as a (very highly developed kind of) computer. The Cognitive Revolution emerged parallel to the Computational Revolution.

Noam Chomsky was the first to address the problem of human language in computational terms. He began by observing that human language, for all its infinite creativity, is not arbitrary but lawful, and not merely statistical but more complex and creative. A simple thought experiment is illustrative of this. Chomsky formulated the following –now famous– sentence: *Colorless green ideas sleep furiously*. In terms of meaning, the sentence is absurd, even arbitrary, and statistically every word of this sentence has very low probabilities of appearing next to the other in actual speech (green things are never colorless, ideas have no color, they don't sleep, and furiously is an unconceivable way of sleeping). Nevertheless, we all perceive it as a perfectly grammatical construction of English Language. Compare that sentence with the following one: *Eats hamburger Paco my*, whose meaning is relatively clear and some of its words are even likely to appear together in everyday language, but we perceive it, nevertheless, as ungrammatical, instead. What this shows is that there must be concrete underlying rules in our minds organizing the ways in which we process language –making us feel the former sentence as more legitimate than the latter–, relatively independent of meaning and frequency. As a model to account for this underlying structure, Chomsky proposed the so-called *Generative Grammar*, a mental program constituted by a finite set of combinatorial rules that can be used to produce and process an infinite set of linguistic arrangements, in a generative manner (Chomsky, 1965).

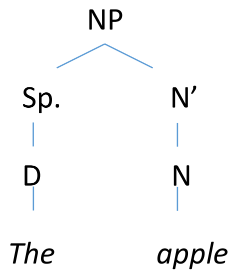
Chomsky elaborated different formulations of this generative grammar through the years. His latest model –called the *minimalistic program*– reduces the whole system to a minimal constructive principle: the X-bar theory (Chomsky, 1993). X-bar is a structure for processing information –what cognitive scientists call a *schema*– which can be recombined in a recursive manner with a minimal set of rules, enabling thereafter the production of the infinite sentences that any human language can express. For example, an X-Phrase (XP) would consist in a head (X') plus an optional specifier, in any order:



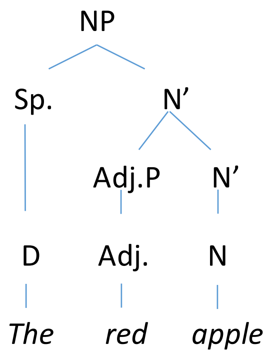
Or:



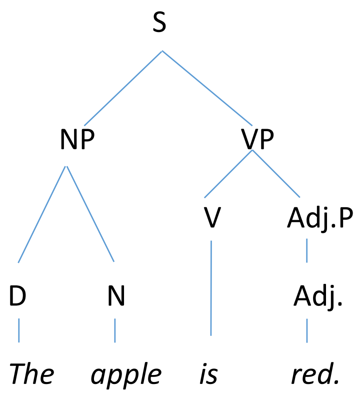
This structure serves for forming any kind of phrase. For instance, a nominal phrase (NP) in English constituted by a determinant article and a noun would be expressed in this way:



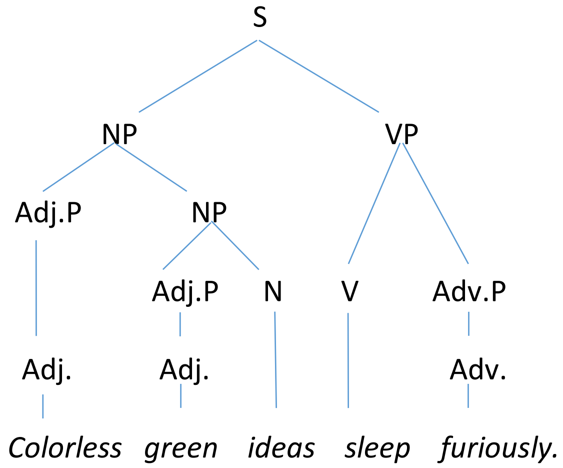
Now, this head (X') can also count as the basis of a new XP (in the form of a head and a complement or a modifier, such an adjective phrase), for instance:



In this way, this schema can be iterated indefinitely, generating thereafter any kind of sentence (hence, the name of *generative* grammar). The resulting sentences can be accordingly graphed in the form of *parsed trees*, where each node expresses an X-bar structure, such as:



Or even:



This minimal principle accounts for a generative mechanism whose productivity can be formally tested and compared with the actual outcomes of human linguistic performances so as to assess its explanatory power. For example, this model explains why we accept the expression “green ideas sleep furiously” but not “furiously ideas sleep green” –which would mix up the branches of the syntactical tree in ways that cannot are not admitted by the rules of the X-bar theory –this kind of proof is known as the *substitution test*.

Once we think of the mental structures that we use to process linguistic information in terms of software, we can also compare this computational model with others. And this is very important, because this generative grammar, as proposed by Chomsky, needs to account not only for the functioning of English or Spanish but also for any other human language. Until the appearance of Chomsky’s theory, studies of grammar were typically focused in the differences between languages: they would describe the specificities of English, Spanish, or French, for instance, that are not obvious to native speakers of other languages. But, after Chomsky’s formulation of a grammar posited as a property of the human mind and formally describable in computational terms, linguists started to pay attention to the shared aspects across human languages. And, indeed, many patterns of communalities became then evident. Unrelated human languages across the world actually show strong regularities in every grammatical level: For instance, among infinite possibilities that can be mathematically conceived, all human languages tend to use between 20 and 50 sounds each, from the same restricted set of around 100 phonemes (s. International Phonetic Alphabet); they tend to use some of the same restricted world classes (verbs, nouns, adjectives, pronouns); verb affixes consistently respond to some of the same restricted categories of information (person, number, tense, mode, aspect, voice), and even though the most usual head and complements of sentences can be ordered in 128 logically possible ways, 95% of human languages use only one of two models.[4] The fundamental parameters in accordance to which our minds process language seems to be, indeed, shared across cultures. Thereafter, Chomsky came to consider his generative grammar as a *universal grammar* (UG), describing it as a species-specific mental faculty (Ibid.). That is, the fact that you speak German and I speak Spanish would be as if you and I were in each of our computers manipulating different files but, at the same time, both using the same brand of computers (our human brains), the same operative system (our human minds), and the same program (our human language processors).

[4] Many more of these linguistic universals can be found in Mark Johnson, 2003.

Chomsky's idea of the universal grammar helps answer, at the same time, a traditional problem in psycholinguistics: How is it that children acquire speech? According to Chomsky's observations, the aggregation of all the linguistic stimuli that children receive from the environment until they acquire a language is not sufficient to model a system equivalent to the system required to reproduce children's linguistic behavior (Chomsky, 1965). On this basis, he concluded that there must be an innate mental endowment that allows the child to perform the learning, to begin with, and that can account from the gap between the scarce input of linguistic stimuli and the sophisticated outcome of a child's linguistic competence – this is called the *Poverty of stimulus argument*. This innate endowment is, of course, the universal grammar. That is, this program would be universal precisely because it would be the innate structure that allows humans (any human) to a acquire a language to begin with. Guided by the principles of this program, when a child hears human speech, he or she is instinctively led to seek certain sounds, word classes, and structures so to parse the information received in a certain predefined ways and to organize it into a human linguistic system. This model explains why children all over the world learn a language so effortlessly and by an analogous process, even at the same age; whereas a parrot, a dog, or a cat, exposed to the same stimuli, would not. This is why Steven Pinker refers to language as *an instinct* more than a faculty; and instead of saying that a child *acquires* a language, Chomsky prefers to say that he or she *grows* a language, because, within the appropriate environment, language grows in the human minds –metaphorically speaking– as naturally as hair grows in the human body.

The same logic has been applied in cognitive science to the analysis and computational modeling of many other cognitive faculties. And by comparing the results across cultures (and even across species, as comparative psychologist do), many universal aspects of the human mind have been discovered.[5] Steven Pinker puts it in a way that stresses how this changes our view of the relationship between nature and culture:

> Superficial categories of behavior certainly do vary across cultures and have to be learned (specific rituals, taboos, superstitions, languages, etc.), but the deeper mechanisms of mental computation that generate them are to a large extent universal and many of them even innate. (…) It varies across cultures what counts as an affront and as a retribution or punishment; but we all seem to be equipped with a program that responds to an affront with an unpleasant burning feeling that motivates us to punish or to exact compensation. People may dress differently, but everywhere they flaunt their status via their appearance. They may respect the rights of the members of their clan exclusively or they may extend that respect to everyone in their tribe, nation, or species, but all divide the world into an in-group and an out-group. (…) The behaviorists got it backwards: it is actually the mind, not behavior, that is lawful. What this model of a universal and generative mind shows is that the frame of traditional debates about human nature was wrong. It is simply misguided to ask whether humans are flexible or programmed, whether behavior is universal or varies across cultures, whether acts are learned or innate. Humans behave flexibly *because* they are programmed: their minds are packed with combinatorial software that can generate an unlimited set of thoughts and behavior. Behavior may vary across cultures, but the design of the mental programs that generate it need not vary. Intelligent behavior is learned successfully because we have innate systems that do the learning. (Pinker, 2007).

Many of our mental programs are intertwined with our language processing capacity: from memory and rationality to emotions and social cognition. We will see in the following chapters some of the diverse ways in which these programs model our interpretation of

[5] A well-documented register human universals can be found in the work of cognitive anthropologist Donald Brown, 1991.

language and affect our experiences and decisions in often systematic and predictable ways. This will lead us to explore in cognitive terms a series of phenomena that are very pertinent for understanding how many literary effects work, and how our minds respond to them, such as: why concrete words are easier to grasp than abstract ones, why we remember plots more easily than incidental details, how do we "fill the gaps" of a story when encountering ellipsis, why things like physical appearance and status can bias our moral judgment of characters, why we are moved by events that we know to be fictional, and even how literature can plant false memories in our minds. In addition, as we will see more thoroughly in the second part of this book, once we understand these computational models, we can code them into computers so as to create artificially intelligent bots capable of performing tasks analogous to (and often much better than) human thinkers, such as retrieving selective information from texts, spotting subjects, translating, and even predicting some of the intentions, emotions, and interpretations of human readers. But, first, let us continue exploring how our capacity to process language is embodied in our biology.

NEUROCOGNITIVE SCIENCE: THE BRAIN AS A HARDWARE

If the mind is a software, what is the hardware on which it runs? The answer is, of course, the brain. The fact that our brains are the material source of our mental life counts with extensive evidence. This is not only a truism inferred from the fact that losing your head necessarily brings your mental life to an end, but also a reality with a very precise functioning that neurologists are progressively unraveling (s. Calvin, 1996; Damasio, 1994; Crick, 1994; Gazzaniga, 2000). For example, localized lesions in specific parts of the brain have proven to make people lose very specific mental faculties, such as the capacity to see, perform mathematical operations, experience emotions, feel pain, recognize people, or even use certain words classes –just as a computer can no longer access certain files or programs when a part of its hard disk gets burnt.

The anatomy of the human brain is quite regular across our species –independently of where one comes from or what one's culture and experiences might be–, it is similarly shaped in each of us by our genes already from prenatal development, and the regions, folds, and wrinkles of our brains are so consistent that been mapped with millimetric precision and can be identified in any person.

Moreover, the localization of many of our mental faculties is similarly distributed in people's brains, as it was already observed by Herbert Jasper and Wilder Penfield in the 1930s. They performed brain surgeries in awake patients (under only local anesthesia), and discovered that, by activating with electrodes different points of people's brains, the patients could be led to move a part of the body or perceive certain sounds and smells, for example (s. Jasper&Penfield, 1954). This way, they identified and mapped for the first time the sensory and the motor cortex of the human brain. Since the 1980s, similar experiments became more usual, namely after the refinement of technology for performing transcranial magnetic stimulation (TMS) –which is a non-invasive procedure in which a changing magnetic field is used to cause electric currents to flow in targeted regions of the brain. Further experiments allowed by these and other technological advances confirmed many of Penfield and Jasper's observations, to the extent that their cortical maps are still used today.

The concrete way in which neurocognitive science considers that our minds are encoded in our brains is by the formation of patterns of neural activations (Calvin, 1996). This mechanism is synthesized by the so-called Hebbian rule, which states that, whenever two neurons fire together, they wire together –in this context, *being wired* means that the activation of one will give *priming* to the other, will increase its likelihood of getting activated as well (Hebb, 1949). In the 1990s, the invention of functional magnetic resonance imaging (fMRI) allowed neurologists to start measuring these patterns and to process the resulting data with computer power. This increased the accuracy of possible predictions in an incredibly manner and accelerated enormously the progress of neurocognitive science. As a result, neurologists can today literally read many of our mental states, such as what emotion a person is feeling or whether one is imaging a place or a face (and even *which face*), by identifying which neural patterns fire in which brain areas (s. Spreng et al., 2014).

To better describe how neurocognitive science accounts for the link between neurons and thought, it is pertinent at this point to come back to the philosophical question that opened this chapter: if none particular neuron of our brains *thinks* in itself, how is it that we, as organisms, do it? What kind of natural phenomenon disappears when you look at its constitutive elements with a microscope? As a matter of fact, a wide range of natural phenomena work that way. In philosophy of science, they are called *emergent* phenomena (s. O'Connor&Wong, 2012). Think of the liquidity of water. None of the molecules of H2O is in itself liquid. However, when several H2O molecules acquire a certain level of kinetic energy (i.e. move at a certain speed), a new phenomenon emerges at the level of the whole, a new general state of matter produced by the interplay between the particular states of the components, and this phenomenon has new properties that are not shared by the individual components –like being liquid or solid. The state of molecules (kinetic energy) really causes the state of the matter (temperature and consistency), even if such state of matter is not directly observable in any particular molecule. Emergent phenomena occur at every level of nature: a molecule of H20 has properties that are absent in hydrogen molecules and oxygen molecules when they are separate; also biological processes are emergent phenomena (a cell can be alive without any of its components being alive), as well as social processes (ant colonies have emergent properties that differ from the properties of each individual ant), and, as said, also mental processes can be described as emergent phenomena. In short, just as we say that liquidity or solidity are states of matter, we could say that *the mind* is a *state of neurons*, that is, an emergent phenomenon that arises when a collective of neurons forms a particular arrangement, and this emergent entity can have properties that are different from those of its constituents: such as consciousness, intentions, desires, values, reason, and some times even language and literature.

People tend to think of neural states and mental states as separate phenomena due to our intuitive way of conceptualizing the world. When somebody asks "Why did Juan cross the street?," people normally do not try to provide a physical-chemical explanation (talking about neurons and hormones, for instance), but people would normally recruit a different kind of knowledge and say, instead, something like: "Because he wanted to get to the other side." We tend to think that physical facts have *causes* whereas mental facts have *reasons*, and that these concepts refer to irreconcilable orders. But this is only the result of an intuitive way of seeing and describing the world. Scientifically speaking, our reasons, our mental states, also have ultimately physical causes. The fact that we talk of intentions, beliefs, values, and interpretations does not mean that these things exist independently of matter –just as the temperature of water does not exist without the water. When the specific correlations between neural activities and mental states are studied systematically, we can discover how the latter emerge from the former: In Penfield and Jaspers' experiments, when patients were led to move an arm or a leg by stimulating their brains with electrodes, these patients really felt as if they had *deliberately decided* to make these movements by themselves; that is, the chemical stimuli did not only produce physical changes but also mental ones, like the experience of intentionality (s. Jaspers&Penfield, 1954). Even our very sense of self has been identified as produced by patterns of neural activations in particular parts of the brain (s. Damasio, 2010). The fact that subjective phenomena are observer-relative, mental, does not mean that these phenomena are really independent of matter. All the mindful reasons that made Juan cross the street can be indeed explained as emergent phenomena caused by the interaction of his mindless neurons, and there is no paradox here: the production of higher-level properties (such as *mindful reasons*) is precisely what characterizes emergent phenomena.

In the last decades, neurolinguists have produced extensive research so as to identify the parts of our brains that process language and to discover how they work. Already in the early stages of neurolinguistics, by studying patients with aphasias (language deficits) produced by brain lesions, Paul Broca and Carl Wernicke identified –each on his behalf– two areas in the left hemisphere of our brains that intervene in language processing (s. Kennison, 2013). They are known today as the Broca area and the Wernicke area, respectively (s. Figure 1). A lesion in any of these areas would affect a person's capacity to process language (either to speak, to understand, or both). But, depending on where the lesion is located, people will experience different symptoms. People with damage in the Broca's area (condition known as *Broca's aphasia*) tend to speak in a telegraphic way: They would say things like "Go, bar, Juan" when meaning "I went to the bar to meet Juan," for instance. More recent studies showed that these patients also manifest difficulties for understanding syntactic information encoded in speech (such as *who did what to whom*) (Caplan, 2006). On the other hand, lesions in the Wernicke's area (*Wernicke's* or *receptive aphasia*) have been observed to produce a different set of symptoms: typically, these people are still able to utter normal sounds and seemingly grammatical sentences, but they often make no sense (s. Brookshire, 2007). These discoveries were crucial for the constitution of neurolinguistics as a discipline. Nevertheless, after decades of research, the particular importance and function of these two brain areas has been reframed into broader and more complex accounts that include also other parts of brain, in both hemispheres, such as the prefrontal cortex (Manenti et al., 2008) the visual cortex, and the sensorimotor cortex (Hauk et al., 2008) (s. Figure 1).

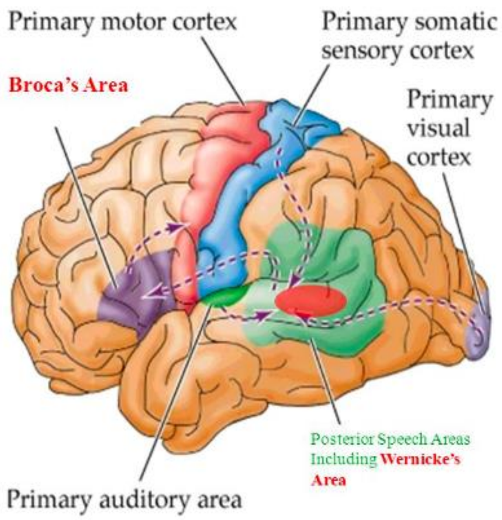


Figure 1. Some relevant brain areas for language processing.

Indeed, although many (including Chomsky himself) have speculated in the past with the possibility that our linguistic faculty was condensed in a distinct brain area, the agreement among neurolinguists today is that language-processing simultaneously recruits several brain areas in a coordinated manner, which has naturally brought up new complexities (Friederici, 2017). One of these complexities is that the same brain area can perform more than one linguistic function at the same time. For instance, we use our hearing system for sensing linguistic sounds, but we also use it when processing the meaning of sound-related words (such as "guitar," "tango," or "sing,"), and just we use our visual cortex to read letters but also for processing the meaning of object-nouns ("sun," "bird," "eye,") (s. Chapter 2 - Words). At the same time, pinpointing specifically linguistic areas proved to be extremely difficult, because most of the brain areas that serve to process language seem to do it only as a particular case of broader and more general cognitive functions. For example, we process the plots of stories by using a particular system in the left hemisphere of our brains, but the general function of this system goes much further: it is what allows us to think in terms of causality, not only when reading novels, but also in our daily life (s. Chapter 4 - Story). Likewise, we use the medial prefrontal cortex for identifying the characters in a novel or a film and imagining their intentions, feelings, and beliefs (and also the author's), but that is only a particular function of the module that allows us to recognize people, empathize, work in teams, form relationships, and ultimately create societies (s. Chapter 5 - Characters).

It is still an ongoing research, the one aimed as discovering exactly our brains attribute meanings to words, process syntactic structures, come up with original interpretations of a poem, make moral judgments of characters and authors, feel beauty, suspense, or horror with a story, or become amused, bored, offended, or persuaded by an argument. But (using Chomsky's terms) these things are no longer *mysteries*: they are now concrete scientific *problems*, because it has become clearer the ontology of these phenomena, where their causes must be sought, what a proper explanation might look like, and what kind of evidence should be taken as valid for confirming or rejecting our hypotheses about them.

EVOLUTIONARY PSYCHOLOGY: THE HISTORY OF OUR MINDS AND BRAINS

Darwin himself championed the idea that evolution must apply also to our mental capacities, our perceptions, emotions, and behavior (1871). The mind is a brain, and the brain is an organ of our bodies. In words of philosopher of the mind Dairon Rodríguez, "cognition is a biological phenomenon" (2017). In this sense, the history of how our species evolved the kinds of brains and the kinds of minds we have, is also a history of natural selection, like the history of any of our organs and biological capacities (Barkow et al., 1992).

This is most obvious regarding certain cognitive characteristics and subjective tendencies that are universal across our species and have a clear adaptive rationale, such as the fact that our minds are capable of experiencing thirst and hunger, which help us monitor our bodies' requirement of nutrients; the fact that we prefer similarly moderate temperatures (not too hot and not too cold); that we like caloric foods; that we tend to feel affection for our kin; or that we have certain universal phobias (heights, snakes, darkness, etc.), which keep us safe from some of the most direct and frequent risks faced by our ancestors –and which act on us even if we live in plains and have never seen a real snake before. All these mental capacities and instincts are clearly not taught to us by our culture: children spontaneously like sugar and reject bitterness, express emotions with similar facial expressions and bodily gestures, are often afraid of heights, feel thirsty when their bodies lack liquid and hunger when lacking food, etc. Nobody would argue that these universal instincts are as adaptive as the shape of our hands or the functioning of our liver. However, in the last decades –thanks to the progress in technology, genetics, and neurocognitive science– evolutionary psychologists have managed to extend this evolutionary understanding of the mind to aspects of our psychology whose adaptive rationale is not so obvious. Things such as language, moral reasoning, religion, politics, science, arts, or literature are being now considered from the perspective of evolutionary psychology, that is, by inquiring how the bodily and mental traits that allow us to perform those activities could have been naturally selected among our ancestors.

The idea that mental capacities are the result of evolutionary processes is also obvious when we think of the mental endowments of other animals: How do sea turtles know, when they are born, that they must run to the Ocean as fast as they can, as they effectively do as soon as they break their eggshells? Nature is full of examples of these kinds of preset instinctive behavior. The European cuckoo (*Cuculus canorus*), to give another example, parasitizes other bird species: they lay their eggs in the nests of other birds –they have actually evolved to produce eggs that mimic the eggs of the species nearby. As soon as the cuckoo leaves its egg (it does it before others), it immediately kicks the eggs of the foster species outside the nest and, this way, it kills its competitors before they hatch (s. Dawkins, 2013). But how do animals "know" these things? Do they learn these behaviors by imitating the example of others? Clearly not. But, still, evolution can provide a perfect account of this. If ocean-eagerness would have increased turtles' chances of survival and an egg-kicking drive would have increased cuckoo's chances of survival, then the organisms that possessed these instincts would have outnumbered the alternatives, and this way, the mental structures that produced these behaviors would have been naturally selected and fostered over time, just as bodily traits are. Organisms are therefore born with particular mental structures that predispose them to perceive the world in certain ways, to behave in certain ways, and to develop certain competences as they develop. If we accept the cognitive model of the brain as a computer, this means that turtles and cuckoos are born with *innate programs* that move them to behave in the described befitting ways.

The brains of human infants are not blank slates (*tabula rasa*) either. Children's minds are not empty amorphous receptacles to be "filled" by cultural nurture only; instead, their brains come equipped with a complex set of innate programs that are developed and specialized by interacting with the world (Pinker, 2007). A blank slate would not be able to perceive, feel, think, and even less would it be capable of learning anything at all. To perform any of these mental tasks, certain innate neurocognitive organization is needed. A child can learn English or Spanish, but it needs first a brain capable of acquiring a language; a child can learn to play the drums, but it needs first a brain capable of spotting and reproducing rhythmical patterns; a child can learn the difference between an apple and an orange, but it needs first a brain capable of retrieving sensorial information and categorizing. These innate conceptual apparatuses with which our minds and brains are equipped constitute intuitive systems of assumptions that made our hunter-gatherer ancestors fitter for coping with their pertinent environment, and were therefore naturally selected. Kant was the first one to formulate the idea that certain categorical distinctions are needed for any kind of reasoning to be even possible (1781). Today, they are called *intuitive* (or *folk*) *theories* (Barkow et al., 1992). Here is a list of some of the most important ones as summarized by Steven Pinker:

- *An intuitive physics*, which we use to keep track of objects and movement. Its core intuition is the concept of the object, which occupies one place, exists for a continuous span of time, and follows laws of motion and force.
- *An intuitive biology* or natural history, which we use to understand the living world. Its core intuition is that living things house a hidden essence that gives them their form and powers and drives their growth and bodily functions.

- *An intuitive engineering*, which we use to make and understand tools and other artifacts. Its core intuition is that a tool is an object with a purpose –designed by a person to achieve a goal.
- *An intuitive psychology*, which we use to understand other people. Its core intuition is that others are not objects but are animated by an invisible entity (mind or soul). Minds contain beliefs and desires and are the immediate cause of behavior.
- *A spatial sense*, which we use to navigate the world and keep track of where things are.
- *A number sense*, which we use to think about quantities and amounts. It is absed on an ability to register exact quantities for small numbers and objects (one, two, three) and to make rough relative estimates for larger numbers. That's why we all tend to "round" numbers.
- *A sense of probability*, which we use to calculate the likelihood of events. It is based on the ability to track the relative frequency of events.
- *An intuitive economics*, which we use to exchange goods and favors. It is based on the concept of reciprocal exchange, in which one party confers a benefit on another and is entitled to an equivalent benefit in return.
- *A mental database and logic*, which e use to represent ideas and to infer new ideas from old ones. It is based on assertions about what's what, what's where, or who did what to whom, when, where, and why. The assertions are linked in a mind-wide web and can be recombined with logical and causal operators such as AND, OR, NOT, ALL, SOME, NECESSARY, POSSIBLE, and CAUSE.
- *Systems connected to the emotions*, which give us models for ethics, for assessing danger and fear, contamination, disgust, etc.
  (Pinker, 2007)

These folk theories are the shared intuitive systems of assumptions and drives that humans have inherited from their ancestors. They can be observed operating in any human infant. Psychologists used to think that the world of children was a sheer "blooming, buzzing confusion" (James, 1890). But now we count with empirical evidence of the fact that, for instance, when a child sees the left and the right side of a puppet, the child doesn't think of these perceptions as corresponding to two different objects, but it manages to recognize that it is always the same puppet, seen from different angles (Rosch&Lloyd, 1978). This shows that the child is already equipped with a capacity to conceptualize the identity of middle-size objects (which is part of our intuitive theory of physics). And the same occurs with all the other intuitive theories.

As the evolutionary account would predict, these mental systems are certainly useful: they give us an intuitive map of the reality in which we are embedded that allows us to cope with it in efficient ways. we prove the utility of our folk theories every time we grab a glass assuming it is a solid object (independently of our formal education on physics) or when we greet our neighbor assuming she is not a robot (independently of our formal education on psychology). And this is even more obvious with regards to our evolutionary acquired senses and instincts: For example, our capacity to sense thirst serves to help us keep our body hydrated, and our evolutionary acquired phobias –e.g. to heights, spiders, snakes, etc. – are useful to maintain us far from danger.

This, of course, does not mean these systems (folk theories, senses, and instincts) are *always* useful. After all, evolution is a massively slow and costly ongoing process of trial and error. Moreover, our evolutionary acquired mental traits were formed in relation to the environment that was pertinent for our ancestors; but the pertinent environment has changed in many ways, which often makes our traits produce undesired secondary effects. For example, being food a scarce valuable resource –as it certainly was for the humans that determined our evolutionary constitution–, having a drive to overeat must have been an adaptation useful for accumulating energies and surviving periods of scarcity. However, having a supermarket open 6 days a week close to our homes –as it is the case in modern cities–, this instinct leads to obesity. This occurs because cultural evolution is faster than biological evolution: humans change their pertinent environment faster than evolution can change their biological constitution and drives. In consequence, many of our evolutionary adaptations that were useful for our ancestors can be today useless or even disadvantageous. For instance, people are biased to overestimate the risk of travelling by plane and to underestimate the risk of travelling by car, because heights were a pertinent risk for our ancestors but high speeds were not. Our taste for sugar was a source of energy crucial for the survival of our ancestors, but a candy-based diet –which is only since recently massively practicable– can lead to diabetes. And our sexual instincts might lead us to feel attracted to bodies fit for hunting and foraging, even though those are no longer the most required skills in modern societies.

We can hardly decide what drives we will have; nevertheless, we have a considerable control on what we do out of it. Our instincts give us automatic responses that function as short-term solutions whose efficiency has been tested through evolutionary time. For instance, if a lion appears suddenly in front of us, our mind experiences fear and that triggers automatic response (*fight or flea*). This instinct is useful is giving us a probabilistically efficient response to a situation where carefully considering all the variables and outcomes in a rational might be counterproductive. The efficiency of this evolutionary acquired response is also evidenced by the fact that many animals share, in fact, this same instinct. However, in relation to particular situations (e.g. facing a boss or a jury) this might not be the best long-term solution. In consequence, a capacity to think rationally on a long-term basis, and to repress thereby particular instincts, was also naturally selected in humans. Therefore, our bodies will desire at some point to overeat a delicious dish beyond the point where our hunger has been satisfied –that is the instinctive dimension–, but it is still possible for us to assess that this might be an inconvenient decision and overcome it, so as to stay in shape. And just as are capable of restraining our instinctive hunger, we are also capable of spending eight hours a day in front of a digital screen doing calculations for somebody else, even if our backs hurts; we are capable of repressing the drive to commit a hateful action, on the basis of a socially-agreed laws; we are capable of eating against our taste on the basis of nutritional or moral convictions; and, even if evolution programed the awakening of an impulse to parent in a given period of our lives, we are capable of evaluating the convenience of that decision and rationally decide not to.

Our folk theories are also not perfectly accurate in the information they give us about the world. Our mental software is full of bugs. For instance, our vision can easily be tricked with optical illusions, our folk psychology can lead us to believe that minds can exist without bodies (*soul*), our intuitive sense of probability is statistically inaccurate (it is biased to believe that past results can affect future chances), and our sense of logic can be tricked with rhetoric techniques (as we will see in detail along the first part of this book). These things occur because evolution only works on the basis of efficiency, not of accuracy. A direct evidence of this can be found in optical illusions. Here is one of my favorite ones:

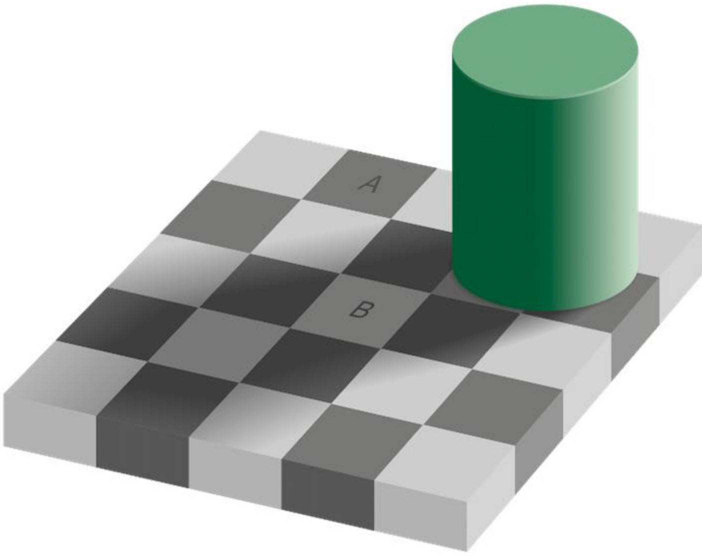


Figure 2. “Checker Shadow Illusion” (optical illusion), by Edward Adelson (1995)

We all perceive the square A as darker than the square B. However, they have both the exact same tone –you can check it by covering the squares around A and B. The reason why our mind fails to distinguish the actual tone has to do with the way we process 3D objects. Our minds interpret the middle of the board as shadowed and the extremes as illuminated, and, in consequence, our minds automatically compensate this contrast by making us see A as lighter and B as darker than they actually are. The point is that this automatic process carried by our vision is clearly flawed. But this does not trump evolution: for a sense (or any other trait) to evolve, it must only be more advantageous than it is counterproductive. In other words, it must only be *efficient enough* to outperform the alternatives. So, since neither our ancestors nor ourselves lived among *trompe-l'oeils* such as this one, the imperfect way in which our brains processes visual information is certainly very efficient.

Likewise, as scientific sources of knowledge, our folk theories can also be very misguiding and unreliable. When trying to make sense of the world beyond our intuitive observations –using more reliable techniques for measuring and contrasting hypotheses– we realize that the reality is much more complex that it first appears to our naked senses. Our minds and brains are directly equipped for hunting, gathering, forming social groups, finding shelter, and building tools; but not for exploring the quantum level or astrophysics, because that is not our most immediately pertinent environment. We are designed to perceive objects as solid masses of matter, and it is only through a huge intellectual process that we can get to conceive that matter is made of atoms –which actually have more void than matter. And, even when we intellectually understand that idea, we conveniently ignore it in our everyday handling of objects. That is why most scientific ideas are *explained* through concrete metaphors that make them more easily graspable to our intuitions: The notion of the atom was famously described using the image of a plum pudding (Joseph Thomson); the theory of general relativity, with the thought experiment of a falling man inside a lift (Albert Einstein); the grammatical structure of sentences is represented as trees (Noam Chomsky); and the mind as a computer (Hilary Putnam).

A clear example of the point I am trying to make here is precisely in the optical illusion we have just seen: we logically understand how the illusion works, we have obtained objective evidence of the fact that the two squares have actually the same tone, we *know* they do, but when our eyes look at the image, we still can't help but seeing the A square as darker than the

B square. This is a bug of our visual sense that anybody can verify; but, since it did not entail a great disadvantage for the survival and reproduction of our ancestors, our intuitive systems are not prepared to even *see* it. We can use technology and reason to obtain a more precise description of the world than the one allowed by our intuitions, but when we cope with the world in our everyday lives, we still do it mostly by following our intuitive systems rather than our scientific theories.[6]

Our mental and neural endowment also includes a capacity to process language. How can evolutionary explain our language faculty? This is one of the most debated and complex topics in evolutionary psychology. Here are two of the most important current hypotheses.

Noam Chomsky has recently published a work, along with and Robert Berwick, where they lay out their hypothesis on the origins of human language (Berwick&Chomsky, 2016). Even though the most evident function of language has to do with communication, Chomsky and Berwick (C&B) consider that this was probably not the original adaptive function that made our language-capacity be naturally selected. C&B consider that what an evolutionary account of language must explain is the acquisition of a particular mental schema: our universal grammar (UG). If we effectively count with a distinctive innate UG, this schema must have emerged as a result of a genetic mutation in the history of our species. This schema would have not been an actual language at this point, but the condition of possibility of a language, a mental mechanism that would allow us to processing information in the way that it is required for creating languages. Even at this stage, this capacity must have already entailed some particular advantages –which are difficult to define from today's perspective–: C&B hypothesize that it could have allowed humans to mentally label concepts and combine them into more complex ideas in a recursive manner. Also, C&B consider that it must have probably been especially useful for increasing the person's memory and competence for planning. On this basis, the fundamental cognitive capacities that enable language development would have been naturally selected to evolve. Once many people in possession of these capacities would have found themselves coexisting, the invention of an external code to communicate these mental operations would have finally made sense. Thereafter, the first natural languages would have been born.

As it can be seen, in this account, only the individual mental capacity to acquire language is an evolutionary adaptation, and its fundamental function would be the enhancement of other cognitive capacities. The communicative function that became manifest with the creation of natural languages would have been secondary –what, in evolutionary terms, is called *exaptive*. However, as said, this is still a contentious subject. A recent alternative model to explain the evolution of language has been proposed by evolutionary and social psychologist Michael Tomasello (2010), who puts socialization and communication back in the table.

As mentioned in the Prologue, humans have the cognitive capacity for mindreading, that is, for imagining the mental states of other individuals –which is an essential component of our intuitive theory of psychology. In a series of behavioral experiments performed in Leipzig that compared human infants with nonhuman primates, Michael Tomasello showed to what extent this capacity if far more developed in humans. These differences in the mindreading competences result in differences in the competence for engaging in tasks of *shared intentionality* (i.e. team work) (s. Tomasello, 2010), one of whose paradigmatic cases is linguistic communication. According to Tomasello, it would have been this mindreading

[6] This phenomenon has great consequences regarding the relationship between intuitions and reason, which will be dealt with more in detail in the chapter on Emotions.

capacity the key adaptation that enabled humans to develop language –and the underdevelopment of it what impedes human-like language development in other animals. In this view, communication would have been the fundamental advantage –and not a secondary function– that made our language-processing capacity be naturally selected.[7]

It is discussed whether human language is the result of an adaptation or an exaptation, but what it is agreed is that our language capacity is universal in our species and that it is an evolutionary legacy. As we will see more in detail in the following chapters, exploring the adaptive rationale that would have led to the constitution of our mental and neural endowments does not only have a historical interest, but it is crucial for understanding particular properties of our cultural systems and of our behavior in relation to cultural practices, such as why human languages and literatures are the way they are: Why are rhymes and metaphors used so extensively in human literatures? Why does every culture have stories? Why do we structure information narratively in the first place? Why do we systematically judge fictional characters with moral criteria? Why do we get emotionally engaged with fictional stories, etc.? Many of the answers to these questions are to be sought in our evolutionary history, in the adaptive rationale that formed all the natural intuitions, instincts, and biases we display when creating or consuming literary works.

HUMANS HAVE CULTURE, NATURALLY

The naturalistic perspectives of cognitive psychology, neurocognitive science, and evolutionary psychology, when considered in an integrated manner, give us an encompassing description of the universal nature of the human mind: they tell us how the minds and brains of our species work and by virtue of what processes they came to be as they are. This does not mean that these descriptive models are blind to the diversity of human cultures. On the contrary: Chomsky's universal grammar is not an attempt to ignore the diversity of human languages and expressions, but an attempt *to explain it*. As it was mentioned, the fundamental problem that Chosmky considered a linguistic theory should account for was *human creativity*: How is it that, with finite elements (some rules and a list of words) humans can produce infinite sentences? That is, the staggering variety of human languages is part of the problem that the theory is meant to answer in the first place. In general terms, the challenge assumed by these cognitive disciplines consists in finding the common biological basis that allowed for the emergence of the diversity of human cultures and practices. In consequence, the analysis of cultures is to be included as a complementary level within this naturalistic account.

So, to sum up. Natural selection allowed for the evolution of brains. The neural connections in the brain tissue allow for information processing. The emergent phenomena resulting from this processing of information is what we call minds. And the interaction of minds in a society produces a second level of emergent phenomena that we call culture. In this manner, these disciplines provide us with an epistemological synthesis between the natural and the cultural domains, between the world of atoms and the world of poetry. This great epistemological achievement of humanity hasn't gone unnoticed by scientists and philosophers. Biologists E.O. Wilson called it The New Synthesis (1975). And physicist David Deutsch called it Theory of Everything (2011).

Once a discipline is integrated with others at lower and higher levels, we can extend the conclusions found in one to further explain phenomena on the other and thereafter gain

[7] We will discuss the notion of *mindreading* more in detail in further chapters, especially in the chapter on Characters.

consilience (that is, evidence-based agreement among disciplines). For instance, the physical model of the atom, at a lower level, gave us further insights about chemical properties, at a higher level, which allowed us to verify and correct many of our chemical theories. Likewise, the consideration of cultural phenomena from a naturalistic point of view has opened the path for new approaches to the cultural domain that bring up scientific models, tools, and insights for describing its properties in increasingly accurate ways. Particularly productive on this regard has been Richard Dawkins's notion of *memetics*, as a way of conceptualizing the dynamics of the cultural domain in compatibility with natural selection and information theory. I would like to make a brief comment about it.

The term *meme* refers to the informational units in which what we consider culture can be devided: words, song, techniques, rituals, stories, typographies, names, etc. The argument of the so-called Memetic Theory is that memes, due to their formal constitution (as information imperfect replicators), are also bound to natural selection (Dawkins, 1976; Dennett, 2017). This theory understands evolution not as a contingency of living organisms, but, more deeply, as a mathematical principle: in any system made of replicating units that inherit some traits plus a percentage of variations, the traits that increase chances of survival and reproduction will in time outnumber the alternatives. And not only genes do this: also recipes, jokes, and ideas replicate from mind to mind (by means of imitation), and some of them survive longer and replicate more than their alternatives: not only viruses but also melodies are contagious. Given this formal constitution, memes are bound to evolve.

The fact that culture works in this evolutionary manner is, in a way, much more obvious to our intuitions than the case of our biology, because cultural evolution is much faster and flexible: in a human lifetime, one can observe the evolutionary change of fashion trends, musical styles, and even political systems. And it is enlightening, on this regard, to notice the fact that Darwin himself possibly arrived to one of his most famous conceptualizations of evolution precisely by paying attention to a case of cultural evolution: language.

Philologists in the 19th Century worked intensively in the trying to reconstruct the history of the languages. When attempting to map the complex similarities and differences among languages, in relation to their change over time, an old idea that became then highly relevant was the *tree structure*, which resulted in the *Family-Tree Theory*. Even today, we conceptualize the evolution of human languages by representing them in family-trees. One of the most influential linguists in popularizing this model was August Schleicher. He made his first publication in 1853:[8]

[8] Schleicher's work was groundbreaking in the field of Indo-European comparative linguistics –of whom he is one of the most important founders. The term *family-tree theory* comes directly from the term he employed in this work: *Stammbaumtheorie*.

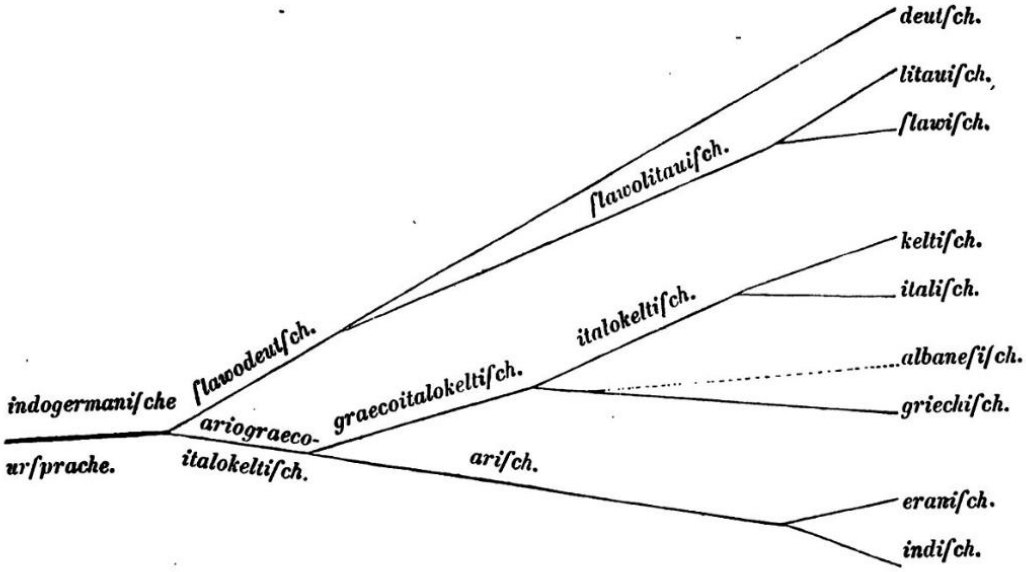


Figure 3. Example of the language family-tree model (Schleicher, 1853)

Only Six years later, in his seminal work (*The Origin of the Species,* 1859), Darwin would famously describe natural evolution with an analogous image: The *Tree of Life*, which applies the same structural principle of language-phylogeny to the relationship between species. The similarity is patent in Darwin's own designs:

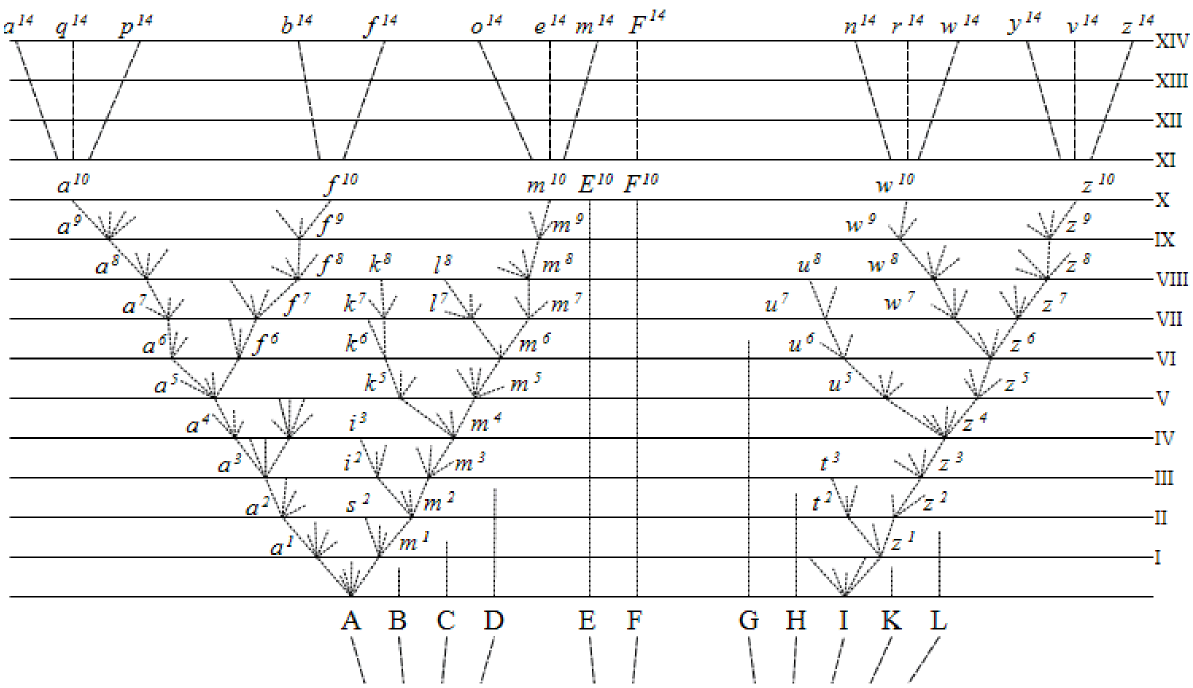


Figure 4. The *Tree of Life* diagram. (Darwin, 1859)

Both Schleicher and Darwin were aware of this structural similarity between cultural and natural evolution, and they even met personally and reflected upon the subject explicitly.

Schleicher expressed his idea on *Darwinism Tested by the Science of Language* (1869). And Darwin expressed it in his own work:

> It may be worth while to illustrate this view of classification, by taking the case of languages. (…) The various degrees of difference in the languages from the same stock, would have to be expressed by groups subordinate to groups; but the proper or even only possible arrangement would still be genealogical; and this would be strictly natural, as it would connect together all languages, extinct and modern, by the closest affinities, and would give the filiation and origin of each tongue. (Darwin, 1859)

Darwin understood the common link between culture and nature. But he understood it in a deeper sense: by realizing that culture (a universal practice among humans enabled by a set of biological endowments) is also a phenomenon of nature –idea that he developed in his later work (*The Descent of Man*, 1872), which pioneered evolutionary psychology one century before it was rediscovered.

Culture is a phenomenon of the natural world that emerges from thinking living organisms. Culture even exists in the natural world beyond humans. Social learning and culture has also been found across different species of primates (White, 2000): biologists have found many cases of communities of primates that developed a particular technique (especially in relation to tool design and usage for foraging) which is shared by the members of that community but not by other communities of the same species in other parts of the world. And the way in which primates maintain these community-specific techniques is precisely by teaching them to each other from generation to generation –just like we, humans, do it with our culture, our traditions, our memes. But the point I want to stress is that, underneath each of our cultural differences (among communities), lays a deeper shared natural capacity for culture (of our species), a capacity to create, consume, and pass on memes. And without an understanding of this fundamental commonality, we cannot get a full understanding of the differences.

Just like our language and any other of our cultural practices, literature and our literary experiences are also part of the natural world. There is a common natural ground that allows human all over the world to produce, consume, and pass on stories, idioms, poems, jokes, characters, and myths. In this sense, literature has also a place in the synthesis of knowledge described in this chapter. I have tried to argue that there is a logical path that can bring us from nature to culture, from atoms to poetry, by integrating the disciplines here presented into a coherent framework. In the following pages, we will see that, when applying these perspectives, many aspects of our most subjective responses to literature can become explainable (and often even predictable) when addressed as natural consequences of our evolutionary endowments, our biological brains, and our biased cognitive systems coping with the world in real-life social contexts.

CHAPTER 2

WORDS

HOW WE ENCODE AND DECODE MEANING

In an evening of 1962, Armenian-American author Aram Saroyan wrote seven letters (i.e. seven characters) on a piece of paper. Little did he know, that minimal text would become one of the most controversial poems in the history of American literature.

Editor George Plimpton selected it for *The American Literary Anthology* one year later. And, as all the other poets included in the volume, Saroyan received a monetary award for his text by the National Endowment for the Arts. However, unlike the others, Saroyan's poem caused a particular uneasiness in some sectors of the public opinion.

This negative reaction was soon voiced by politicians, such as Representative William Scherle, who launched a national campaign against the wastefulness of the NEA, objecting the fact that Saroyan's seven-letters text could have been awarded as a poem. As a journalist report summarizes: "Pretty soon, Michael Straight –deputy chairperson of the Endowment at the time– was personally called to the offices of 46 members of Congress to explain the matter, and mailbags of letters from fuming taxpayers clogged the agency's boxes, most of them variations on a theme: *We cannot afford to lower taxes but we can pay some beatnik weirdo to write one word… and not even spell it right?!*" (Daly, 2018).

Here is the full text of the poem written by Aram Saroyan (1962):

lighght

The word appeared exactly like that: isolated in the middle of the page, written in lower case letters, with that extra "gh" in the middle.

Many questions can be drawn from this historical anecdote: What does such a text mean? How is it supposed to be read? What is there in a word? How do we discriminate what is and what is not a poem? In this chapter, we will explore how cognitive science describes our capacity to process words and we will consider what insight this view can give us into this kind of questions.

RECOGNIZING LETTERS AND SOUNDS

Humans have had spoken language for at least 50 thousand years –and probably much longer–, but the earliest evidence of written language is only 5 to 10 thousand years old (s. Daniels&Bright, 1996). We cannot be evolutionary specialized for reading and writing, it is too recent a technology. Although most people in the world is today literate, there are still many countries where this is not the case. Unlike learning how to speak (which is a skill that our biology is programmed to develop), learning how to read and write requires especial lessons and a considerable amount of effort and time. Nevertheless –as you are proving it right now– we certainly are capable of learning how to read and write. But how do we do it? How can we learn new skills that our biology wasn't naturally adapted to do?

Neurologist Stanislas Dehaene formulated the theory of *neuronal recycling* so as to account for phenomena of this kind at a neurocognitive level (s. Dehaene, 2009). He proposed that our brains have the capacity to reuse (*recycle*) neural circuits –in particular, from the cortical areas–, which were evolutionarily designed to perform certain tasks, in order to learn novel –but structurally similar– tasks. Following this logic, all the culturally transmitted skills that humans are capable of inventing can be explained in a way that is consistent with our evolutionary history. Reading would also be one of these skills. When we read a word, several visual stimuli are captured by our photoreceptors. From this material input, our brains undergo a series of processes that ultimately allow us to retrieve meaning. Dehaene's concrete hypothesis is that, at this fundamental level of the reading process, our brains recycle mainly the cortical areas that originally evolved of object recognition –especially from the visual cortex to the left inferior temporal cortex (Ibid.).

So as to read, a brain must solve two fundamental problems. Firstly, it must be capable of identifying letters and words; secondly, it must associate those identifications with particular meanings. The first task is known as the *invariance problem*. Reading requires us to identify which are the aspects that do not vary of a letter or word, in spite of the infinite shapes in which these letters or words can be found. And we do this by recruiting our system of categorization (the same system that we use for learning how to recognize objects). Our minds search for patterns of commonalities and thereafter form criteria of membership on the basis of which we become capable of assessing whether an object is or is not of a given kind. Once we have a category such as *ball*, defined, for instance, as referring to "spherical objects," we become capable of recognizing tennis-balls and foot-balls as instances of the same kind, despite their noticeable differences (s. Rosch & Lloyd, 1978; Harnad, 2010). By means of an analogous process, we get to recognize A, A, and A as versions of the similar letters, by forming a mental category of the letter A. And, in some cases –as it is usual in contemporary writing, which makes such an extensive use of digital text processors–, our brains must also perform a further kind of categorization: encoding different typographic fonts as subcategories of the same mental address, so as to identify a, *a*, and *A* as still referring to the same letter. And we are so skilled at this that, without too much effort, we can be come capable of recognize words even when they are written with v3rY 5tR4ngE ch4R*a*cTeR5.

In languages like English, which use alphabetical writing systems (e.g. Latin, Greek, or Cyrillic), our brains learn to read by training the conversion of sounds into letters and vice versa. This creates very strong neural links between the mental representations of these letters (*graphemes*) and our mental representations of sounds (*phonemes*), which we acquired when we learnt to talk. And, actually, the two are active when we read. It is believed that, for this reason, silent reading must be a relatively modern practice –as some historical evidence suggests (Ibid.). But, even though adult readers are capable of reading in silence, still neurons in our auditory cortex get activated when doing it, and also in the parts of the motor cortex that we use for pronouncing the words we are reading (Ibid.). Even if we are not really verbalizing what we are reading, at a certain level our brains are. This neural activity even sends electrical signals to the muscles that we use for pronouncing the words in question. And, in fact, recently, NASA created a machine capable of tracing these electrical sings and decoding them into strings of sounds. The process is called *subvocal speech recognition* and, in a very direct sense, it is a way of reading people's minds. Extravagant as it sounds, the aim of this technology is a very concrete one: that astronauts can communicate in difficult situations –e.g. under high pressure or when surrounded by a lot of noise– without even moving their lips (Brakus&Bluck, 2004).

With psychological experiments some mental biases can be recognized that evidence the extent to which visual and auditory processing are entangled when reading. For example, considering only the spelling, try to identify which of the following words is a real English word (I take the examples from Dehaene, 2009):

rabbit, culdolt, karpit, money, nee

In this kind of experiments, people are requested to read each word and to state their answers. And what is measured is their velocity of response. What it was found in this case is that people systematically take more time to decide if they are reading a real word when the written word *would sound* like a real one (e.g. *carpet* and *knee*), even if explicitly requested to only pay attention only to the spelling (Dehaene). The explanation would be that *karpit* activates –via our mental representation of its sound– our mental address for *carpet*, which interferes with our realization that, actually, it is a different word. That is, it requires our brains to second-guess. This is a simple proof that we cannot help but processing sound even when we read silently.

Another proof of the deep entanglement between letters and sounds is our fcilitty ffour reedhing mizpxlled werds. However, this only seems to occur when the spelling alterations do not produce great alterations in the sound, as in the previous sentence. When the opposite occurs, reading the words becomes much more difficult. Try to guess which is the word misspelled in each of these cases:

iaughert, teexh, gvltar, beliuer, amolb[9]

Even though the words in this list *look* (graphically) very similar to the ones written in the footnote, their shapes correspond to very different *sounds*, and it would therefore be expectable that it takes longer for people to recognize these words. Indeed, the most usual spelling mistakes are not precisely cases of that kind, but instead the ones that alter the writing *without* altering the pronunciation (e.g. *it's* and *its*, *truly* and *truely*, your and you're, *holiday and holliday*, etc.). This seems to show that, although our minds decode visual and auditory cues simultaneously, the auditory cues have a certain priority for identifying the word –which actually makes sense, considering the oral evolutionary origins of our language faculty.

However, this property of our linguistic mind is more advantageous than hindering: The same mechanism that makes us commit spelling mistakes so easily is the one that give us the enourmous flexibility we have for easily recognizing anomalous spellings (as long as they preserve sufficient sound similarities with the words we know): such as when a text intends to imitate a particular accent (*Ziz is how Germans sbeak*), when people chat (*I luv u 2*), and when poets invent words (such as *lighght*). This already allows us to predict a particular reader-response regadring Saroyan's poem: trained English readers would not be able to help seeing the word *light* in *lighght*, and many of them would probably judge it is a plain misspelling –which is, indeed, what actually occurred: people discussed whether the misspelling was deliberate and what it could mean, but not whether Saroyan's poem referred to the English word *light* or to another word in an invented language, for instance.

[9] The answers are *laughter*, *teeth*, *guitar*, *deliver*, and *arnold*.

In any case, the reading process does not stop here. Our minds do not go directly from this step to the meaning of a word. The visual and auditory cues are further categorized by our minds in levels of increasing complexity. To begin with, it does not always occur –especially in English– that each single sound matches a single letter: Graphemes can be complex, such as the *th* in "the" or the *sh* in "show". However, the fact that we associate each of them with single phonemes makes us also encode these complex graphemes as units. Try to recognize which of the words in this list contain the letter *a* (Dehaene, 2009):

garage, metal, people, coat, please, light

It usually takes it longer for English speakers to recognize the letter *a* in *coat* and *please*, because, in those cases, the letter is embedded in complex graphemes (*oa* and *ea*) that we associate with single sounds.

Also the grapheme *gh* is one of these complex cases; nevertheless, a further phenomenon occurs with it, as well. What determines the sound represented by the grapheme is not only its shape, but also its position in the word. Our mind immediately decodes this so as to distinguish that, at the beginning of a word, the *gh* stands for the phoneme /g/, as in *ghost*; in the end, it stands for /x/, as in *laugh*; and, in most other cases, *gh* is either silent or it affects the sound of the previous vowel, as in *light* –where the *gh* allows us to distinguish the pronunciation of the *i* from the one it would have in *lit*, for instance. In consequence, recognizing whether the word *light* contains the letter *a* probably takes also longer than in the first words of the list, because, even thought it does not contain the grapheme *a*, its pronunciation does contain an *a* sound: /laɪt/.

At the same time, our reading minds also automatically parse the words into syllables. We tend to think of syllables as units that only poets care about (when counting meter), but every reading brain is sensible to them. Try to identify whether the middle letter of each of the following five-letters words is in normal or bold type (Ibid.):

List 1: HO**RNY RID**ER **GR**AVY LIG**HT**

List 2: **VOD**KA ME**TRO** HAN**DY SU**PER

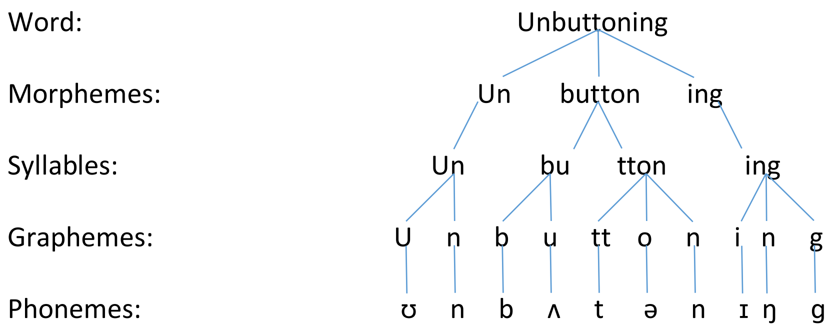
Expectedly, the task is easier to perform in the second list, because in those words the change in the font (normal vs. bold) matches the syllabic divisions of the words –which shows the extent to which our minds take this kind of parsing into account.

At the same time, our minds parse the words we read according to yet another criterion: morphemes. Morphemes are the smaller linguistic units of meaning. A morpheme is not the same as a word. Some words are made of only one morpheme (e.g. *cut*), but some are constituted by many: *unkindness* includes the morphemes *un-* (which means "") *kind* (which is an adjective, synonym of "polite") and *-ness* (which turns the word into a noun: "the state of"). The combining the meanings of these three morphemes (un-kind-ness) builds up the meaning we attribute to the word as a whole ("The state of non being kind"). As it can be seen, some morphemes can act as autonomous words ("cut," "kind,") but some others (suffixes) can only appear attached to others ("un-" "-ness").

Our minds also parse the words we are reading in accordance to morphemes. For instance, seeing *departure* written in a computer screen makes you faster in recognizing *depart* (cognitive scientists say that *it primes it*)[10]. However, out morphemic parsing does not respond strictly to visual or auditory cues: *can* primes *could*,–even though they look and sound very different–, whereas *aspire* and *aspirin* do not prime each other –even though they look and sound very similar. And our morphemic parsing does not respond strictly to meaning-related cues either: *hard* still primes *hardly*, and *depart* primes *department* –even though their meanings are unrelated. Dehaene's explanation of this phenomenon would be that our minds parse in accordance to the morphemes we already know, as if predicting that those units are likely to be pertinent for decoding the text we are reading, even if some times they might mean something else (like the *-ment* in *department*) (Ibid.).

The fact that our minds are trained for parsing morphemes is what allows us to easily make some sense of words we ignore –as long as they use familiar morphemes. If we read, for example, a made-up word like *cleavest*, even though we don't what it means, we can imagine that it might be an adjective meaning "the most *cleave*" (independently of what *cleave* might mean). Likewise, we can imagine that *asdfyxcvqwerality* might be a noun meaning "relative to *asdfyxcvqwera.*" Playing with the mental structures that lead us to make sense of inexistent words is a something that literary authors have done for a long time. In Spanish-speaking literature, for instance, there is a significant tradition of poems made of invented words, which are called *jitanjáforas* (a made-up word in itself, invented by Alfonso Reyes in 1929). The way in which these poems make sense is, most prominently, by means of morphemic markers that make us feel we are reading nouns, adjectives, or verbs, even though we ignore exactly what they mean. Saroyan's poem can be also considered a minimal example of this.

As we can see, each word is constituted as a tree of categorical levels of analysis that respond to the way in which our brains encode them and decode them. Dehaene illustrates these different levels with the example of the word *unbuttoning* (Ibid.):



When reading a single word, our minds and brains execute a complex processes of categorization, at different levels of analysis, and using different kinds of information. By considering the ways in which these processes work, we have started to gain some insight on what might happen in people's minds when reading Saroyan's poem. But what happens when, after having identified the written word, we are directed to the mental address of its content? How is the content of a word structured and processed in our minds?

[10] I am referring to the so-called *priming effect*: when a neural activation makes another neural activation more likely we say that X primes Y, which would be evidence of a neural pattern (i.e. a Hebbian link) between the two.

THE STRUCTURE OF MEANING: FUZZINESS, FAMILY RESEMBLANCE, AND PROTOTYPES

Words demarcate their meaning also through a process of categorization. This strategy has clear evolutionary advantages, especially in terms of economizing knowledge. Each of our categories includes a list of common features that characterize the members of that category. Therefore, once we identify an object as a member of a category (e.g.: "This animal has a beak, it must be a *bird*"), we access a mental data basis of information encoded in the category, that allows us to anticipate information about the object ("If it is a bird, it may lay eggs, fly, sing, etc.").

But how are these categories structured? How do our minds form the concepts by means of which we sort different things as members of categories like *bird*, *light*, or *poem*? The classical view assumed that our mental categories were constituted by logically precise binary definitions. Mathematical functions are a good example of this: The category "even numbers" is formed by a precise definition ("being divisible by two"), which permits us to establish which number is even in a binary way, that is, without ambiguity nor vagueness. However, this conception of categories is problematic when applied to other words; since, as we know, ambiguities and vagueness abound in human language.

A clear example of these problematic cases is the word *bachelor*. According to the definitional view, a category like *bachelor* would be formed by a definition like "unmarried man" –the words *unmarried* and *man* would also be, in their turn, formed by clear definitions of this kind. If that was the case, there should be no doubt as for what qualifies as a bachelor: it would be sufficient to establish whether somebody is effectively a man and is not married. However, if we observe how people judge what is and what is not a bachelor, we see that a different thing occurs: There are many cases of men that abide to the above-given definition of *bachelor* but people still do not consider them as bachelor (e.g. Adam, The Pope, male children), and at the same time, there are men that do not abide to the definition but still are considered as *bachelor* (e.g. a man in an open marriage or a man married for exclusively legal reasons but not emotionally engaged). Figure 1 illustrates this:

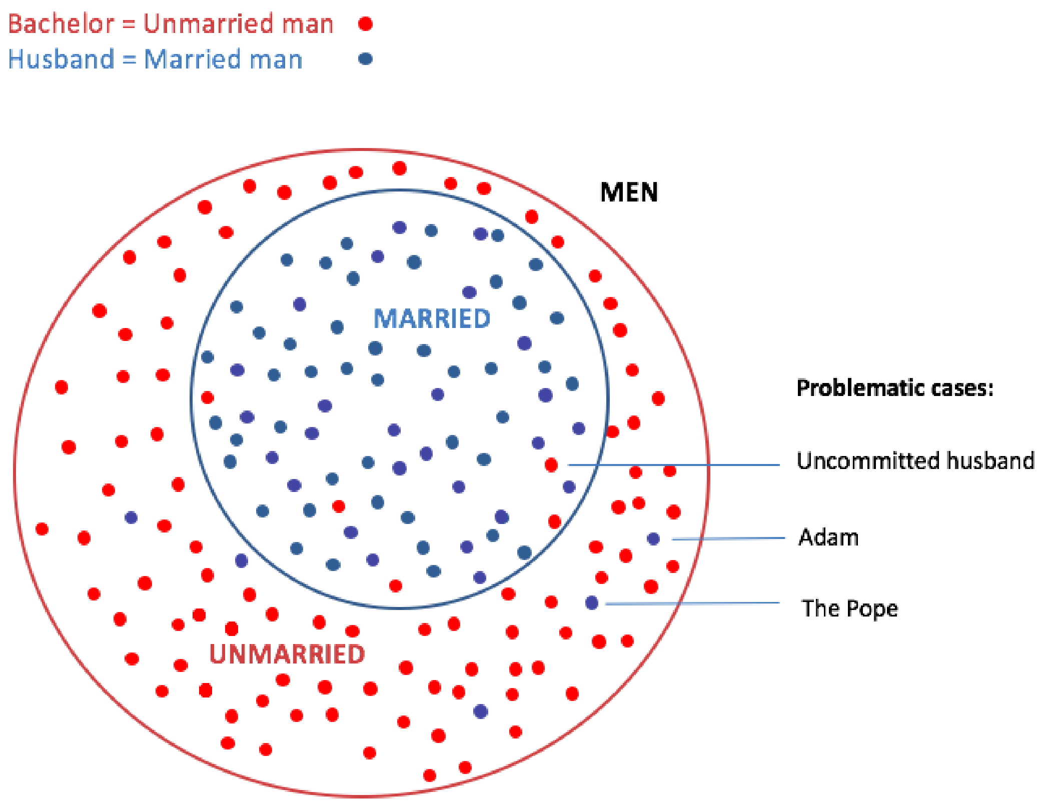


Figure 1. Definitional version of the categories *bachelor* and *husband* in terms of civil status (binary membership)

The same problematic cases can be found around most of the categories we use in our every day lives: If a 4-year old child takes a soap from a hotel, is he or she *a thief*? If all of my grandparents were Italians, but I was not born in Italy, I've never visited Italy, and I do not speak Italian, am I *Italian*? If a woman gives birth to a dead child, is she a *mother*? If Saroyan text has only one word (and misspelled), is it *a poem*? As surveys verify once and again, people doubt and disagree with regards to many questions of this kind. But, if words' definitions were really binary, there would not be such doubts: each of us would have a sense of perfect clarity with regards to what qualifies as a thief, an Italian, a mother, or a poem. Since in many cases we do not, it means there must be something wrong with the classical descriptive model. Due to these inconsistencies, the classical view was in last decades challenged by Eleonor Rosch and Roger Lloyd (R&L, 1978). Inspired by Wittgenstein's conception of language (1953), they proposed another mechanism to explain the structure of categories that can account for these problematic cases.

R&L made a famous experiment to show these kinds of semantic ambiguity are not eventual, but actually constitutive of many categories. They asked several people to qualify "How good an example of a bird is… X," presenting the test subjects with pictures of different varieties of birds. If categories were binary, people should consider every exemplar as "a perfect example" or "a wrong example" of any given category. But the experiments revealed that people consistently considered degrees of membership instead. They considered, for instance, that a robin is "a very good example of a bird", a chicken is "a mediocre example of a bird," and a penguin is "a bad example of a bird." What was perhaps even more revealing was that the results were also consistent with regards to the degrees of membership attributed to each example (all the participants considered the robin as a good example and the penguin as a bad one, for instance). Now, the very fact that people can assess "goodness of example" shows the inadequacy of the classical view –if categories were constituted by precise definitions, no member would have any special status.

On the basis of this experiment, we came to a better understanding of the ambiguity of categories, which was conceptualized by R&L as *fuzziness*. Fuzziness refers to the fact that, at least some categories, have degrees of membership and no clear boundaries. Categories like *bird* or *poem* are fuzzy categories. The term was taken from the mathematical model of fuzzy sets: sets that are not binary, and to which elements can belong by degrees, by moments, etc.[11] That means that you don't think of poems and birds by recruiting an binary logical definition of *poem* and *bird*. You think, instead, of different objects you find in the world as resembling "more or less poem-like" and "more or less bird-like."

Once the phenomenon was identified and defined, it had to be explained how it worked and why: If categories are not constituted by binary definitions, then how are they constituted? What kind of structure can account for these fuzziness-effects? As a solution, R&L proposed the Family Resemblance model and the Prototype Theory. Their claim was that a category, instead of a precise definition, would rather be a collection of commonalities shared by different members of the category, and this collection would often assume the form of a *mental prototype*.

Figure 2 represents a prototype of bird (central member). This prototype is a mental construction that aggregates the set of traits that are frequent (or salient) among real birds. The utility of this prototype is that it allows us to categorize the rest of the birds we may find as more or less bird-like, in accordance to their similarity to this prototype –which in the graph is represented as distance from the center.

Figure 2. Prototype version of the category *bird* (fuzzy membership).

[11] It is however Lotfi A. Zadeh who has been credited for first applying the notion of "fuzziness" to concepts ("fuzzy concepts") in a paper of 1965 where he attempted to give a mathematical account of the phenomenon.

The mechanism by means of which our minds select and organize the commonalities of a series of object and condense it into a prototype is described by R&L as *Family Resemblance*, because it echoes the way we think about members of the same family. Imagine that you encounter the González family (see Figure 3). After meeting some of its members, you will mentally select some of the recurring properties that you observe in them: most González have big noses, dark hair, big beard, no moustache, wear glasses, etc. None feature has to be necessarily shared by all the members, but as long as a feature is salient, you will tend to select it, and you will build thereafter a prototype of the González-family member. This prototype needn't exist (is not a real family member), but it would be a mental generic representation of an individual made out of the average features across the González you know. Then, when you meet a new person, you will assess if this person looks more or less González-like, by measuring how much it resembles your mental prototype.

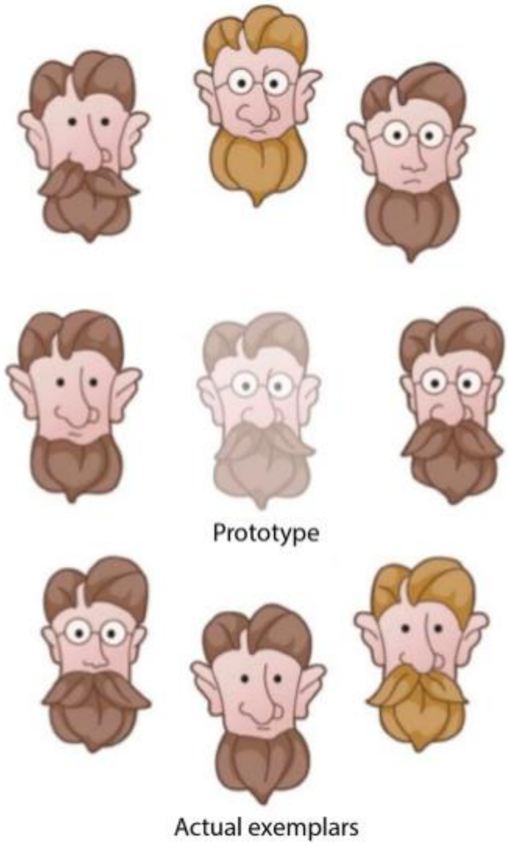


Figure 3. Family-resemblance categories. E.g.: The González Family.

As in the case of the González family, in many cases –especially when referring to material objects– a family resemblance category would actually constitute a *visual* prototype: a mental image of how the representative case "looks like." This is certainly the case of categories such as *bird*, *chair*, or *ball*. R&L called them basic-object categories, and they claimed that they are one of the most fundamental kinds of categories of human cognition. Basic-object categories are among the first categories that children learn. And examples of these categories abound in our every day lives: If one would look up for the word *bird* in an encyclopaedic dictionary, one would probably find an illustration of a bird designed with generic features (such as the one in Figure 2). That generic bird need not exist –in reality one can only find concrete robins, parrots, doves, etc., not generic birds. But, by sharing the most common traits of concrete birds, that prototypical drawing gets to function as a parameter of membership, as a category.

We can now return Saroyan's poem: Is that word a poem? Clearly, it is typical of poems to have more than one word. Moreover, since poetry is considered a verbal art, it is typically associated with qualitative writing skills –the opposite of what is associated with misspelling, which is an interpretation that Saroyan's poem effectively elicited. It would be possible to perform psychological experiments to measure the relative weight that people attribute to these and other variables (length, spelling, lexical variety, profusion of metaphors, etc.) when judging whether a text is or is not a poem. But people do not only hold diverse prototypes of poems, they also attribute different *prototypicality* to different aspects of the prototype: Even when agreeing that Saroyan's is atypical in relation to the prototype of *poem*, this might mean different things for different readers. For some, it may mean that, therefore, *it is not a poem*. For others, the text would still qualify as a poem because it has the social credentials of a poem: it was written in a poetry anthology, it was written by a poet, and it was awarded. Moreover, other readers might see the very atypicality of the text as a feat meant to make us wonder "what is a poem?," and some may consider this effect as being intriguing, funny, or shocking enough to make them categorize the text nevertheless as *a poem*.[12] The bottom line is that the diversity of people's interpretations does not mean we are unpredictable, but only that we are complex. Indeed, at least some of this diversity is explained by the fact categories are fuzzy.

In addition, there is a variable that can be used to further analyze why some categories are fuzzier than others: institutionalization. Institutionalization is the process by means of which a concept becomes conventionalized within a society (s. Berger&Luckmann, 1966; Searle, 1996). The value of money, the legality of a contract, and the definition of technical words (such as *biometrics* or *gigahertz*) are highly institutionalized; whereas the beauty of painting, the funniness of a joke, or the definition of colloquial words (such as *friend* or *sandwhich*) are categories with much lower levels of institutionalization. My hypothesis on this regard is that the more institutionalized a field becomes, the less fuzzy categories behave when used in it. And this can be clearly seen when considering words that exist both as a technical term and as a colloquial one. For example, the word *demand* is used both in economy and colloquially. In the first case, its meaning is very clearly defined ("Consumers' willingness to pay for a specific good or service"). But when used colloquially, it means (among other things) "to ask for with authority," and it is applied in a much fuzzier way: e.g. If someone asks you for a direction in the street without saying "hi" nor "please…" Is that a *demand*?

This principle seems to be a necessary consequence of the nature of meaning: most words have many meanings, and the technical meanings are only a subset of these. Therefore, any specialized meaning will be necessarily more limited than the sum of all the other possible meanings of that same word –as it can be verified by reading any dictionary. If this hypothesis is correct, it would be expectable that, if R&L's experiment would be replicated with professional biologists as test subjects –who were instructed to act as such, that is, abiding to the technical concepts established in their discipline–, much lower fuzziness effects (if any at all) would have appeared in the results. This is, indeed, something that could be tested.

Some important conclusions can be drawn for this. We do not cognize words with formulaic definitions, but by constructing mental prototypes. These prototypes assembly commonalities: that is, features that we consider typical of the elements of the category. And

[12] Unlikely as it may sound, this line of reasoning is the one that grounded the so-called *Anti-Poetry* movement, lead by writers like Nicanor Parra and Elias Petropoulos: a movement that produced atypical poetry, which (expectedly) most readers do not find poetic.

we assess the membership of new elements to the category by considering how similar they are to the prototype. These features are hierarchically ordered: we consider some of them as more or less defining of membership –and this, of course, may vary across cultures, time, and individuals. This functioning produces expectable fuzziness effects. This is the first characteristic to have into account when trying to make predictions of what meanings people attribute to words: People do not interpret literary works as "romantic" or "not romantic," as "offensive" or "not offensive," as "poetic" or "not poetic": but as more or less romantic-like, more or less offensive, and more or less poetic. We do not experience the meanings of words as rigid definitions, but as fuzzy clouds of resemblances and differences across mental representations. And this theory provides a crucial insight not only for understanding the disagreements that can exist around what qualifies as a poem, but also for understanding much more serious social debates constituted around the fuzzy boundaries of words such as marriage, citizenship, war, democracy, and many others. By means of which prototypes do we cognize those words? What parameters do we use for discriminating what abides and what does not to the categories labelled by these words?

COGNITIVE BIASES OF CATEGORIZATION

We have said that, in principle, we identify commonalities across objects and thereafter we create prototypes that we use to categorize other objects. But this description seems to portrait humans as if we were professional statisticians that collect data, make a probabilistic averaging, calculate the corresponding p-values, register the results, and the apply them to objectively classify the world. And this is not how our minds work.

Our mental system of categorization is designed to allow us to predict likely information in a fast way. Our category of *bird* allows us to automatically assume that if an animal has feathers and a beak, it will probably fly. It is efficient in giving us useful estimative map of the world, but it is not designed to be scientifically accurate. In consequence, it systematically produces predictable distortions that become evident through psychological experiments. These systematic biases affect how we constitute categories and how we apply them.

On the one hand, the prototypes your minds create are significantly biased by your personal experiences. If some features are salient to your direct experience of an object, you will be more prone to pick up those features as part of your prototype, even if you know they are not really representative or meaningful of the class in question. People that live in a homogeneously black community, for example, will be more prone to have a black prototype of "man," even if they are rationally aware that other ethnicities exist. Likewise, if most poetry one has read has rhymes, regular metric, and correct orthography, one would intuitively expect to find these features in a poem, even if one is aware that not all poetry necessarily abides to these parameters. In this sense, it is expectable that people would consider Saroyan's poem *lighght* immediately as non-poetic in reason of its atypical length and orthography, for example, even if none all the people that argued this would have answered that length and orthography are criteria they really use to judge whether a text is a poem.

On the other hand, the constitution of each prototype is affected by other prototypes of our mental system of categories. That is, prototypes influence each other, across categories. For example, the prototype of the category "husband," can affect your prototype of a supra-ordinate category (such as "man"), of a sub-ordinate category (such as "French husband"), or of a contrastive category (such as "wife"). If your prototype of "husband" would include being "drunk," that could bias you into thinking of "men" in general as being more alcoholic, of "French husbands" as being "wine enthusiasts," and of "wives" as being less alcoholic than the actual average of the wives you have observed. Likewise, if in one's mind the prototype of "poem" is defined as opposed to "prose" –and if we consider *being narrative* as a typical property of prose-, then we would expect poems to be less narrative than the actual average of poems we have read. This might lead us to judge stories in verse (such as the *Iliad* and the *Divina Comedia*) as non-prototypical poems. The case of expected orthography is also accountable in this terms: If one's notion of poetry is conceived as opposed to colloquial or informal language, and bad orthography is associated with informal language, then we would expect poetry to be more scrupulous with the orthography than the poetry one has read actually is. This explains the fact that readers were more outraged to find misspelling in Saroyan's awarded poem –bringing the observation to a Congress debate- than people typically are when finding misspelling in other contexts (e.g. chats, letters, graffiti, newspaper).

Many other biases result from the nature of our categorizing minds. We have salience-biases: we tend to assume that the more salient the property of a prototype is, the more likely it is to be shared by the members of the category ("If this is a poem, it will express beauty"). We have over-attribution biases: we tend to think that an object is more similar to others than it actually is, only because it we have placed it in the same category ("Now that I know this is a poem, I see there is some beauty to it"). And we even have normative biases: we tend to think that the properties collected by the prototype do not represent what the members effectively have in common but what they *ought* to have in common, by virtue of belonging to the category ("To be a proper poem, it *must* be beautiful") (Social psychologist Jonathan Haidt calls this the instance of the *the sacred*, s. Haidt, 2012). In this sense, if a property like *beauty* was part of our prototype of poem, it would be expectable that some people find an unsuspected beauty in *lighght*, once disposed to read it as a poem, that they would not find if they would see it as the name of a lamps brand, for instance. And it is also expectable that people that doesn't find beauty in *lighght* get to consider it not only non-poetic but even offensive.

Writers, musicians, filmmakers, stand up comedians, politicians, journalists, and marketers make use of these bias constantly, so as to better appeal to their audiences and guide their subjective experiences. These biases are part of the material of every author, because the way we perceive a story or a poem is not determined by the independent meanings of words that we can find in a dictionary, but it is the result of the ways in which these words are processed by our subjectivities, how they activate our mental categories and how they make our prototypes interact with each other. The nature of our mental categories and the biases they introduce to our interpretation of a text have also consequences for the ways in which we respond to rhetoric forms, cognize stories, judge characters, and react emotionally, which will be explored in the following chapters.

THE NEURAL GROUNDING OF MEANING: EMBODIED COGNITION

How is the process of cognizing words implemented in the brain tissue? How do neurons encode and decode the information of the words we know (our *lexicon*)? These questions are studied in neurolinguistics. By using different measuring technologies, neurologists are mapping the neural circuits that are responsible for processing the meaning of words. A crucial concept in relation to this is *embodied cognition*.

The notion of embodied cognition emerged in the 1970s (v. Varela, Thomson, & Rosch, 1993, and it is the theoretical attempt to integrate in a single model our understanding of the mind and the body, which have been traditionally considered as opposite dimensions. A crucial pioneer experiment bumped into embodied cognition by accident.

It happened in the 1990s in Parma, Italy (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). A group of neuroscientists was doing research with macaque primates. They studied how their neurons controlled their body movements. They did it by introducing electrodes in the brains of the monkeys, targeting individual neurons in motor regions, and measuring the electrical activity that they received when the monkeys performed different motor tasks, such as grabbing bananas. The remarkable finding occurred when one of the scientists took himself one of the bananas that were used for the experiment. At that moment, electrical activity was registered in one of the monkeys' brain in the exact part of the motor region that got activated when the monkey grabbed a banana itself. But the monkey was not grabbing anything this time, it was only observing the scientist doing it. What was discovered then are the so-called *mirror neurons*: neurons that encode both the performance and the recognition of specific actions.

These results lead to a provocative hypothesis about the way our brains might encode and decode meaning. Maybe we have a mirror system that allows neurons that we use to perceive and act to also process the recognition of perception and actions. Maybe the words we use to name what we feel and act are processed by the same parts of the brain that actually perform the feeling and the acting. (S. Lakoff&Johnson, 1999).

Experiments with humans –where their brains would be scanned with fMRI while making them cognize particular words– proved exactly that: namely, the processing of many verbs and nouns recruits specific parts of the sensorimotor cortex that are semantically associated with those words (S. Hauk et a., 2008). For instance, when we read the word *leg*, we get activation in neurons of our motor cortex that are responsible for leg movement.

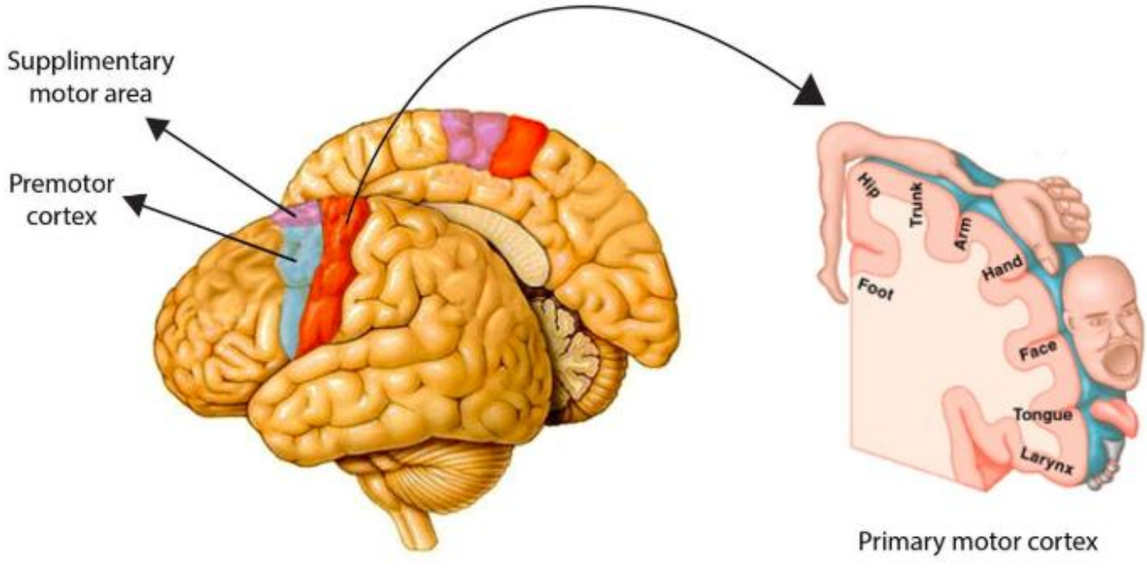


Figure 4. The motor cortex.

This provides a possible answer to the so-called *symbol grounding problem*. There seems to be a difference between merely manipulating symbols by following a set of rules (like a computer would) and feeling that these symbols have meaning (which is how we seem to experience language). But how is that experience of meaning grounded in our brains? After the aforementioned observations, the embodied-cognition view proposed a model for explaining the symbol grounding problem: its idea is that, when processing words, we create an embodied mental simulation, a recreation of what it would feel like to experience in our own bodies what is being told in language (S. Lakoff&Johnson, 1999). For instance, if you read "Maradona kicked a ball," you would create a mental simulation of that scene and of what it would feel like for you to experience (imagining yourself in presence of Maradona, observing him kicking a ball). And, so as to perform this embody simulation, your brain would recruit the pertinent sensorimotor areas (the neurons responsible for leg movement, visual perception, etc.).

Once this model for explaining meaning processing was formulated, its predictions had to be tested in further experiments, so as to verify if the provides an adequate description of how our brains process meaning and how they do it. In this process, many properties of the embodied-simulation program of our minds were revealed.

Our embodied simulations are not as detailed as the reality we perceive directly from our senses. This can be proven with a simple experiment: If you try to draw a 10 EUR banknote (or any other currency that is familiar to you) only on the basis of your memory, the result will be much less detailed that if you have an actual exemplar in front of you. Nevertheless, our embodied simulations are often more detailed than one might think. When we read a sentence that mentions an object, we do not imagine a completely abstract and features-less representation of the object. Instead, we imagine object with lots of particular properties, engaging all our senses.

An experiment proved an aspect of this by asking people about parts of objects. The assumption is that, if they are making a detailed visualization of the object in question in their minds, then the more visible a property is, the faster the test subject will identify it. A simple way to measure this is to consider parts with different size: the larger ones should be more easily identifiable than the smaller ones. And that is exactly what was shown by an experiment performed by Solomon and Barsalou (2004). The participants were asked to imagine, for instance, a bear, and then were asked about body parts of the bear. Systematically, the participants answered faster about the bear's *head* than about the bear's *nose* –and about every bigger part, in general, faster than about every smaller part.

Another experiment meant to analyze how detailed our embodied simulations are presented participants with pairs of objects and properties. Each of the properties corresponded to a different sensorial modality: sound, vision, taste, smell, touch, or motor control. What the participants had to do was to match the mentioned object with the mentioned property. That is, if they saw *blender-loud*, they had to state "the blender is loud." What was discovered is that the participants were faster at saying "the blender is loud" when the previous pair presented included a property of the same modality (auditory), such as "the bell rings," than from another modality (such as "lemons are sour"). The interpretation is that the embodied cognition of an object recruits also the activation of our senses in relation to the pertinent properties of the object.

This means that our embodied simulations are much deeper and extended than suggested merely by the aforementioned correlation between reading the word *leg* and the activation of our leg-neurons: Also, reading *ball* can activate our leg-neurons, and also our visual cortex, and perhaps even our auditory cortex. The general thesis derived after these experiments and many others is that our embodied simulations have the form of an *immersed experiencer view* (Bergen, 2012). That is, we imagine what it would feel like to be actually experiencing with our whole body what its being described in speech.

EMBODYING LITERATURE

We can find profuse literary cases that make productive uses of our embodied cognition. The properties of embodied cognition will be more developed in the following chapter (Rhetoric). But we can advance some observations.

Let us consider only the first strophe of Oscar Wilde's *The Ballad of Reading Gaol* (1897):

> He did not wear his scarlet coat,
> For blood and wine are red,
> And blood and wine were on his hands
> When they found him with the dead,
> The poor dead woman whom he loved,
> And murdered in her bed.

In this single strophe, Wilde employs several kinds of embodied-cognition strategies. By opening the poem with a pronoun (*He*), Wilde immediately sets a viewpoint, a perspective from which we can start to elaborate our embodied simulation: We know we will be the observers of a scene about a man. He adds visual references to characters, setting, and objects (the scarlet coat, the stained hands, the dead woman in the bed). He mentions bodily actions (to wear, to find, to murder). He builds up sensorimotor metaphors (comparing blood with wine). And he even introduces the bodily experience of rhythm through the meter of the verses, creating thereafter the effect of an auditory eco by using rhymes. As a result, our minds can generate after this single strophe a quite rich embodied simulation of what is being described through language.

What about Saroyan's poem (*lighght*)? First, we can naturally expect it to evoke the word *light* in the readers' minds, due to its graphic and phonetic similarity. Secondly, we might expect the word *light* to recruit their visual cortex –since the content of this word is associated with visual experiences. However, this word is isolated in the text, it has not particular linguistic context, it does not mention any concrete object nor action, and it is it does not even determine any particular perspective from which to set our embodied simulation –because it is not clear who is speaking to us in the text. In consequence, we can perhaps expect the embodied simulation resulting from reading this poem to be rather weak and vague instead of engaging and detailed. Indeed, the extended accusation of *meaninglessness* attributed to the poem by a significant sector of the public opinion seems to be consistent with this account: people did not accuse the poem of being offensive, repulsive, or immoral in its content; the accusation fundamentally argued, instead, that what was offensive was that a poem perceived as meaningless, as semantically empty, would receive be awarded as qualitative poetry ((s. Daly, 2018). Furthermore, this suggests that producing rich embodied experiences might be a property that people typically expect or even require of poetry. In fact –as we will see more in detail in the chapter on Emotions– this "embodiment requisite" is not arbitrary, but it is deeply related to the origin of the artistic instincts in our species (s. Dutton, 2010; Miller, 2001).

But what about the words that do *not* refer to physical perceptions or actions? How do our brains process abstract words such as *meaning*, *democracy*, or *poetry*?

A SECOND SYSTEM FOR PROCESSING MEANING

For a long time, neurolinguistic research was focused in what is called referential language. When a child is learning how to speak, many words are taught to him or her by pointing out to particular instances of the referent: “This is a hand,” “This is eating,” etc.; so that, in time, the individual generates an association between the term and the thing or action. And this process can be accounted by the embodied view. However, this explanation, by itself, is incomplete, because an even greater part of our vocabulary is not learnt like this, in presence of the referents of words, but in absence of them. Most of the vocabulary is learnt by hearing people talk about things and actions that are not available to our observation –like when hearing a fictional story. In consequence, neurolinguists proposed a model that accounts also for this second way of encoding word information.

Computational theory models proved that the semantic information of a word can be formally accounted by calculating the frequency with which the word appears related to other in language use (speech, texts, etc.). On this basis, neurolinguists have hypothesized that the brain would have a system for performing this kind of calculation: inferring the meaning of a word from its linguistic context (recent evidence of this can be found in s. Carota et al., 2017). If this was true, then patterns of similarity should be found between the co-occurrence of words in speech and the neural wirings between words. And this is what the experiments revealed.

A team of neurolinguists –integrated by F. Carota, N. Kriegeskorte, H. Nili, and F. Pulvermüller– performed recently a study using fMRI and representational similarity analysis –a technique aimed at the comparison across patterns– (Ibid.). They targeted words from two macro-categories: actions and objects. Actions were chosen in relation to body-parts: as arm-related, leg-related, and face-related. And objects were chosen of three kinds: animals, food, and tools. They averaged the searched words in relation to their average occurrence in language use. They made a group of people hear these words, and the patterns of neural activations were registered and compared with the patterns of co-occurrence of these words in linguistic context.

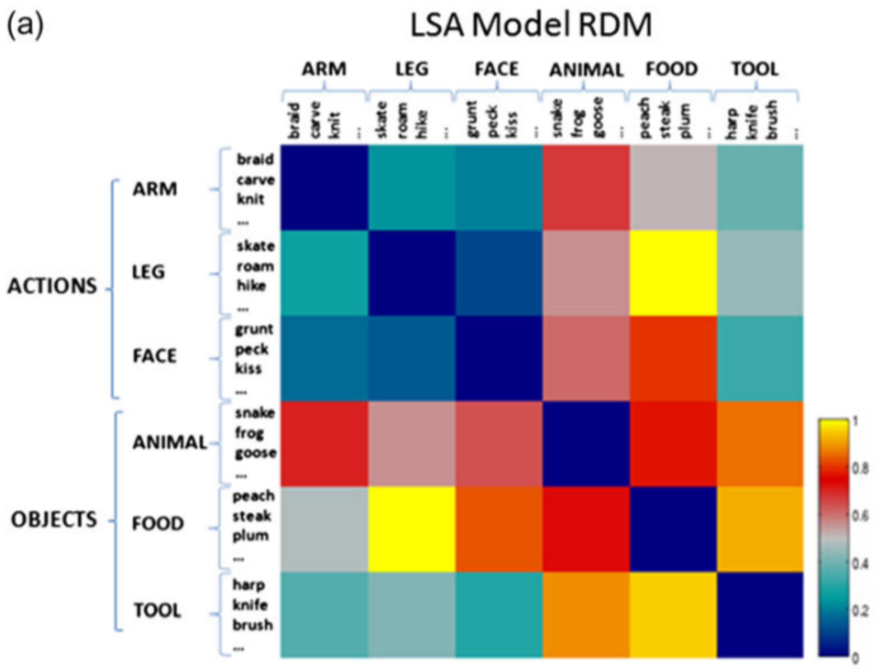


Figure 5. Latent-Similarity Analysis Model RDM (Carota et al., 2017).

As it can be observed in the results of this test, there was high similarity in the patterns of brain response of action-related verbs (as the embodied view predicts), but also across word classes, and with a patterned distribution: namely, more between action verbs and tool-related nouns and, less between action verbs and food-related nouns, and none between action verbs and animal nouns. That is, arm verbs like "braid" "carve" and "knit" were similar not only to other action related verbs (such as "grunt" "peck" and "kiss"), but also to words of a different semantic category, such as harp, knife, and brush (tools).

The model that has been proposed to account for these results is two-folded. On the one hand, they argue, we must have a system for registering referential meaning: that is our embodied-cognition system, which would create a *ground kernel* of words that are encoded on top of our sensorimotor area (Ibid.). But, on the other hand, we must also have a neural system that is capable of calculating the likeability of a word of appearing in a particular linguistic context, and thereafter infer elements of the meaning of the word and wire it to our ground kernel. The first system determines that certain areas of our brain are category-preferential (the motor cortex for processing verbs, for instance) and certain areas that are multi-categorial or amodal (like the circuits that allow us to link verbs with tools). Indeed, just as the embodied-view predicts the fact that action verbs are similarly patterned, this second system predicts that arm-related verbs will also have a positive correlation with tool nouns and negative correlation with animal nouns, in reason with their actual distribution in the context of linguistic use.

HOW CAN THIS MODEL BE USED TO ANALYZE LITERARY INTERPRETATIONS?

The integrative account presented in the last section suggests that, when processing a word like "light" (suggested in Saroyan's poem *lighght*), we would not only recruit the parts of the brain that process the visual perception of light, but we would also recruit terms that are frequently associated with "light" in speech. Where can we find these other terms? Many data basis of language-use exist currently. A particularly rich one for identifying trends in real time is GoogleTrends (GT).

GT calculates the frequency with which a keyword is searched in Google Search Engine and with what other queries it appears related –which gives a parameter of people's interest. If we search in GT the word "light," today (29 July 2018), worldwide, we find that it has the following curve of frequency:

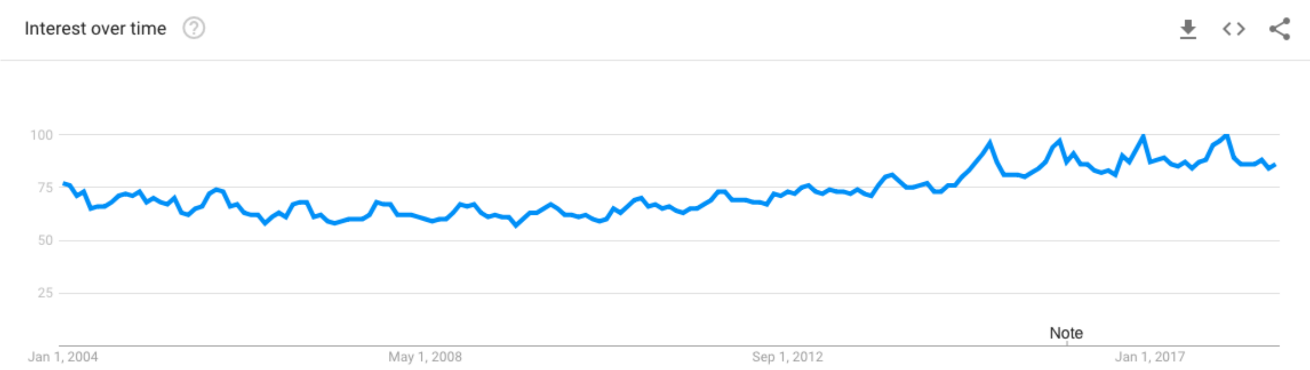


Figure 6. Worldwide searches of the keyword "light" in Google SE. (GoogleTrends, 2018).

It has the following geographic distribution:

Figure 7. Geographic distribution of the searches of the keyword "light" in Google SE. (GoogleTrends, 2018)

And it is most usually embedded in the following queries (these are only the first 20 results): *metro last light, pokemon light platinum, light between oceans, light yagami, dying light, via light, all the light we cannot see, led tube light, light scribe, hyper light drifter, light png, bud light lime, bud light platinum, lara croft and the master of light, segunda via light, Philips wake up light, infamous first light, zuzka light, star shower laser light, light novel* and *light in a box* (GoogleTreds, 2018).

If our brain has a system for encoding the meaning of a word in relation to its linguistic context, we could interpret that these statistics of linguistic use identified through GT would be showing us at least a part of the map of meanings that people are likely to relate with a word like *light*. In consequence, if we would want to enquire the meaning that Saroyan's poem might have for current audiences, this kind of quantitative inquiry would provide us with useful data for elaborating informed hypotheses. By observing the data results, we can observe immediately some expectable results, for instance, that the word light is more frequently searched in English speaking countries, which indicates that they would be more familiar with it. But it also gives us information of which non-English-speaking countries have also certain familiarity with the term (v.g. Western European countries, India and Pakistan, Brasil and Argentina, etc.). At the same time, by analyzing the related queries, we can discover particular aspects of the meaning people attribute to the term.

The related queries, in which "light" appears embedded might not necessarily give us direct information about the full meanings people attribute to the keyword in different situations. But it can show us some interesting aspects and tendencies. For instance, we know that the word *light* is ambiguous in that it can work both as a noun and as an adjective. And this is correlated with two different sets of meaning. As the Cambridge dictionary registers, when used as a noun, *light* refers to brightness –or to something that produces brightness, such as a device or a flame–, but when used as an adjective, *light* means not only bright, but also pale, not heavy, not much, not serious, and not severe.

The typical procedure of close-reading criticism (as performed mainly in the tradition that goes from the Russian Formalists and the New Criticism School to the traditions of post-structuralism and cultural studies) would consist in speculating on all potential meanings in which that word (in that poem, for instance) can in principle be interpreted. However, using the perspective that the cognitive model gives us, plus the aid of quantitative tools, we can address a further question: what meanings are the actual readers more likely to attribute to that poem? This question is particularly pertinent if we want to understand the effective role and impact that a piece of language can have in a particular cultural context.

Indeed, if we take the stated queries in which the term "light" appears embedded in GT, and count how often it is used as a noun or as an adjective, we discover a particular distribution: "light" appears in 65% of these 20 cases used as a noun, whereas only in 25% of the cases it is used as an adjective –and in 10% it refers to personal nouns: names of people and places, etc. On this basis, we can hypothesize that, if current readers would encounter Saroyan's poem, a majority would be likely to relate it with concepts like *brightness* and *flame* (noun meanings); but we should also expect to find a proportionally smaller group that might be more likely to relate it with concepts like *weak* and *unserious* (adjective meanings), instead.

As it can be appreciated, although this hypothesis refers to subjective responses, the hypothesis poses an objective question that, with adequate techniques, can be empirically addressed. This way, it's predictive power can be assessed, and that can guide us into formulating explanations about the phenomenon and also into formulating new hypotheses, such as: Could the fact that people read Saroyan's poem as a noun or as an adjective have affected how much they liked it in the first place? Is there a different proportion of noun- and adjective-interpretations in the favorable and unfavorable audiences?

This exploration of quantitative data guided by cognitive models that has been presented here was meant to offer a very small sample of the potential of combining cognitive science with data science for the study of literature –in favor of which I will argue along this book. As said, this will be done in two main parts: the first one focusing in cognitive science, the second one in data science. In particular, the relationship between meaning and linguistic context that has been considered in this last section will be more deeply developed in the case study of the third part of this book –where, instead of analyzing a single word, a whole text will be studied.

But, before going into that, we must still explore many more aspects about the reading mind. We have seen in this chapter an introduction to the problem of how we process words. In the next chapter, we will analyze how words affect each other when put together in phrases, sentences, and discourses, and what does tell us about the nature of language and literature.

CHAPTER 3

RHETORIC

HOW PHRASING AFFECTS OUR INTERPRETATIONS AND BEHAVIOR

Why did Dylan Thomas write *Do not go gently into that good night* (1947) instead of writing – more straightforwardly– something like: *Try not to die*? Why do people repeat the proverb *The pen is mightier than the sword* instead of directly saying that ideas are more effective than enforcement? Why did Miles Raymond entitled his novel *The Day After Yesterday*, instead of calling it directly *Today*?[13] In short, why is it that we care so much about the ways in which we phrase language, making so specific word choices, so elaborate rhetoric turns? What criteria do we use for deciding these things? And how do these rhetoric decisions really influence our responses to texts?

THE FRAMING EFFECT

In 2003, cognitive psychologists Eric Johnson and Daniel Goldstein (J&G) performed a study in which they analyzed the distribution of organ donors across European countries. They obtained the following results:

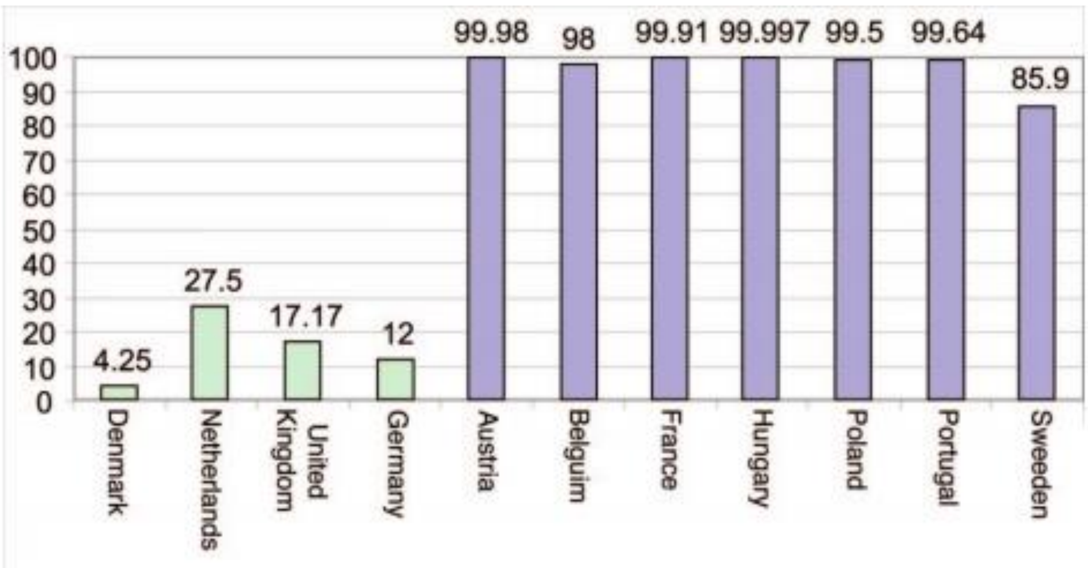


Figure 1. Effect consent rates by country. The four leftmost bars (green) are explicit consent (opt-in). The seven rightmost bars (blue) are presumed consent (opt-out). (Johnson&Goldstein, 2003).

[13] Miles Raymond is the protagonist of Rex Pickett's novel *Sideways* (2004).

As it can be seen in J&G's graph, in some countries organ donation seems to be consistently frequent (right, blue), whereas in others it seems to be consistently rare (left, green). But what can explain this patterned difference?

When asked this question in surveys, people usually answer that the differences in organ donation must be produced by cultural differences: they might reflect the fact that these are societies with different values, whose people care differently about others (s. Ariely, 2008). However, J&G's graph shows that culturally similar countries exhibit very different behavior regarding organ donation: Sweden is on the right (many donors) while Denmark is on the left (scarce donors); Austria is on the right while Germany is on the left; French is on the right while UK is on the left.

If cultural differences do not seem to have much correlation with organ donation, what might produce the differences? The case of Netherlands clearly illustrates the magnitude of the effect discovered by J&G. The Netherlands arrived at the 28% shown in the graph only after the Government mailed every household in the country begging people to join their organ-donation program. They were not indifferent to the subject, but actively trying to make the rate grow. Nevertheless, their strategic actions –massive as they were– did not bring them very far. Could it not be, then, that the countries in the right part of the graph did not have very diverse people but were instead employing more efficient strategies for orienting people's decisions? Indeed, that seemed to be the case, and the different strategy identified by J&G was in the text of the forms used for deciding the organ-donor status.

The countries in the left (scarce donors) were using *opt-in* forms, such as the following one:

☐ Check the box if you *want* to participate in the organ donor program.

People tended not to check the box and, in consequence, they would not donate their organs.

The countries in the right (many donors) were using *opt-out* forms, instead:

☐ Check the box if you *do not want* to participate in the organ donor program.

People also tended not to check the box but, in these cases, they would become donors.

Contrary to what common sense might lead us to think, people did not behave differently in countries with scarce or with many organ donors. In both cases people did the exact same thing: they tended not to check the box. But, since the forms were differently designed, the same behavior had different outcomes in each case (becoming an organ donor or not).

J&G explained this phenomenon as a cognitive bias: When faced with decisions that people perceive as serious and difficult (e.g. decisions that involve life and death, such as organ donation), they tend to chose the default option (in this case, leaving the box blank). J&G interpreted this bias as a coping strategy: avoiding to decide by accepting whatever was chosen for us (Johnson&Goldstein, 2003).

The bottom-line of this discovery is an alarming realization: Many of the judgments and decisions that we feel we take in our lives are actually decided, more than by ourselves, by the people that write forms –as well as contracts, signs, labels, newspapers, speeches, tweets, articles, warning sings, idioms, and also literature. This case shows a particular instance of the many ways in which phrasing choices can affect people's decisions. These phenomena were first empirically tested and systematically theorized in the 1980s by Atmos Tversky and Daniel Kahneman under the name of *the framing effect.*

Classical economy considered humans as agents that act by assessing gains and costs in a purely rational manner. But psychological research has been progressively showing that our decisions are largely irrational –although, not random: we are, in some ways, *predictively* irrational (in words of Dan Ariely, 2008). By integrating the psychological findings with economical inquiry, the field of behavioral economics emerged. Atmos Tversky and Daniel Kahneman are two of the founders of this discipline. One of their most groundbreaking experiments was precisely aimed at –as said– measuring how different phrasing alternatives affect people's interpretations and decisions, and according to what principles this occurs. This experiment is today known as The Asian Disease Problem (Tversky&Kahneman, 1981).

In this experiment, test subjects were presented with a hypothetical scenario: they were asked to chose between two kinds of treatments for 600 people that were described as suffering from a deadly disease. The first kind of treatment (A) offered a certainty (*X number of people will live/die*), whereas the second (B) offered a gamble (*X% chances of X number of people living/dying*). But, at the same time, each of these treatments was described in two different ways, either with a positive or with a negative frame, as follows:

Treatment A (Certainty)

(Positive frame): *If this program is adopted, 200 people will be saved.*

(Negative frame): *If this program is adopted, 400 people will die.*

Treatment B (Gamble)

(Positive frame): *If this program is adopted, there is one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.*

(Negative frame): *If this program is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.*

Each participant was presented with one version of A and one version of B. If people were purely rational agents, their preferences of either A or B should be based in the actual differences between A or B (certainty or gamble), not in the differences of the framing –which did not introduce any logical change in the content. However, the difference in the word choices did matter and showed a huge effect in the participants' responses. Namely, people tended to chose certainty (A) when this option was the one presented with a positive frame, but they preferred to gamble (B), instead, when this option was the one presented with a positive frame. The remarkable discovery was that people's decisions were more systematically guided by the framing –even in a life-and-death scenario like this one– than by the actual content of the sentences (i.e. the nature of the treatments described).

The so-called framing effect is a cognitive bias that describes the fact that people's choices depend largely on how the options are presented to them. In this particular case –which was built in relationship to a decision about risk: certainty (*risk-aversion*) vs. gamble (*risk-seeking*)–, the framing effect showed how focusing on gains (positive framing) or losses (negative framing) affects our behavior in a predictable direction: namely, people are risk-averse for gains and risk-seeking for losses –which is today a central concept of behavioral economics (Ibid.).

The framing effect has consistently proven to be one of the strongest biases in decision making (s. Thomas&Millar, 2011). It is, therefore, extensively used currently in every field where people behave in response to communicational inputs, to the alternatives they are offered, be these in the form of language or any other. By designing models on the basis of the framing effect, nowadays student cafeterias carefully set the order in which the food is sorted and displayed so as to guide the students into making healthier dietary choices (Ensaff et al., 2015); supermarkets vary the location of their products in accordance to the demands of targeted classes of customers (Sulakatko, S., 2014); and governments use framing techniques as social-engineering tools, so as to orientate people's decisions, from garbage recycling and energy consumption to voting intentions and organ donations (as we have seen in J&G's study).

The framing effect was further developed in cognitive linguistics –especially in works like George Lakoff's (1990). And it soon became clear that an exemplary field where framing is –and has always been– highly productive and intensely used is literature. We could say that literature is a field naturally designed for elaborating different forms of framing, since literature is a practice that is not only concerned with conveying information but with conveying it in appealing ways, producing particular effects in the readers. This can be found in all the elaborate rhetorical strategies employed by literary authors –as mentioned in the beginning, with the examples of Dylan Thomas' verses, popular proverbs, and novel titles. In the following sections, we will explore some of these linguistic framing techniques that are frequently leveraged in literary writing for modeling our reader-responses.

EMBODYING LINGUISTIC CONTEXT AND GRAMMAR

> One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed into a horrible vermin. He lay on his armor-like back, and if he lifted his head a little he could see his brown belly, slightly domed and divided by arches into stiff sections. The bedding was hardly able to cover it and seemed ready to slide off any moment. His many legs, pitifully thin compared with the size of the rest of him, waved about helplessly as he looked. (Kafka, 1915).

This is the opening of Franz Kafka's *The Metamorphosis*, which is an exemplary literary case of embodied cognition. We have said in the previous chapter that, when processing language, our minds create mental representations –called *embodied simulations*– of what is being told to us. Thus, when we read *One morning, Gregor Samsa (…) found himself transformed in his bed into a horrible vermin*, our minds start immediately to imagine that morning, that bed, and that individual. And, when we do so, our brains recruit neurons responsible for our senses and bodily movement (sensorimotor cortex), so that we get to see those things in our minds' eye, imagining what it would be like for ourselves to experience in our flesh what the text describes. These embodied simulations are a crucial part of how we process and experience meaning.

Many experiments of embodied cognition have been performed in the last decades so as to discover in which particular ways these embodied simulations are affected by our rhetoric decisions, by the ways in which we frame the content of a text, by our word choices, our phrasing strategies and our style. And what was discovered is that these rhetoric decisions are deeply influential, namely in indicating us not only what to simulate in our minds, but also how to simulate it, and from which perspective. Exploring these phenomena can help us better understand the nature of our embodied simulations (of our experiences of meaning) and also many of the rhetoric and stylistic decisions of literary writers.

To begin with, our embodied simulations are measurably sensible to linguistic context. A series of experiments run in the University of California obtained evidence of this (Sato et al., 2012). They exposed participants (English and Japanese speakers) to sentences that mentioned objects and implied that they had a certain shape. For instance, a sentence would say *The egg was put in the fridge* (suggesting a whole egg), whereas another one would say *The egg was put in the pan* (suggesting an open egg, instead). Afterwards, participants were asked to identify images of eggs (whole or open). And people were effectively faster at recognizing the eggs whose shaped matched the shape implied in the sentence they had read (fridge-whole, pan-open). The interpretation of these results, in light of the embodied-cognition theory, would be that using the word *fridge* or *pan*, in the context of *egg*, leads people to activate different mental representations of the eggs –the adequate for each case–, which consequently primes the recognition of that actual shape when found afterwards in a picture. The same effect occurred in another version of this experiment where the cue that indicated the shape of the object preceded the mention of the object (e.g.: *In the fridge I put the egg*), which means that both the preceding and the subsequent linguistic context can influence our embodied simulation of a given concept.

These results suggest that our embodied simulations have a particular structure: when confronted with a sentence, we do not simulate each word of it individually, in a linear succession (like in a comic-strip), but, instead, we seem to aggregate the meaning of many words and merge it all together, forming thereafter a dynamic immersed-view experience (like in a virtual reality movie). In consequence, when we read the opening of *The Metamorphosis*, we do not simulate in our minds successively the concepts of *morning*, *bed*, and *vermin* (*Ungeziefer*, in the original German), as individual ideas, but we progressively integrate the concepts we encounter as we read into the imagination of a whole unified mental scene: including a particular vermin, in a particular position, on a particular bed, on a particular morning.

Another interesting example of the extent to which rhetoric alternatives model not only what we simulate but also *how* we simulate it, is shown by an experiment performed in 2007 by Richardson and Mattock. They exposed participants to two different sentences. The first one said “Through the *clean* goggles, the skier could *easily* see the moose” (italics, mine). The second one: “Through the *fogged* goggles, the skier could *hardly* identify the moose.” After reading one of these two sentences, the participants had to identify the picture of a moose. One would think that the fact that the skier’s goggles had been described differently in each sentence (clean/fogged) should not have any effect in people’s speed at spotting moose pictures. The incredible result is that it did: the participants that read about the moose as being *hardly* seen by the skier’s through *fogged* goggles was consistently slower in spotting moose pictures afterwards than the participants that had read the *clean*-goggles and *easily*-seen sentence. These results suggest that our linguistic choices can even model the granularity (level of detail) of our embodied simulations: that describing the *blurry* vision of an object really makes the listeners have a blurry mental visualization of that object.

This last experiment suggests also something more: that language determines our *view point.* In this case, the participants were allegedly taking the perspective of the skier, since the difference in his or her goggles affected the detail with which the participants visualized themselves the moose.

Building up the viewpoint of the reader is a crucial rhetorical procedure that narratologists call *focalization*. Stories have not only a narrator (*a voice that tells*) but also a focalizer (*an eye that sees*) (s. Martinez&Scheffel, 2009). Kafka tells *The Metamorphosis* in third person (*When Gregor Samsa woke up…*). That is the narrator. But, in terms of focalization, it can be noticed that –at least in this first paragraph– we are only shown what the character Gregor Samsa can see (*if he lifted his head a little he could see his brown belly*), so the text is displaying a so-called *internal* or *participant* focalizer. This manipulation of the viewpoint is not trivial but it affects reader-responses in significant ways: Imagining the Titanic sinking as seen from the shore is a very different experience from imagining it sinking with us inside.

Now, even though the narrator and the focalizer are conceptually different phenomena (as narratologists indicate), they seem not to be completely independent at a psychological level: Cognitive experiments have shown that the grammatical markers of person (which indicate who is talking to whom) have a measurable effect in the viewpoint that we take in our mental embodied simulations. That is, the kind of voice we imagine speaking seems to bias the kind of eyes we imagine seeing. Let us explore how this phenomenon works by considering the following novels' openings:

1. Vaughan died yesterday in his last car-crash. (Ballard, *Crash*, 1973)

2. You have put your left foot on the grooved brass sill, and you try in vain with your right shoulder to push the sliding door a little wider open. (Butor, *Second Thoughts*, 1957).

3. Through the fence, between the curling flower spaces, I could see them hitting. (Faulkner, *The Sound and the Fury*, 1929)

4. I am an invisible man. (Ellison, *Invisible Man*, 1952)

Each of these stories opens with a different rhetoric strategy, each using a particular grammatical person: He, You, I. And, as said, these markers affect reader-responses in the predictable ways.

What studies have shown is that, when a story is told in the third person (*He*, *She*), readers tend to assume an observer perspective (external focalization) (s. Nigro&Neisser, 1983). This means that, in the example 1, readers would imagine what it would be like to *see* (from "outside") Vaughan dying in a car crash. On the contrary, when the story is told in the second person (*You*), readers tend to take a participant perspective (internal focalization) (s. Zwaan et al., 2004): in the example 2, readers will imagine what it would be like to *be* (from the "inside") the character that puts the left foot on the grooved brass sill, etc.

In one of the experiments in which these effects were tested, participants were exposed to sentences that used different grammatical persons for describing objects or actions (v.g. *You/He are/is throwing a ball*), and afterwards they had to respond to images of balls in motion –following a similar logic to that of the aforementioned experiments (s. Bergen, 2012). By measuring which images the participants recognized more easily, it could be inferred which viewpoint they were assuming in their embodied simulations. And, as described, when people heard a sentence like "*You* are throwing a ball," they effectively became faster at recognizing images that showed balls moving away, distancing towards the background –which is consistent with a participant viewpoint. Whereas, when the participants read sentences like

"*He* is throwing the ball," they became faster at recognizing the images that showed balls moving from left to right, instead –which is consistent with an observer viewpoint.

Interestingly, there was also a pattern, in these latter results, in the kind of direction in which the participants imagined the balls moving (from right to left or from left to right). Nevertheless, this was not predicted by grammatical markers but by a different kind of variable: the participants linguistic background. People that speak English, Spanish, or Italian tend to imagine the ball moving from left to right, whereas Arab and Chinese speakers, tend to picture it from right to left. An experiment was done about this phenomenon (describing *people passing by*), which showed that the conventional directionality of the writing system of our language seems to influence considerably the default directionality of our embodied simulations (Maas&Russo, 2003; s. also Chan&Bergen, 2005).

Coming back to the effects of the grammatical person, in conclusion, the general rule would be that the second person (You) primes a participant perspective (it makes readers prone to put themselves in the shoes of the character, and imagine what it would be like to *be* the character), whereas the third person (He/She) primes an observer perspective (it predisposes people to imagine the scene as seen from outside, as if they were external observers watching the characters).

But what happens with the first person (*I*)? What perspective do readers assume when reading stories that begin like the ones of our 3rd and 4th examples? A variation of the aforementioned experiment discovered that the use of the first person can lead either to an observer or a participant view point, depending on a further variable: how much information is given about the character in question (Brunye et al., 2009). Two texts were set up to test this. The first one said simply *I am slicing a tomato*. People responses after reading this sentence were consistent with a participant-perspective. But, in a second version of the text, some information about the character preceded the sentence:

> I am 30-year-old deli employee.
> I am making a vegetable wrap.
> I am slicing a tomato.

After reading this version, the participants responded better to the observer-perspective pictures, instead (Ibid.). What this suggests is that the information that is displayed about the character (a richer or a poorer description) also biases people's viewpoint in predictable directions: a richly-described character primes an observer viewpoint, and a poorly-described character primes a participant viewpoint. We can interpret these results as indicating that, what makes people assume an observer perspective, is knowing with whom they are talking –or about whom they are reading. The more the *I* of the story is personalized as an individual, with concrete recognizable traits, the more we will see this character as an *other*. But, when the information about the character is scarce, people seems to take the content of the story in a more generic sense, as describing something that could happen to anybody, and in consequence they become more prone to imagine what it would be like to experience the referred content by themselves.

This tells us something about our examples 3 and 4. Both are told in the first person (*I*), but they vary in the kind of information they give us. In Faulkner's opening (4), we are given information about the things the character is seeing: *Through the fence, between the curling flower spaces, I could see them hitting*. But we are not given much information about the character itself. In consequence, as readers, we can easily place ourselves in the position of the character (participant viewpoint), and imagine what it would be like to experience what he experiences and how he experiences it: the referred flower spaces, and the people hitting, as seen through the fence. Ellison's opening works in the opposite way: *I am an invisible man*. The very first information we receive is about a feature of the character. The novel continues extensively in the same mode: "…I am a man of flesh and bone … I am invisible, understand, simply because people refuse to see me…" (Ellison, 2952), depicting with all this information a particular character that we are brought to imagine seeing from the outside as a particular individual (observer perspective), different from ourselves, to whom all these things happen.

Skilled writers use these techniques not randomly, but as strategies for building up more complex and precise effects. Indeed, the pronominal decisions taken by each of the referred novel openings, and the kind of viewpoint they elicit in the reader, are directly related with the kind of story each of these authors tell. (1) Ballard elicits an observer viewpoint to make us secretly gaze into the underworld of symphorophilics –people that get aroused when observing car-crashes. (2) Michel Butor elicits a participant viewpoint to makes us embody the experience of taking the train from Paris to Rome –an effect that is supported by many other features of the novel, such as a timeline in the margins of the page measuring the ideal reading duration of the novel in parallel with the duration of the train trip. (3) Faulkner (participant perspective) uses a *stream of consciousness* (narrative technique invented by James Joyce) so as to make us imagine what it would be like to possess the mind of a mentally-challenged man and to think what he himself thinks –emulating a real-time thought process as a flux of juxtaposed sentences. And, finally, (3) Ellison (observer perspective) draws our attention to the living conditions of a socially marginalized man, in an attempt to make us see what is typically unseen, to make the invisible visible. In this sense, each of the rhetoric decisions taken by authors produces particular effects that are coherent with the subjects of their works and, this way, contribute to the general literary effect.

Another productive way in which different rhetorical alternatives affect our embodied simulations is by the use of metaphors.

METAPHORS IN THE FLESH

When we read *Do not go gently into that good night*, we might interpret that Dylan Thomas is suggesting something like *resist death*, which can in turn be also interpreted as a way of saying *do not give up*. If we are capable of decoding meaning in these ways, it is because we understand that ideas can be phrased *metaphorically* –i.e. talking about X in terms of Y. Following our aforementioned interpretation, in this verse, we are interpreting that Dylan Thomas talks about failure in terms of a biological state (death), about biological states in terms of times of the day (day: life, night: death) (*that good night*), and about the process of changing biological states in terms of movement across space (*do not go gently into…*). Literature, indeed, is abundant in metaphors. Empirical literary studies have actually identified metaphors as a true literary universal: each society might conventionalize different metaphors, but no human society has been found that does without metaphors (s. Lakoff&Johnson, 1999).

The phenomenon is even deeper, because metaphors are not an exclusively literary procedure. They are also used in every other domain of language. Nutritionists talk about the food pyramid (which implies a version of the metaphor *hierarchy is a vertical structure*), lawyers talk about public defense (implying the metaphor *trials are battles*), linguists talk about branching (*sentences are trees*), mathematicians talk about real and imaginary numbers (*numbers are objects that exist along a line*), and neurocognitive scientists talk about neural circuitry, networks, activations, and modules (*brains are computers*).

As George Lakoff and Mark Johnson (L&J) observed in their book *Metaphors We Live By* (1980), even our most colloquial language is deeply metaphorical. In fact, many metaphors are so frequent that we can hardly notice when we use them that we are talking metaphorically (Ibid.):

Ideas are objects: He has an idea, He took his idea from her, She found/lost the idea in a book.

Understanding is seeing: Now I clearly see what you mean, I cannot discern the meaning.

Morality is cleanness: He has a spotless record, He didn't want to get his hands dirty

Society is a body: The health of society, Social recovery, The backbone of society, Social paralysis.

Desire is hunger: He has sexual appetite, She is starving for success.

Affection is warmth: He is a warm-hearted man, She is a cold woman, He is cold-blooded.

These are only a few of the more than 200 cases analyzed by L&J in English language. These kinds of everyday metaphors are so common and deeply rooted in our language that it is certainly difficult to find a completely metaphor-free piece of speech.

After these observations, L&J formulated a provocative hypothesis: Perhaps –they deemed– metaphors are not simply superficial linguistic phenomena, but deep cognitive operations. That is, maybe we do not only talk metaphorically, but our mind really thinks metaphorically; maybe metaphors are *figures of thought* more than *figures of speech*. L&J called this phenomena *conceptual metaphors.* What this would mean, in concrete cognitive terms, is that our minds possess a system that enables them to translate information across different cognitive domains, which would underlie our linguistic expressions.

L&J conceptualized the notion of metaphor as a mechanism by means of which we process information about one domain (called *target*) in terms of another domain (*source*). For example, a metaphor like "ideas are objects" permits us to talk about ideas (target-domain) as if they would have the properties and affordances that objects have (source-domain): i.e. as it they were material things that can be seen, measured, possessed, traded, fund, lost, sold, destroyed, upgraded, etc. When analyzing metaphors in these terms, L&J noted a singular pattern: the target-domain is typically more abstract than the source-domain. This is clear in the aforementioned examples: the target domains are clearly abstract notions (ideas, understanding, morality, society, desire, affection) that are accounted for in more concrete bodily terms (objects, seeing, cleanness, body, hanger, warmth). This suggested a point of departure for inquiring L&J's hypothesis. If their theory was correct, if more than *talking* metaphorically we *think* metaphorically, then a particular prediction could be made: It would be expectable that, when people read "She broke my heart," they really picture in their mind the idea of an object getting broken, and when people read "Do not go gently into that good night," their minds really picture the idea of physically walking into a dark space –this would occur even while they understand perfectly that no heart was really broken and that Dylan Thomas is not talking about a nocturne promenade. On this basis, many experiments started to be designed in order to test this kind of phenomena.

One of the most direct experiments in relation to this was done by Wilson and Gibbs at U.C. Santa Cruz in 2007. They had a group of people perform different tasks that involved physical actions, such as grasping, swallowing, etc. But none of these words was mentioned, the participants were only induced to physically execute the actions. After each of these tasks, the participants would be shown sentences in a screen that they had to recognize as proper English expressions. Among the sentences shown, there were metaphorical idioms, such as "grasp an idea" or "swallow your pride." If metaphors were simply a superficial linguistic procedure, the cognitive processes related to the physical action of grasping should not have any relationship with understanding the expression *grasp an idea*. But they did: people were consistently around half a second faster in recognizing the metaphors whose pertinent action they had executed earlier. Performing the action of grasping or swallowing made them recognize faster the expressions that used those terms metaphorically. What this suggests is that some of the same mental and neuronal processes must get activated when think of *grasping ideas* and *swallowing pride* than when we actually perform these actions. That is: that when we use bodily notions (e.g. break a heart, starve for success, walk into the night) as a source for talking about more abstract target notions (grieving, ambition, death), our minds do not simply access the implied content but they really do it by recruiting the bodily notions that were mentioned. A second version of the experiment showed that the same phenomenon occurred even when the participants only *imagined* the actions of grasping or swallowing, even in absence of the actions themselves (Ibid.), which reveals to what extent these bodily actions seem to be encoded in our minds in relation to the abstract notions which they metaphorize.

This phenomenon suggested something even deeper. If –as seen in the previous chapter– our language is grounded in our bodily experiences, how is it that we understand abstract, disembodied concepts (words such as *idea*, *society*, *morals*, etc.)? Maybe the answer to this fundamental question was, precisely, in our mental system of conceptual metaphors. Integrating the theory of conceptual metaphors with the theory of embodied cognition, the theory of metaphorical simulation then emerged (S. Lakoff&Johnson, 1999). This theory explains that, once our minds ground some fundamental embodied concepts, they use these structures for processing other kinds of information, by means of a metaphorical faculty that allows them to translate information across domains (affection in terms of temperature-perception, time in terms of special-localization, etc.). This faculty would be what makes us capable of processing abstract concepts to begin with. And this theory entailed concrete predictions: If this was the case, then neural activity would should occur in the pertinent areas (sensorimotor cortex) when people talk metaphorically about abstract things. And this is what was tested next.

Many experiments were done using fMRI scanners so as to measure people's neural activity when reading metaphorical expressions. These experiments would typically expose patients to sentences that included action-related verbs. Some of these sentences would use these verbs in a literal way (e.g. *bite the apple*, *kick a ball*, etc.) and some in a metaphorical way (e.g. *bite off more than you can chew*, *kick the bucket*). While the patients would read the sentences, the researchers would measure the neuronal activation in the patients' brains. In the first experiments where this was tested, literal sentences activated the pertinent areas of the motor cortex predicted by the embodied-cognition theory. But the metaphorical ones did not (Aziz-Zadeh & Damasio, 2008; Raposo et al., 2009). However, the activations appeared in a third experiment, which enlarged considerably the sample of sentences, and also showed the sentences to the patients progressively (word by word) (Boulenger, 2009). And the activations were certified when they appeared in a fourth experiment that considered a further variable: the popularity of the metaphors (Desai et al., 2012).

This fourth experiment used not only metaphorical idioms (usual metaphors, such as *bite the bullet* or *kick the bucket*) but also less familiar metaphors (such as *bite into this idea* or *kick this meeting for tomorrow*). And, what they discovered was that, even though both cases activate our motor neurons, the less familiar ones do it much more. This means that, for instance, both *swallow your pride* and *swallow your love* will expectedly activate swallowing-related neurons, but the second one will do it much more –for it is less frequent. This could perhaps explain some of the difficulties of previous experiments in finding the pertinent neural activations when the patients read metaphors: they were using mostly very conventionalized idioms.

The hypothesis that when a metaphor becomes very familiar its embodied effects decrease had already been advanced by Bowdle and Gentner in 2005. And it constitutes an aspect of the reading mind that is particularly insightful for understanding the use of metaphors in literature.

We have presented the theory that embodied effects of metaphors seem to diminish as the metaphor becomes familiar. This phenomenon can provide one of the explanations for the constant search, by literary writers, of novel metaphors or novel ways of combining and applying them.

A common metaphor in English (and in many other languages) is *dying is sleeping*. Many idioms are evidence of this metaphor: e.g. *put to sleep*, *born asleep*, *sleeping with the fishes*, etc. Writing schools often advise writers to avoid idioms like the plague. However, more than avoiding them and creating radically new metaphors, what literary authors most often seem to do is to re-elaborate them in innovative ways. The following strophe by Robert Frost (from "Stopping by Woods on a Snowy Evening," 1923) provides a minimal example of this:

> The woods are lovely, dark, and deep,
> But I have promises to keep,
> And miles to go before I sleep,
> And miles to go before I sleep.

In this case, the first mention of *And miles to go before I sleep* can be read literally: as referring to a real distance, a real walk, and a real sleep. However, the repetition of this verse draws us to reconsider the meaning (it breaks Grice's maxim of quantity: Why is he repeating the information? What else is he trying to say?), which brings up the metaphorical content: the realization that *to sleep* might also mean *to die*. This way, by means of an unfamiliar presentation, the common metaphor becomes visible again. Following the aforementioned experiment, this literary effect can be explained as follows: The familiarization with a metaphor (through idioms) reduces its cognitive effect (the embodied experience produced by the sensorimotor activations), but that effect can be re-created by using the metaphor in an unfamiliar way.

The pleasure of encountering unfamiliar presentations of familiar concepts had already been identified as a characteristically literary effect by the Russian Formalists. The famous example given by Viktor Shklovski to explain this notion was Tolstoy's story "Kholstomer" (1888) which describes the world of men (something very familiar to us) through the eyes of a horse (an unfamiliar perspective) (Shklovski, 1917). We can find variations of this procedure in virtually every literary work: Shakespeare tells us the familiar story of two people falling in love in the unfamiliar situation of having enemy families; Conan Doyle tells us the familiar story of a criminal inquiry with the unfamiliar intervention of a super intelligent detective; and the movie *The Martian* tells us the familiar story of a man surviving a wreck but in an unfamiliar place (Mars). Viktor Shklovski called this procedure precisely *defamiliarization* (*ostranenie*). And one of the most elementary examples of defamiliarization can be found precisely in the creative usage of metaphors that abounds in literature: When Chesterton describes the night as “A monster made of eyes” (“Second Childhood,” 1938) he is taking familiar metaphors (*the night is a living being, the stars are eyes*) and combining them into a novel original arrangement. If using unfamiliar metaphorical expressions effectively increases our sensorimotor activations, this would constitute a physical evidence of *ostranenie*, not only as a theoretical description, but as an actually traceable psychological and physiological phenomenon.

By systematically studying the functioning of the metaphorical system of our brain, more specific principles that govern it were discovered. A very interesting one revealed the possibility of manipulating people’s perceptions. The idea would be that, if metaphors create actual patterns across mental domains, then we should be able to bias how people process information of one domain by means of producing alterations in the other one. For instance, in many languages, people talk about time using diverse spatial metaphors (e.g.: The Winter is coming, back in the past, in the end of my life, etc.). if people truly *think* of time in terms of space, then altering spatial cues should have some influence in people’s perceptions of time, even independently of language. And this is exactly what the pertinent experiments found. One of them was performed by Casanato and Boroditsky in 2008. In it, the participants had to observe a line growing in a screen for a period of time. After watching the video, the participants had to make an estimation of how much time had it took for the line to grow to its final size. In different scenarios, the speed of growth and the final size varied. And, what was found, was that the final size of the line systematically biased people’s judgment about the time passed, and it did it precisely in the direction that the metaphor *time is space* would predict: the longer the final size of the line, the more time people deemed it had taken, and vice versa. The groundbreaking discovery verified by numerous experiments of this kind is that many metaphors are not simply ways of talking, but they can affect our cognition even to the level of our sensorial perceptions.

This suggested a further hypothesis that is highly pertinent for understanding reader-responses to literature: If our metaphors are so deeply embodied that they can bias our perceptions, maybe they can also bias other responses, such as our interpretations and judgments. A series of experiments on this topic were done in relation to the very familiar metaphor *affection is warmth* (e.g. “He is a warm-hearted guy”, “You are so cold-blooded”, etc.). In one of them, performed by Williams and Bargh in 2008, participants were asked to describe an imaginary character. But, before doing this, each participant received a cup of coffee –unbeknown by the participants, this was part of the experiment. Some of the coffees were hot and some others were cold. When the coffee was hot, the participants were more prone to describe the imaginary character in positive and personal terms (as loving, happy, generous, sociable, etc.). And the opposite happened when the coffee was cold. (Williams & Bargh, 2008). This means that, just as cognizing affection recruits our temperature sensors, altering perceived temperature also predisposes people for a certain kind of affection judgment. And even a similar experiment was done in the opposite direction: people were asked to describe a moment in their past, and they were asked afterwards about the temperature of the room. When they had evoked a memory of inclusion, they felt the room warmer; when they had evoked a memory of exclusion, they felt the room colder. (Zhong & Leonardelli, 2008).

This bidirectional influence (from temperature perception to affection judgment and vice versa) shows the extent to which metaphors are embodied, inscribed in our flesh. Understanding these connections between abstract ideas and our sensorimotor perceptions gives us, in consequence, an empirical basis for exploring systematically the form of literary metaphors and their effects in our minds and bodies.

THE RHETORICAL INSTINCT AND THE ORIGIN OF BEAUTY

The rhetorical skills we have analyzed so far have a clear functional rationale. Being able to recognize a positive or a negative frame is useful for anticipating potential gains and loses –even if, for this very reason, it can be manipulated for priming different kinds of interpretations and behavior. Our sensibility to grammatical structures and markers is useful for guiding us into what kind of embodied simulation to perform. And our system of embodied metaphors facilitates our grasp of abstract ideas.

However, the rhetoric procedures that we can find in literary language go way beyond these informative functions. [14] Literature also exploits language in many ways that have no obvious direct utility and can, in consequence, be described in principle as *ornamental*. Poetry across the world, for example, recurrently uses procedures such as rhymes ("Once upon a time you dressed so fine /
You threw the bums a dime in your prime." –Bob Dylan, "Like a Rolling Stone," 1955), alliterations ("For the sky and the sea, and the sea and the sky." –Tennyson, "The Rime of the Ancient Mariner," 1798), complex rhythmical patterns (as illustrated by these famous iambic pentameters: *To be, or not to be–that is the question /, Whether 'tis nobler in the mind to suffer*, etc. –Shakespeare, c. *Hamlet*, 1599-1602), and even visual patterns –as shown not only by modern so-called *concrete poetry*, but also across history, for example in the case of this hatchet-shaped poem written in the Ancient Greece:

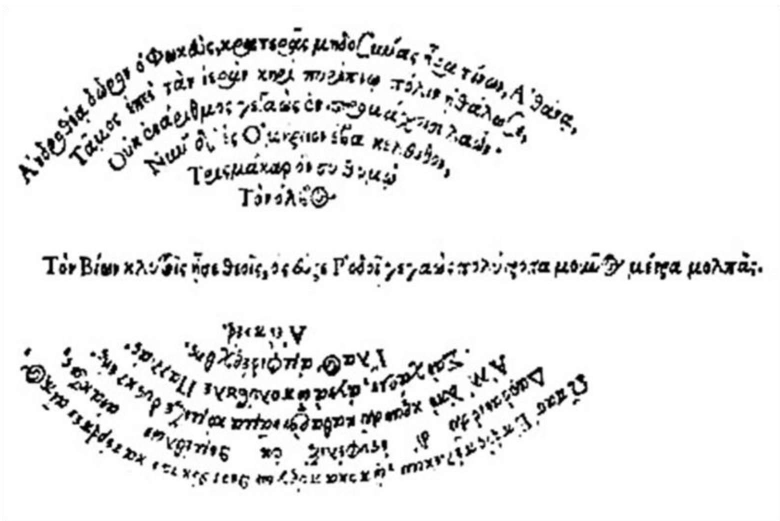


Figure 3. Simmias de Rhodes, "Pelekys," c. 300 BC.

[14] The term informative is used here in a technical way: in reference to the theory of philosopher of language H.P. Grice (s. Grice, 1975). He observed that human conversation follows a series of rules (*maxims*): informativity, truthfulness, relevance, and clarity. These rules are not prescriptions that we always respect, but they work as the implicit default assumptions that speakers normally share about what is to be expected from a linguistic exchange. In consequence, each ostensive violation of these maxims that is produced by the speaker reveals secondary meanings and intentions to the listener (*implicatures*): If someone says "What a beautiful day" in the middle of a deluge –thus, ostensibly breaking the maxim of truth–, the listener would interpret that the intention of the utterance is not the literal meaning but a secondary one, such as irony, for example.

Ornamental rhetorical elaborations at the level of the sentence are not only numerous across poetry, but also across narrative. We know, for example, that in the title *The Book of the Thousand Nights* (VV.AA., 1706-21), *nights* is a metonymy to refer to the stories that are told during that time, and *a thousand and one* is not an accurate sum but a rhetoric way of expressing *many*. However, if these rhetoric turns would be removed, and we would talk directly about *The Book of the Many Stories*, the result would be perhaps clearer, but also less appealing. The frequent elaboration of ornamental rhetoric in narrative can be illustrated by recalling virtually any famous story opening, from the classic *Once upon a time*… to *It was the best of times, it was the worse of times, it was the age of wisdom, it was the age of foolishness…* (Dickens, *Tale of Two Cities*) or *Lolita, light of my life, fire of my loins...* (Nabokov, *Lolita*, 1955)

The ornamental use of rhetoric is actually universal (Gottschall&Wilson, 2005; Dutton, 2010). Everywhere, especially when producing or consuming literature, people tend to elaborate, expect, recognize, repeat, standardize, and invent ornamental rhetorical and formal feats. And they do these things even when dealing with forms of language that are not typically considered as literary: such as the usage of puns and word games in jokes, anecdotes, and slogans –*Don't be trashy, don't you care? Recycle, don't just stare*; *Pick up the pen and vote for Ben*; *Reading while sunbathing will make you well read; Insect puns bug me.*

Not only we use profusely this kind of ornamental rhetorical procedures, but our very linguistic faculty seems to be designed for favoring such practices. This can be seen, for example, in the size of our vocabularies. Statistical measurements across languages show that each person knows an average of 60,000 words; however, in 98% of her speech, each person uses an average of 4,000 words (Miller, 2001). And it is a universal phenomenon: humans regularly tend to learn 15 times more words than they actually use. This fact does not seem directly useful, but it is, indeed, consistent with our rhetoric behavior: the display of a variety of synonyms is a typically a valued linguistic feat –often independently of whether talking about "Homer," "The author of *The Iliad*," or "The Greek poet" really introduces more clarity to a text.

But why is it that we have all these strange tendencies in the first place? Why is it that the minds and brains of humans are designed to make us so skillful at using ornamental rhetorical language, and why are we so prone to do it and so sensible to enjoy it? In short, the question I am trying to raise here is the following: How did evolution made of us rhetoric creatures? Evolutionary psychologists have actually explored these questions intensely in the last decades. And the account they provide can give us a great insight onto why human literatures are the way they are and why we respond to them in the ways we do.

As said, we are not only very skillful at elaborating rhetorical uses of language, but we also feel attracted to them, we have an evident impulse to perform them and consume them. So, why is it that we have this *rhetoric instinct*? Evolution is fundamentally utilitarian. In consequence, if we have an instinctive tendency to rhetoric practices, there must be some value to it. And, effectively, psychological experiments showed that the exposure to and practice of playful and ornamental uses of language have measurable cognitive advantages: language play facilitates language learning (Bebout&Belke: 2017); people that read stories more frequently score better in psychological tasks related to empathy, imagination, and memory (s. Abbot, 2010); and even nursery rhymes have been proven to have strong effects in children, such as helping them to overcome dyslexia (Goswami&Bryant, 2016). These advantages are also reflected in the extreme beyond-use size of our vocabularies: we derive an advantage, a cognitive training, from developing our verbal skills.

The hypothesis that many evolutionary psychologists formulated, on the basis of these observations, is that our rhetorical instinct could simply be a training mechanism, a kind of mental sport that was naturally selected for being useful for improving other cognitive skills (Boyd, 2009; Miller, 2001; Dutton, 2010). If being inclined to engage in rhetoric play effectively makes us smarter, it would make sense that the individuals inclined to do it more would become thereafter fitter and would eventually outnumber the alternatives.

This hypothesis has also been applied to other cognitive capacities that are in the basis of all our artistic performances (s. Dutton, 2010). For example, our capacity and tendency to produce visual arts has been considered to be acquired by our species because this practice trains our vision, which is advantageous for better judging our surroundings (Ibid.). And our capacity and tendency to produce music and dance has been attributed to the fact that these practices train our sense of sound recognition and rhythm, which would be advantageous for making us more skillful in every task that requires repetition, such as walking, hammering, swimming, or even having sex (Ibid.; Pinker, 2007).

Now, this could explain our tendency to practice rhetorical uses of language, but why do we feel so attracted to these practices? Many other activities also train our cognitive skills: playing chess in your mind, doing algebraic calculations, memorizing cards… Nevertheless, these kinds of things do not give us the kind of pleasure we seem to take from artistic practices –for which we even have coined a specific name: *aesthetic pleasure*; and, in its verbal form, *poetic pleasure*. Expressions that are common across languages, such as *chat somebody up* or *sweet talker*, suggest that our rhetoric abilities even have a significant role in courtship: we feel, indeed, *attracted* to good stories and good storytellers (Miller, 2001). How can this attraction be explained? A more recent evolutionary hypothesis addressed this problem by integrating the aforementioned account within a larger framework, which considers not only natural selection, but also *sexual* selection.

Organisms do not only adapt to their physical environment, but also to their social environments, that is, to each other. This second form of adaptation was conceptualized by Charles Darwin as *sexual selection* –first formulated in *The Origin of the* Species (1859) and then further developed, especially in relation to humans, in *The descent of man* (1871).

Darwin considered sexual selection as a competition that took place in two realms: within members of the same sex (for gaining mating priority) and in relation to members of the opposite sex (for obtaining mating choice). The fundamental principle of sexual selection can be summarized as follows: Some mating choices produce fitter offspring than others; then, the choosing criteria that are more advantageous will in consequence get amplified over time as instinctive attractions to certain traits. These traits, in their turn, will also get amplified over time –because they will be more frequently selected– as *fitness signals*: indicators of genetic quality to be sought. This way, sexual selection designs strategies in every species for guiding our mating choices, making peahens attracted to peacocks with large and colorful tails, bower birds attracted to skillful bower builders, and –hence the hypothesis– humans attracted to individuals that exhibit artistic skills. Indeed, many evolutionary psychologists, consider that human arts –including the verbal ones– would be an expression of a phenomenon of this kind (s. Miller, 2001; Dutton, 2010): "A way to attract prospective mates with one's skills and, indirectly, with one's genes." Let us analyze the distinctive characteristics of sexually selected traits so as to consider to what extent they could account for our rhetorical instinct.

For a trait to work as a fitness signal, to begin with, it must be observable and it must effectively predict a genetic advantage –this property is called *principle of honesty.* Some fitness signals indicate physical advantages, and these ones typically manifest in observable bodily traits. For example, peahens prefer large trains with colorful plumages –which predicts strength (Petrie, 1994). But, at the same time, also mental qualities (such memory, visual cognition, etc.) are pertinent for organisms. Therefore, a mechanism had to be selected to allow organisms to recognize cognitive fitness, and this is the function that *behavioral* signals have: instinctive ways of behaving that are correlated with cognitive fitness and can be observed by other organisms. For example, male satin bower birds (*Ptilonorhynchus violaceus*) create stick structures (called *bowers*) which they carefully decorate with colorful bright objects –v.g. stones, leaves, etc., especially blue and yellow. These structures are not functional in themselves –bowers are not nests–, but they are useful for indicating the skills of the builder: indeed, female bowerbirds visit the bowers during courtship, and –in response to many variables, including symmetry, colors, and size– their mating choices are influenced (Endler, 2012).

The same principle seems to be applicable to human artistic practices –in fact, the complex criteria by means of which female bowerbirds judge the bowers has been often described in terms of *aesthetic experience* (Ibid.). Humans also have consistent attractions to particular bodily traits that are evolutionary explainable in that they are correlated with genetic fitness: for instance, we tend to judge as attractive faces that are symmetrical –which predicts healthy development– and faces that look closer to the average of our society –which predicts genetic diversity (Little et al., 2011). In fact, cross-cultural research has verified a wide variety of aesthetic universals, in addition to these, that humans share with regards to bodily traits: e.g. female waist-hip ratio, male upper-body mass, tallness, age (Etcoff, 1999). BuT we also give a high value to psychological traits –an extensive cross-cultural survey showed that, in fact, the most valued characteristics by men and women are of this kind: intelligence, humor, honesty, and kindness (Lippa, 2007). So, it would make sense that we use some kind of behavioral signal to indicate the possession of these fitting cognitive traits.

Our rhetoric practices are a good candidate for this: since –as mentioned earlier– they are effectively correlated with fitter cognitive capacities (s. Bebout, Belke: 2017; Abbot, 2010; Goswami; Smith et al., 2005). This means that rhetoric skills could, in principle, function as an effective criterion for guiding humans into choosing fitter mates. This idea reveals that pehaps a deeper truth lays in the traditional notion of hermeneutics according to which "the art of interpretation consists in discovering the subject behind the object" (Scruton, 2013).

But there are also other characteristics that are typical of sexually selected traits, which must be considered. As listed by evolutionary psychologist Geoffry Miller, these would include: a wasteful virtuosity, a quest for status, an instinctive sense of attraction to them, and the manifestation in courtship (Miller, 2001). Let us consider each of them.

Sexually selected behavior typically evidences different forms of virtuosity and wastefulness. Showing an excess of a resource indicates that the organism must have much more of it. This is how sexual selection explains the display of ornamental behavior that can be found among a great diversity of species: If a bower bird is capable of investing a lot of effort in decorating a useless bower, it must be a very resourceful bird. This same logic would be applicable to the ornamental rhetorical usages that we have mentioned in this section. And it would also explain the extreme beyond-use size of our vocabularies (also mentioned earlier): Counting with an unrequired variety of words that one can strategically display (as a wasteful virtuosity) in particular situations might be useful for indicating that one has enough resources to spend time learning these extra words and acquiring the ability to use them creatively, which indicates in its turn other cognitive qualities, such as conscientiousness and memory, and makes of this display an effective fitness signal that an individual can use to appear more attractive (Ibid.; Dutton, 2010: 146-9). In short, a large vocabulary might represent for humans what large tails represent for peafowls.

Sexually selected traits are also related to a quest for status: since fitness signals increase the mating opportunities of organisms, possessing these traits ends up being valued as a criterion for establishing the position of the organism in the social hierarchy (Miller, 2001). Indeed, verbal skills (including the exhibition of lexical variety and rhetorical ability) are highly valued not only in courtship, but also in most fields of human culture: from comedy, politics, and education, to religion, law, and even business –Pierre Bourdieu analyzed this and other markers of status in terms of *cultural capital* (1984). Also cultural productions such as dictionaries, rhetoric treatises, *ars poeticas*, writing schools, manuals of style, literary canons, literary contests, and even spelling contests are evidence of the value that humans give to linguistic skills as status providers. A further evidence of the relationship between language and status can be seen in the attention people tend to pay to certain linguistic norms that do not affect the informativity (*sensu* Grice) of the utterances. For instance, in English, no actual informative disadvantage is produced by ending a sentence with a preposition. However, the capacity of a person to observe that (subtle and useless) rule is consistently considered by people as a criterion to assess the speaker's cultural capital (*sensu* Bourdieu), which is a measure of status. The same occurs with regards to the correlation between the level of status and the level of acceptable orthography –the higher the status, the higher the orthographic expectations, independently of the informative intention (recall Saroyan's poem in the first chapter and the dimension of the reactions to his misspelled *lighght* (Saroyan, 1962), which fewer people would reprimand in more informal contexts, such as a chat or an email.

Finally, considering the human arts as results of processes of sexual selection provides us also with a frame for understanding the kind of enjoyment we associate with these practices: *aesthetic pleasure*. The fact that a fitness signal becomes selected to facilitate mating choice entails that also the attraction to that trait is acquired by the pertinent members of the species. The peacock's train serves as a fitness signal not only because it predicts genetic quality, but also because peahens feel attracted to it. So, why would we be programmed to find any pleasure at all in producing and consuming literature or any other artistic product? Pleasure and pain constitute the systems of rewards and punishments by means of which evolution fosters fitter behavior in organisms. The behavior fostered by aesthetic pleasure can be explained, in this sense, as the adaptation that drives us into seeking and displaying the signs that can increase our chances of finding a cognitively fit mate, or being chosen by one. This is how –we could say– nature created the subjective experience of beauty, and this is also why aesthetic pleasure can be considered a *natural* component of human arts.

CHAPTER 4

STORY

HOW WE PROCESS NARRATIVE

Honduran-Guatemalan author Augusto Monterroso wrote in 1954 the micro-story "The Dinosaur," which consists of merely one line:

> When he woke up the dinosaur was still there.

One could argue that no story is really narrated in this line. Nevertheless, after reading those nine words, most people are capable of recreating rich storylines in their minds. We can conceive somebody thinking, for instance:

> "…*Still* there"? This means the man must have encountered the dinosaur before... It must have been quite a surprise! He must have wondered if he had traveled to the past, if the dinosaur had traveled to the future, or if everything was a dream... Did he fall asleep afterwards or did he passed out immediately due to the astonishment of seeing a dinosaur…? In any case, how shocked must he have been when finding the dinosaur again, since, "when he woke up, the dinosaur was *still* there"!

You might have imagined a similar or a different storyline; but you most probably imagined one. And the incredible thing is that the original text only states the brief final line, the rest of the story is created by our minds, intuitively and effortlessly, as if filling the gaps of an imaginary structure by a literary inertia. But how do we do this? How can such a rich cognitive process be elicited by such a limited input? How do our brains allow us to *think* narratively? And by means of what rationale did evolution give this singular capacity to our species? These questions are the subject of this chapter.

A significant part of human culture is constituted by stories: legends, myths, fables, novels, song-lyrics, theater plays, jokes, anecdotes, comics, films, series, advertisements, political speeches, journalistic reports, biographies, gossip, history, ideologies, etc. Actually, psychologists, anthropologists, linguists, and empirical literary scholars have recurrently pointed out to the fact that stories are universal in our species: every human society intensely produces and consumes stories (s. Brewer, 1984; Brown, 1991; Mooiji, 1993; Bruner, 2002; Hogan, 2010). Our minds seem to be especially programmed for stories and a great part of our culture and our mental life is organized as stories. Stories are such a singularly human feature, and they are so crucial in our lives, that evolutionary psychologist Jonathan Gottschall has decided to call us *the storytelling animal* (2013).

Inquiring the natural origins of our narrative skills will enlighten many aspects of our literary practices and our reader-responses, such as why most literature is narrative, why we remember more easily story-events that contribute causally to the plot, why stories are so full of conflict, why we tend to judge stories morally, why we enjoy stories so much, and why we see stories even where no stories are being told.

THE NARRATIVE SCHEMA: THE UNDERLYING STRUCTURE OF NARRATIVE COGNITION

What is the mental software that allows us to organize information in a narrative form to begin with? In cognitive terms, narrative can be described as a kind of *schema* –i.e. a mental structure for processing information (Hogan, 2003). There are schemas of many kinds: Thanks to representational schemas, we can recognize an object by only seeing one of its parts (e.g. when seeing a hand sticking out of a window, we immediately realize that it must belong to a person –and that it is not just floating in the air– due to our human-body schema); a script-schema guides us through the protocols of conventional situations (e.g. when eating at a restaurant, it lets us know when to sit, when to order, to whom, etc.); a procedural schema guides us into skills that require know-how (e.g. riding a bike, dancing, or playing piano); and it is a narrative schema what allows us to understand that Theseus entering the labyrinth, killing the Minotaur, and escaping are three interrelated events of the same story. The narrative schema is the mental mold made of predefined rules and roles that we use for cognizing information in that particular way: namely, in term of agents that perform intentional actions along a timeline (s. Brewer, 1985). And, as with any other schema, if we discover its structure, we can predict its outcomes.

In 1932, Frederic Barlett designed an experiment meant to trace patterns in the way people modify received information when they recall it later, and he bumped into the effects of the narrative schema (Barlett, 1932). He presented a heterogeneous group of people with information unfamiliar to their cultural backgrounds –the Native American folk tale "The War of the Ghosts." Afterwards, he analyzed how different pieces of information about the story were recalled. What he found was that all the participants modified the information they had received, and not randomly, but in systematic ways that he categorized as follows (Ibid.):

- Omission: Excluding or ignoring information the test subjects considered irrelevant.
- Shift of focus: Emphasizing information the test subjects considered more relevant.
- Transformation: Altering details of the story (such as order of events or description of places).
- Rationalization: Redeveloping aspects of the story that did not make sense for the test subjects (e.g. creating new information, adding explanations, etc.)
- Cultural shifts: Altering content and style of the story to provide more coherence and appropriateness in terms of the cultural backgrounds of the test subjects.

Since the modifications introduced into the story were not random but followed consistent patterns, Barlett inferred that people must use similar principles to organize the information they receive. These principles of organization are today known as schemas, and further experiments have verified the effects observed by Barlett and discovered others with regards to the narrative schema (s. Brewer, 1985). The following sections will explore how the human mind might has acquired this schema and the ability to use it productively.

THE STORY INSTINCT

Some aspects of narrative cognition make an evolutionary account particularly pertinent. The first one is that –as mentioned in the beginning– narrative seems to be a human universal (s. Brown, 1991). In addition, the capacity to cognize narrative is manifested by children at a very early and regular age and without need of special lessons. These are distinctive characteristics of the kind of behavior whose development is determined by innate structures (such as learning to walk or to speak) in contrast with the behavior whose acquisition depends more on a particular cultural nurturing (such as reading, writing, or playing the guitar). According to some psychologists, evidences of narrative cognition appear in infants even before the development of language, in fact, as soon as children start to perform pretend play –i.e. from ages 1 to 3– (s. Russ&Wallace, 2004; Boyd, 2009; Donald, 1991): including the elaboration of scenarios, characters, goals, obstacles, and even drama (Russ&Wallace, 2004; Dissanayake, 2011). The inquiry on the biological grounds of narrative cognition has also fostered research in comparative psychology, and some researchers have claimed that even certain nonhuman primates show –particularly in their social behavior– signs of an underlying rudimentary form of narrative cognition (Dautenhahn, 2017) –which would be consistent with the view of *mental continuity* already formulated by Darwin (1871): the idea that the cognitive differences between humans and other animals are of degree, not of kind.

A last crucial characteristic of our narrative cognition is that we are not only capable of cognizing stories, but we also have a strong drive to do it: our minds *narrativize* perceived information frequently, automatically, and effortlessly. The first and most famous experiment aimed at measuring this phenomenon was performed by Fritz Heider and Marianne Simmel in 1944. It consisted in showing to a group of people an 80-seconds video of geometric figures moving across a white background (s. Figure 1). When asked what they had seen in the video, the answers of the test subjects varied widely. But what was most significantly consistent was that 97% of them reported having seen a *story* (where each figure represented a character that performed certain actions), whereas only 3% reported having seen only what was literally shown in the screen: geometric figures moving across a white background (Ibid.).

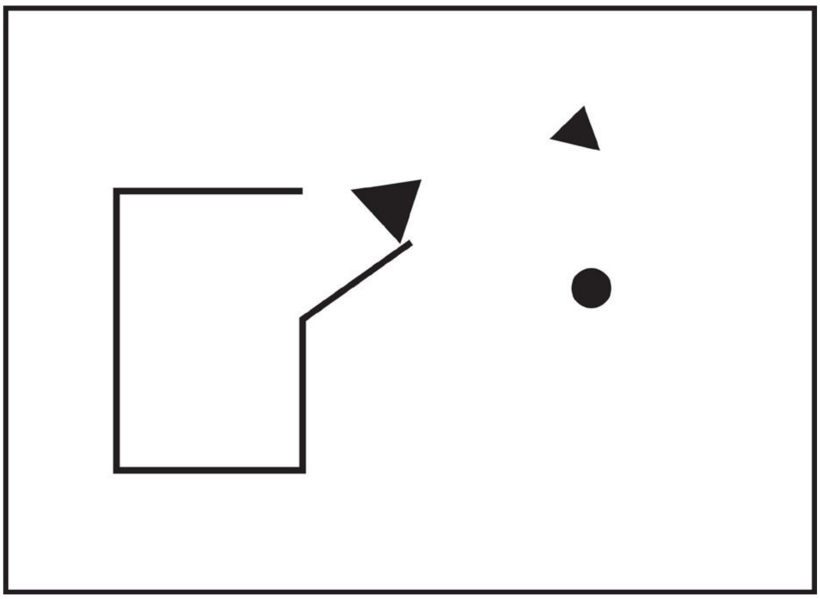


Figure 1: Exposure-objects displayed in various positions and configurations from the moving film. Large triangle, small triangle, disc, and house. (Heider & Simmel, 1944).

For example, this readiness to cognize information narratively is what allows us to –among other things– use and understand the rhetoric procedure of *ellipsis* (i.e. the omission, within the exposition of a story, of a part of the sequence of events). By placing each piece of information into a narrative schema, our minds display a structured series of inferences that guide us into how to interpret the input, what to expect from it, and how to complete the missing information –in the form of a story–, even when very little is explicitly told. If our narrative cognition functions is such a consistent and predictable way, it can also be manipulated –and this is exactly what skilled writers do, as we have seen in the aforementioned example of “The Dinosaur.”

But the effect can be triggered by even scarcer stimuli. In fact, we do not even need a whole sentence. Here is a limit hypothetical example. Try to read the following words and *not to think* of a story:

LION - MEAL - CHILD

I have been running this informal experiment, asking people to tell me what they think this three-words text is about. And most people answer by telling me a micro-story: “Ah, I get it: a lion ate a child, as its meal” or “It is about a lion and a child who share a meal” or even “It could be about a child that eats lion-meat.” I did not mean to write any specific story. I deliberately did not include sufficient elements for the text to be conventionally considered as an actual story in the first place –the words are all nouns, there are not even verbs or any other grammatical markers of the relationships between these words. Nevertheless, people still see a story there: they automatically recruit a structure of agents performing intentional actions on objects along a storyline, and they fill those roles by considering their knowledge about the mentioned nouns. This suggests that minimal linguistic stimuli seem to be enough to trigger our story-schema. Evidently, we are biased to automatically narrativize linguistic cues and organize them into a story-form, even if we have no logical grounds to assume that the relationships and information supplied by this schema is actually pertinent for the case –even if we are explicitly requested *not to think* of a story: Did you manage to repress your story instinct when cognizing *lion*, *meal*, and *child*?

On the basis of the observed universality of narrative cognition, the way in which it is developed by children, and the readiness with which we apply it, some psychologists have considered that narrative cognition abides to the description of an actual *instinct*: an “innate, species-specific biological force that impels an organism to do something, particularly to perform a certain act or to respond in a certain manner to specific stimuli” (s. *APA Dictionary of Psychology*, 2018) (s. Gottschall&Wilson, 2005; Gottschall, 2013). If narrative cognition is effectively an instinct –and not simply a contingent mental technique transmitted by some human societies purely through cultural nurturing–, an evolutionary account is certainly due. Then, the central question to be asked would be about its functional rationale: What advantages could an instinct to cognize narrative have provided to our ancestors? Some psychologists consider that our storytelling instinct is not really an adaptation but an *exaptation*[15] (e.g. Pinker, 2007b). However, all agree that it must have entailed actual advantages because it is so costly –otherwise it would have been long wiped out by natural selection as a hindering. As for the advantages attributed to narrative cognition, these are four of the most important current hypotheses:

1- Narrative cognition is useful for processing chains of causal relationships,

---

[15] *Adaptations* are traits that evolved for entailing a direct advantage for the survival and/or reproduction of organisms. *Exaptations* are secondary consequences, or by-products, of the adaptations (Gould&Vrba, 1982) Exaptations are still evolutionary acquired, but not directly for the benefit they represent on their own. For instance, the for of our hands and the constitution of our skin is *adaptive* (directly functional), but the lines in our hands are *exaptive* (a secondary consequence of the form of our hands and the constitution of our skin).

2- It is a source of vicarious experience,
3- It helps us bond socially,
4- It is a sexually selected fitness signal for facilitating mating choice.

Considering the evolutionary functions of our capacity and tendency to cognize narrative is very important for another reason: each of these hypotheses will give us a particular insight into the nature of human language and literature.

1. CHAINS OF CAUSALITIES: STORIES AS TOOLS FOR MAKING SENSE OF DISTANT CAUSES

The first hypothesis states that we use our narrative instinct to process chains of causalities and thereby spot distant causes to explain different effects (s. Broek, 1990; Herman, 2003; Gottschall&Wilson, 2005; Gottschall, 2013). That is, to *make sense* of the world (s. Bruner, 2002), in terms of causal structures.

Recent research in comparative psychology suggests that –in contrast to the traditional view– the capacity to perform causal inferences is not exclusive of our species. It has been identified in nonhuman primates as well (Rodríguez, 2016), especially in the ones capable of using tools (in terms of the associations between a tool and its effects) (Povinelli, 2000), and, according to some, even in rats (Blaisdell et al., 2006). Blaisdell and his team trained a group of rats to distinguish between two kinds of cause-effect relationships, that they called *observation* (seeing) and *intervention* (doing). In the first scenario, a light would turn on and the rats would observe that, then, a piece of food would appear in a little well. In the second scenario, the rats were offered a lever, and by pressing it themselves, the light would also turn on, but in this case no food would appear in the well. Finally, the rats learnt to distinguish between the two variables: They would only check for food when they observed the light turning on by itself, but they would not do it when they saw the light had turned on by the intervention of another rat. This behavior cannot be accounted by a mere Pavlovian association: even though the light was the same in both cases, the rats could interpret it as produced by two different kinds of causes and successfully predict different outcomes thereafter, as evidenced by their consequential behavior. Therefore, the explanation seems to be that rats can think in terms of causality.

We may not be the only animals that understand relationships of causes and effects, but, still, we do it with a singular mastery. Whereas a rat might be able to cognize, for instance, different connections between presence/absence of light, the presence/absence of food, and observational/interventional causes, humans are capable of understanding much more. Namely, we can understand that the casual relationships identified by the rats –albeit effective in their environment– are false or at least superficial: the food did not appear because the lights turned on, but because a researcher put it there, and he put it there because he wanted to make an experiment, and he made the experiment so as to prove a hypothesis, etc. That is, we can think of every event as the effect of a cause, and this cause as the effect of a further cause, in a virtually indefinite manner.

This generative property is characteristic not only of the human mind but, accordingly, also of human languages, which provide us with a finite set of words and rules with which we can create and understand infinite utterances (s. Chomsky, 1968). The claim of the hypothesis presented in this section is that it is the narrative schema the one that enables us to perform this generative mechanism in relation to causal structures. In this sense, the popularly known fact that children tend to ask "Why…?" to every statement and "…But why?" again to every answer, insistently requesting causal explanations, could be interpreted as a naturally driven behavior oriented at developing narrative cognition –just as asking for the names of things makes part of children's vocabulary development.

By allowing us to extend causal schemas, story cognition is extremely useful beyond literary uses: it fundamentally makes us more skillful at making sense of the world, increasing our capacity to organize information, and to make predictions and inferences. Indeed, we normally apply narrative cognition also to make sense of individual static elements that we encounter in the physical world: We see a brown spot in the carpet and our mind immediately explains it by a reverse-engineering narrative: "I guess my colleague just came to my office, spilled the coffee that was on the desk, and took away my cup." This kind of narrative inference, as it can be appreciated, is analogous to the one we used at the beginning for interpreting "The Dinosaur." The enormous source of knowledge that narrative cognition represents for our minds, merely by means of extending our causal schemas, would have been advantageous enough to justify the natural selection of this mental faculty.

If narrative cognition is fundamentally an extended form of causal reasoning, then, neural correlations should be found between these two mental performances. And, indeed, they have: Michael Gazzaniga identified the brain circuitries that produce both causal reasoning and narrative cognition as being in the left hemisphere of our brains. He called this neural system the *Narrative Interpreter* (Gazzaniga, 2000). In a today famous experiment, Gazzaniga and his team showed to a group of split-brain patients flashes of images from their right side and from their left side (so that they were processed by the left or right hemispheres, respectively): the image of a chicken claw was fed to the left brains of the patients and an image of a snowy landscape to the right brains. Then, they asked the patients to pick up, from a set of pictures, the ones that were related to the ones they had seen. As expected, with the hand controlled by the left brain they picked up the picture of a chicken, and with the hand controlled by the right brain, the picture of a shovel –this manifests the typical behavior of split-brain patients, each of whose brain hemispheres works in a seemingly autonomous way. Now, when asked to justify why they had chosen a chicken, the patients naturally explained "Because you showed me the picture of a chicken claw." But, when asked why they had chosen the shovel (for which they had used the right brain), for some reason they could not recall the snowy landscape that had been fed to that brain hemisphere. Gazzaniga's hypothesis for accounting for this phenomenon was that the cognitive task of processing explanations (i.e. "Why did you chose that image?") is performed by the left hemisphere of the brain –hence, since these were split-brain patients, their left brains could not access the information of their right brains, and therefore they failed to bring up the image of the snowy landscape as the cause of their choice. This brain system that Gazzaniga identified would function as a processor for making sense of the constant flux of information we receive from the environment. Further neuroimaging evidence showed that this system is effectively employed each time we cognize narrative (Ibid.)

In addition, Gazzaniga's experiment also tells us a crucial aspect of how our narrative cognition works: when the left brain lacks information, it *invents* reasons with the only information it does have. Even though when trying to explain their actions (left brain) the test subjects could not access the information about the snowy landscape that had led them to chose the shovel (right brain), the patients would not say "I don't know why I chose the shovel;" but, instead, they would answer things like: "Because a shovel is needed to collect chicken coop." That is, the left brain searched for the information it did have (the image of the chicken claw) and immediately made up a connection with the shovel to justify the previously performed actions. But the patients did not experience these answers as lies they had just come up with, instead, they were convinced that those were the actual reasons of their actions (Ibid.) –which means that our left brain is so insistent in trying to make sense and elaborate explanations and justifications that it has a downside: when it does not find an explanation, it creates lies with which it persuades our own conscious selves (s. Gottschall, 2013: 46).

Now, postulating this function for the narrative instinct allows us to hypothesize expectable reader-responses and literary and linguistic universals. Namely, if allowing us to conceptualize long chains of causalities is the fundamental evolutionary function of narrative cognition, we can thereafter hypothesize, firstly, that, when reading a story, we should expect people to pay particular attention to causally-related events and to remember them more easily; secondly, that it would be expectable that literatures around the world be dominantly causally-structured; and thirdly, that human language should show sings of being designed to express chains of causalities.

As for the first prediction, that is precisely what Frederic Barlett had proved (as mentioned earlier): people do tend to focus on the causal chains that structure stories, to recall them more easily than other pieces of information, and even to automatically correct the causal deficiencies the stories might have (Barlett, 1932). At an intuitive level, this thesis is also certainly plausible: we may pay attention to many aspects when reading a novel or watching a film (e.g. characters, style, images, etc.), but we only feel that we are *following* the novel or film inasmuch as we are making sense of the plot, that is, of the causal structure.

As for the second prediction, the dominance of chains of causalities is verified in the widespread dominance of plot-based literature across cultures. If there is a shared formal aspect across literatures of the world, it is the profusion of causally-structured narrative. Novels, short stories, novellas, theater plays, popular legends, myths, and even a great deal of poetry (such as epic poetry) are built up around the causal structures of plots. Even when stories are not realistic, when they include fantastic elements, they typically are still causally-structured in terms of the rules set in the hypothetical worlds where they are set. For instance: The events told in Ariosto's *Orlando Furioso* (1532) or J.R.R. Tolkien's *The Lord of The Rings* (1954-5) do not correspond to the events that can possibly occur in our world, but still, the narration of these events is structured in the form of chains of causality of which our mind can make perfect sense. Supernatural or fantastic literature does not mean absurd, irrational, or non-causally-structured literature. Non-narrative literature obviously exists (such as contemplative or introspective poetry), but to a significantly lesser extent. The dominance of narrative literature is also confirmed in the literary market: novels have much larger audiences than poetry. And, even though in the past (in the oral origins of literature), poetry was the dominant form (i.e. versification), poetry was also more narrative (as it can be verified in ancient epics, for instance).

The predominance of causally-structured plots in literature can also be shown in that these plots usually work as a condition of possibility for making sense of other formal or semantic structures. For instance, it is true that many stories rely heavily on symbolic structures that we process through analogic reasoning. Dante's *Inferno* (Alighieri, 1320), for example, has been often considered as representing a satire of the politically ruling class of the Florence of the 14th Century (s. Ferrante, 2014), but that symbolic dimension cannot be understood without relying primarily on the causal structure that makes it possible for us to conceptualize those secondary meanings. Likewise, we cannot argue that Herman Melville's *Moby-Dick* (1851) symbolizes the struggle of man against his inner demons, for instance, without referring to the logically-prior causal storyline of Ahab chasing the whale and its consequences.

At the same time, if our minds are instinctively driven to seek causal structures in stories, this would explain why literary works that resist being so interpreted –such as completely surrealistic or absurd novels– are more the exception than the rule. There are, of course, many historical examples of such literatures –from André Breton's novels to Eugène Ionesco's plays (s. 1994)–; but their popularity has always been restricted to minoritarian cultural groups. The Surrealist movement (between the 1920s and 1930s, especially in France), for instance, promoted this kind of non-causally-structured writing, but their use of this and other techniques was precisely aimed at shocking the common sense (*épater la bourgeoisie*). So, we could ask: Why would causal-structured narratives would be an expression of common sense to begin with? What we are arguing here is that, it is indeed common sense, although it does not have a culture-specific origin, but an evolutionary one. It is, in a way, the "natural" form of literature. And, with regards to it, the uneasiness that many people effectively reported experiencing when confronted to surrealistic works can be accounted as a form of frustration resulting from the repeated attempt and failure of people's minds in trying to find the causal structures of texts that were purposely designed to oppose that logic.

The functioning of open endings, as well as the particular effect they produce, are also accountable in relation to this. An open-ending story suggests the existence of a causal structure but, at the same time, fails to provide sufficient elements to reconstruct it in its complete form, thus leaving us eagerly wondering: Was Josef K actually guilty of anything (in Kafka's *The Trial* -1959)? Was Miles killed by a ghost (In Henry James' *The Turn of The Screw* -1898)? Did Melanie survive the attack of crows and seagulls (in Hitchcock's *The Birds* -1953)? This human eagerness for giving closure to the causal chain of the events told in stories is also the principle by means of which cliffhangers work: It is the curiosity of Prince Shahryar that Scherezade leveraged –exciting it and leaving it unfulfilled with every unconcluded story– so as to persuade him to keep her alive for yet another night (in the *One Thousand and One Nights* –VV.AA.,), and it is the same curiosity that moves us still today to crave for the next episode every time we watch a series.[16]

Finally, the hypothesis that our narrative instinct is an adaptation for processing chains of causalities introduces a new explicative principle for language evolution. In particular, it means that human language should show marks of design fit for expressing causality. And, indeed, the kinds of word-classes and syntactic structures more universally shared (v.g. nouns, pronouns, verbs, adverbs, etc.; subject, object, predicate, complements, etc.) seem to respond quite explicitly to the requirements of expressing semantic arguments in terms of causal relationships: that is, of describing the world in terms of agents that perform actions onto objects (or other agents) with consequences in a particular timeline.

2. CONFLICT: STORIES AS LIFE-SIMULATORS

The second hypothesis is that stories are low-cost sources of information and vicarious experience (s. Baxter, 1997: 133; Gottschall, 2013: 45-67). Stories' evolutionary function would be to simulate life so that we can practice in a safe mental spaces on how to solve problems that we might face in the future. This function allows us to predict and explain another literary universal: it would be then an expectable outcome that stories be mostly about people in conflict.

Intuitively, conflict seems to be a quite prototypical feature of stories: Stories are, indeed, typically constituted by human (or human-like) characters dealing with human (or human-like)

[16] This technique will be further explored, in relation to its neurocognitive nature, in the chapter on Emotions.

difficulties in the pursuit of (human pertinent) certain goals. Popular stories across cultures and time are markedly built up around human conflicts: war, death, injustice, tragedies, and natural catastrophes are as abundant in classical as in and modern literature. Love stories are, more than about love in itself, most usually about the difficulties of finding, consummating, or losing love (e.g. From "The Song of Songs"[17] and the myth of Dido and Aeneas to *Romeo and Juliet*[18] and Hollywood movies). Even comedies typically portray people in conflict trying to overcome obstacles (e.g. Tartuffe struggles to obtain the goods of Orgon,[19] Chaplin faces poverty and jail, and the Roadrunner is constantly being chased by the Coyote). And, as research in child psychology suggests, already children stories are mostly about conflict (s. Paley, 1988).

The psychologist Vivian Paley performed an extensive a research on children's pretend play (Ibid.). She was particularly interested in finding out whether sex differences were correlated with differences in the stories children create when playing.[20] She worked with large groups of three-year-olds during several months. Paley discovered, indeed, many correlations –as one would expect if stories were effectively simulators for potential life problems, in a species in whose evolutionary history sexual differences have been consistently correlated with differences in lifestyles (s. Geary, 2010)–: namely, male infants' play tends to be organized around stories of physical forms of conflict (e.g. war, fight, chase), whereas female infants' play tends to be organized around stories of interpersonal and social forms of conflict (e.g. treason, tragedies, misunderstandings).[21] These differences were systematically manifested across varied purposely designed scenarios (v.g. creating gender-mixed groups, placing each child in the playground arranged by the children of the opposite sex, depriving children of toys, etc.). And, yet, a common element was also evident: children's play (independently of sex) was always structured in the form of a story-structured simulated conflict. This is coherent with the fact that pretend play is a common behavior across many animal species that responds to a clear adaptive rationale –being fundamentally considered as useful to prepare infants for adult life (s. Mitchel, 2007; Russ&Wallace, 2004).

Conflict is, indeed, a plausible candidate for a literary universal (s. Baxter, 1997: 133; Gottschall, 2013: 45-67).[22] As said before, it is indeed a shared element of our mental prototype of stories: the less conflicting a story is, the less story-like we tend to feel it. The story of a person that is doing well and then does better, without any troubles, effectively sounds like an unlikely literary plot. Samuel Beckett's *Waiting for Godot* is one of the scarce examples of a famous story with vague or no conflict: Two men wait for an uncertain visitor in an uncertain place for uncertain reasons. No incident occurs: the visitor doesn't even show up. Even though it is academically considered –for many other reasons– as a valuable literary work, it was never a massively popular story –and, even in this limit case, a form of conflict can still be recognized in its plot: in the unfulfilled expectation of the visitor's arrival. If the

[17] VV.AA., *The Bible*, 4000BC-96AD

[18] Shakespeare, c. 1599-1602

[19] Molière, 1664

[20] For this research, Paley purposely selected children whose parents declared being actively opposed to imposing gender stereotypes to them.

[21] These results are also consistent with the studies that show that men tend to be more interested in *things*, whereas women tend to be more interested in *people.* (S. Su&Rounds, 2009).

[22] Narrative cognition is claimed to be an *absolute* universal, that is, a capacity shared by all humans, like the capacity to see or to speak. When talking about literary universals, instead, we are in most cases referring to *relative* universals, that is, to features that are not necessarily shared by all cases, but whose appearance, nevertheless, is not dependent on cultural differences.

evolutionary function of stories is indeed to prepare us for potential life dangers and conflicts, the unpopularity of this kind of works is expectable: popular stories would expectably *seek* rather than *avoid* conflict. This hypothesis could be empirically verified across world literatures if a way of quantifying the presence of conflict would be designed –a lead for this could be found in the method of *sentiment analysis*, for example, which is largely used in digital humanities and serves to discover progressions of positive and negative emotions in texts by measuring proportions of words labeled in advance as positive, neutral, or negative.

3. SHARED IDENTITY, VALUES, AND INTENTIONALITY: STORIES AS COMMUNITY-MAKERS

Humans are social animals. How did evolution produce this? Would not evolution favor necessarily selfishness and individuality? It is, indeed, true that, if in a group made of only perfectly cooperative self-replicating organisms a single self-replicating organism would mutate to selfishness, this organism would profit from every interaction with the cooperative ones and, at the expenses of them, the selfish organisms would soon outnumber the others. But then, again, why is it that not all individuals are perfectly selfish, individualistic, and competitive? How can we explain the fact that, as we can observe in nature, also social animals exist, such as ants, bees, wolves, and humans? This is often called *the free rider problem* (cita).

Using mathematical models, evolutionary theorists have discovered that, although an indiscriminate altruism instinct is unlikely to evolve from our selfish genes, a so-called *parochial altruism* can effectively be adaptive (s. Haidt, 2012). Parochial-altruistic organisms are the ones that have a selective cooperative instinct: they tend to cooperate with their own in-group members but to compete with the others. This constitutes what, in a nutshell, can be described as selfish groups of cooperative members. When this variable is introduced, we discover something very similar to what we actually observe in nature: when members start to develop certain parochial altruism, they unite in forming larger collective organisms that become more successful than the purely selfish free riders, and end up outnumbering them. This discovery provides an insight for understanding a great range of evolutionary phenomena at different levels. It explains, for instance, why the mitochondria in every cell has a DNA different from the one in the nucleus: They used to be free-living bacteria that eventually grouped into a super-organism that was fitter than its two components left by their own. Furthermore, it explains the evolution of our parochial social instincts, and the possibility of group selection (if some groups become fitter than others in the game of life, then their most adaptive traits will be selected) (Wilson, 1975).

We are social creatures, and our parochial altruism has two expressions, two biases, which evolutionary psychologists refer to as *in-group prosocial behavior* (cooperate with your own) and *out-group antisocial behavior* (compete with others). We can distinguish this dynamic in every interaction among collectives. We debate dividing ourselves in groups (politics, religion, etc.); we have historically competed physically at a group level (wars); and when we don't have to, we entertain ourselves by simulating group competition (sports, videogames, etc. – social psychologist Jonathan Haidt considered that, actually, sports are to war what porn is to sex (2012). Evolutionary psychologists also refer to the interaction of these instincts as *tribal behavior.* These biases are so strongly rooted in our psychology that they are triggered independently of the criteria by means of which we get to feel part of a group. Several psychological experiments have proven that it does not even matter by mean of which criteria our group is a group. Even if participants that do not know each other are grouped randomly by the researcher at the beginning of the experiment, as soon as they find themselves sorted as "us" and "others," they immediately start to act out the tribal biases: unconsciously favoring their own and disfavoring the others (Mladinow, 2012).

But group structures are more complex than these simple biases. Groups have also identities, values, goals. How do we form, organize, and communicate this content? The answer, for many evolutionary psychologists, is: through stories (s. Gottschall&Wilson, 2005; Oatley et al., 2006; Boyd, 2009; Haidt, 2012).

IDENTITIY

Most of the species that evolved social behavior organize themselves by means of kinship. All wolves of the same pack are blood relatives to each other: they are bound together by their genes –and they recognize this kinship by diverse biological cues, such as their smell. The same occurs with bees and ants: social animals tend to organize themselves on the basis of their kin, as families. Humans are the only species that has managed to create gigantic social structures, that bring together thousands or millions of individuals that are not necessarily directly related. Humans are capable of creating social groups beyond kinship, around our sports teams, our political affiliations, our nationalities, our religions, our taste in art, etc. And one of the ways in which we have managed to do this is precisely by stories.

The hypothesis of this section is that this is, actually, one of the fundamental evolutionary functions of our capacity to cognize stories. Besides the aforementioned functions, stories also seem to be community-makers, a technique that we use for organizing people around a sense of identity, values, and beliefs: religions do it with myths, nations do it with epic legends, and ideologies do it with narratives of history (s. Haidt, 2012). And this view can give us a new perspective on many properties of our literary production.

As we have seen by considering the theory of embodied cognition, we are more skilled at judging concrete things, individuals, and situations than abstract concepts. That is why the usage of metaphors and examples for conveying abstract ideas is so pervasive. Stories build up simulacra of social situations (with concrete characters and events) that portray abstract ideas such as identity, goals, and values. We can observe this function by analyzing the grammar of stories.

As folklorist anthropologist and linguist Vladimir Propp showed in his classic work on Russian folk tales, when a complex human experience is displayed through a plot, we segment it into recognizable components that our intuition finds easy to label and typify (v.g. the origin of the hero, the violation of a rule, the confrontation with an enemy, etc.) (Propp, 1928). This allows us to condense and organize complexes abstract processes –where external actions and internal states are intertwined– into embodied concepts (such as "Oedipus' parricide," "The Passion of Christ," "Hamlet's dilemma," etc.)[23] that we can more easily grasp and apply to new situations. Stories, in this sense, give symbolic organization and flesh to our intuitions and observation of ourselves and others experiencing life as societies; they embody the content of the social structure.

The fundamental typological components of narrative identified by Propp have been found to be largely shared by folktales around the world (s. Gilet, 1998), and extensive examples can be found on the use of these components for portraying group-defining stories of identity, origins, and values. The myth of Abraham, for instance, gives Jewish people a

[23] S., respectively, Sophocles, *Oedipus The King*, 429BC; VV.AA., *The Bible*, 4000BC-96AD; Shakespeare, *Hamlet*, c. 1599-1602.

conceptualization of their origins, on the basis of which they derive the categorization of the in-group (*The Chosen People*) versus out-group (*goyim*) individuals. It is, indeed, a quite typical property of social groups to posit the story of a founding member, and to ground thereafter identity labels that allow them to define themselves in opposition to outsiders of the group, and to conceptualize their values, goals, and criteria of membership by means sharing a common story.

This can be coupled with the way in way in which Carl Jung viewed literature, myths, and legends: as the primarily way in which humans intuitively conceptualize the *collective imaginary*, by creating stories that express ancestral patterns of behavior –which he called *archetypes* (Jung, 1969). A very good example of one of this archetypes that appears universally across literatures is the one *the hero* –conceptualized and described by Joseph Campbell (1949)–, which is used by social groups to portray a unified ideal of virtue to follow. Not only superhero's stories have heroes, but also the myths of religions, the official histories of nations, and even music fan groups and football teams.

We use narrative cognition to create stories about ourselves –that give us a unifying idea of who we are, where do we come from, and what our values are– and stories about our social groups –that give us a unifying idea of who our group is, where if comes from, and what its values are.

VALUES

An evidence that confirms these observations about the social nature of stories is the tendency of our literature to portray moral issues and our tendency to judge literature in moral terms. Literature created with an explicit moral aim is certainly abundant across cultures (v.g., parables, fables, exempla, etc.). But that is not the only kind of stories where we find moralization. The portrait of moral dilemmas appears in virtually every human story: novels, movies, song-lyrics, children's tales, even in videogames. Stories are often structured in relation to a group of characters paradigmatically portrayed as good and another group paradigmatically portrayed as bad, for instance (protagonist/antagonist). And the fact that our literature is organized in this manner is correlated with our behavior towards them. Psychological experiments showed that we have a great propensity to evaluate stories morally, we enjoy judging characters in moral terms: celebrating when their good actions are rewarded and their bad ones are punished, when the enemy is defeated and the hero triumphs.

We judge stories morally, and we also do it with a bias: towards reaffirming the values of our own group –which is naturally correlated with our aforementioned parochial altruism and tribal tendencies. In a series of experiments, social psychologist Jonathan Haidt showed that whether we find our own moral values reaffirmed in a story constitutes, in fact, one of the strongest limits in our capacity to engage with stories and enjoy them (Haidt, 2006 & 2012). Haidt tested this behavior by making different groups of people read stories that portrayed social taboos (such as incest, gratuitous murder, cannibalism, etc.), which became apparent at specific points of each narrative. Whenever a story seemed to show acceptance towards scenarios that people found inacceptable, the test subjects would immediately manifest rejection of and resistance to the story.

This thesis does not claim that we cannot enjoy stories that portrait suffering and pain, but only that the valuation of these events that a story displays is key for our moral judgment. We certainly can enjoy stories of misfortune and despair. Whole genres (such as tragedies or horror stories) are centered around emotions of this kind. But a story being tragic does not necessarily mean that it promotes immorality. Oedipus, for instance, is portrayed in Sophocles' play as a hero that submits himself to be punished when he discovers he has committed atrocious acts.[24] The story ends up very badly for everyone, but it doesn't promote incest in crime. On the contrary, it is a deeply moralizing in that it positively promotes values of its community, condemning incest and crime and saluting the socially agreed corresponding protocol: the proper inquiry, legal process, and due punishment. As a counterpoint, it is illustrative to mention the first story used by Haidt in his experiments: it did not include any tragedy, it was instead a positive love story, but about two siblings. Despite the romantic topic and the happy resolution of the plot, around 80% of the participants (cross-culturally) still showed rejection and resistance towards the story from the moment they knew that the relationship acclaimed in it was incestuous (Haidt, 2006). Our appetite for stories can accept tragedy (undesired outcomes), but it systematically rejects alien morals (unjust judgments).

This tendency is precisely what the literary concept of *poetic justice* describes: our bias to favor the "good ones" in stories and expect their success, our parochial altruism expressed in our stories.

INTENTIONALITY

This view of the function of narrative cognition as being useful for socialization is also consistent with one of the most recent hypothesis about the evolutionary origin of our language capacity. As mentioned in the first chapter (Language as a natural phenomenon), in a series of behavioral experiments performed in Leipzig that compared human infants with non-human primates, Michael Tomassello showed that humans have a much more developed capacity to *mindread* (i.e. to formulate hypotheses about the mental states of other individuals); and this difference in our mindreading competence results in different competences for engaging in tasks of shared intentionality –which Tomassello considered as the crucial trait that enabled us to develop language (s. Tomassello, 2010).

Humans are capable of engaging in a great range of social activities: from dancing, playing, and flirting, to making commercial interactions, contracts, debates, elections, and ideology wars, etc. All these are examples of what philosophers and psychologists call *shared intentionality* (or *we-intentionality*). Shared intentionality is not simply an aggregation of two or more subjectivities, but an emergent phenomenon that arises from the mutually-aware coordination of two or more subjectivities under a common goal (Bratman, 1992): walking in parallel on the same direction is not the same as walking *together*, for instance, which involves a degree of synchronization and mutual awareness. Shared intentionality is the form that our social interactions take, and this process is enabled by our mindreading capacities.

Linguistic communication is one of the paradigmatic cases of shared intentionality (Grice, 1975). And we can see evidence of this social nature of language in its very grammar. Our languages effectively count with specialized systems for expressing not only information about the external physical world, but also information about our interpersonal relationships, that is, about our social worlds. Among many cases of this, we can mention two examples: the markers of formal/informal speech (e.g. *Usted/tu* in Spanish; *Sie*/*du* in German), which can indicate difference of status or intimacy among speakers; and also the verbal systems of deontic modality –*may*, *could*, *should*, *must*, etc.–, which convey the possibility of an act in terms of someone's freedom, rights, and responsibilities within a social order.

[24] Sophocles, *Oedipus The King*, 429BC.

Also creating, telling, and consuming stories are activities of shared intentionality, where we are brought to imagine the minds of others (Galgut, 2010; Zunshine, 2010)). In fact, as shown by psychological experiments, the practice of consuming stories effectively increases our mindreading skills (Berns et al., 2013). And this has been considered by many scholars as a crucial evidence of the social function of narrative cognition (Zunshine, 2010).

This also opens the question of what other mental capacities the exercise of narrative cognition might improve, which introduces us to the next hypothesis.

4. ORNAMENT AND ATTRACTION: STORIES AS A MATING STRATEGY

In the last chapter we have talked about rhetoric procedures at the level of the sentence. However, there are also rhetoric procedures that are performed at the narrative level, at the level of the structure of the story. And, also in these cases, our performance goes usually way beyond a functionalist account –that is, beyond the attempt of conveying information clearly, which is the default intentionality assumed in human communications (what Grice called *informativity* –s. Grice, 1975). Just as we use ornamental rhetoric like rhymes, alliterations, and puns, in order to create aesthetic effects, at the narrative level of human literature we can also find complex rhetoric elaborations that seem to fit an ornamental function. We see this, for example, in the elaboration of symmetries in plots: e.g. Oedipus finding in himself the criminal he was tracking (Sophocle, 429BC), The predator (Ahab) dying with his pray (*Moby-Dick*) (Melville, 1851), Lovers from adversarial families (Shakespeare, c. 1599-1602). Rhetoric elaborations can often be found in the patterns of similarities and dissimilarities on the basis of which characters are designed: e.g. The intellectual Dr. Jekyll and the monstrous Mr. Hyde (Stevenson, 1886), the pious yet criminality-savant Father Brown (Chesterton, 1910-36), the twin enemy of William Wilson (Poe, 1839). And they can also be found in the use of symbolic echoes across the elements of the story: e.g. the portrait of a man that gets broken in the beginning of a story and prefigures his death in the end (Borges, 1932). All these and other kinds of rhetoric elaborations at the level of the narrative structure do not have obvious informational functions –figuratively anticipating the death of a man with the premonitory break of his portrait some pages earlier, doesn't really give the reader plot-relevant information about the event. Nevertheless, we do value, expect, seek, and repeat this kind of narrative rhetorical games. But why exactly?

A hypothesis that got popularized in the last decade is that –just as in the case of rhetoric ornaments at the sentence level– these narrative rhetorical features can also be the evidence of stories functioning as a sexually selected fitness signal. That is, as a behavioral trait selected for exhibiting cognitive fitness, which would have made our species take it as a criterion for selecting mates. Thereafter, we would have acquired an instinct to practice storytelling and a sexual attraction to storytelling performances. Let us consider the typical features of sexually selected traits, so as to evaluate to what extent this description fits the stories of human literature and our responses to them.

As indicated in the previous chapter, so that a trait can function as a fitness signal, it must be *honest*, that is, it must be truly correlated with actual fitness. And, indeed, narrative cognition is not only useful (recall all the uses stated in the aforementioned paragraphs), but people that deal more with stories effectively perform better in particular cognitive tasks, especially the ones related with mindreading, empathy, imagination, and memory (s. Abbot, 2010). Even at a neurological level, the evidence indicates that creating and consuming stories enhances connectivity in the brain (Berns et al., 2013).

At the same time, we had mentioned that sexually selected traits have also other typical characteristics. These include: an instinctive drive to perform it, a sense of wasteful virtuosity, an attraction to it, a quest for status, and an observable manifestation of it –especially in courtship rituals (s. Miller, 2010). All these things seem to be evident regarding human stories: narrative cognition functions insistently and automatically (as an instinctive drive), ornamental rhetorical displays are very frequent in narrative literature (wasteful virtuosity), humans feel certainly appealed by good stories and good storytellers (attraction), in reason of which being a skillful narrator effectively functions as a marker of social distinction among not only writers but also politicians, journalists, teachers, lawyers, priests, artists, etc. (status). Finally, storytelling is effectively suited for being exhibited (observable): in fact, as said earlier, storytelling was for most of human history a communal oral activity, meant to be directly shared, and we still use it in this way whenever we share an anecdote in a meeting, whenever we stage stories (in theater plays, movies, or concerts), and also whenever we publish a literary work.

With regards to the manifestation of sexual traits in courtship rituals, some further observations could be made. At an intuitive level, the usage of stories for seducing is intuitively evident; in fact, many idioms (such as *chat someone up* or *sweet talker*) suggest the significant role that linguistic abilities seem to have in human courtship. Also the fact that narrative cognition involves such a high degree of mindreading makes is intuitively suitable for courtship –behavior in which others' mental states are at stake. Geoffrey Miller brought this hypothesis further, and proposed that it would then be expectable that *love* constitutes a predominant literary subject –he even called it a "natural subject" of literature (Miller, 2010: 72). That is, if our capacity to cognize narrative is fundamentally a sexually selected trait that manifests in storytelling displays, it would be expectable that the products of this practice take the natural function of the practice (courtship) as one of its central topics. And, indeed, the evidence seems to be consistent with this view. A recent example of this can be found in a study performed by empirical literary scholars Jonathan Gottschall and Marcus Norlund. They performed a large quantitative analysis of narrative literature around the world that revealed massive references to romantic love spread across highly diverse and isolated culture areas, as a true relative universal (Gottschall&Nordlund, 2006). Incidentally, this study refuted the spread social-constructivist assumption that the romantic-love discourse is an exclusively Western invention (e.g. Bloch, 1991). Even if the romantic-love literature written in Europe might have particular differential features (every literature does), the topic of romantic love is certainly not solely European. In fact, European literatures were not even close to the highest scores in the world, even though they did rank higher than the average (Gottschall&Nordlund, 2006). Love stories, and the use of stories for seduction, seems to be one of the dominant expressions of storytelling, which might tell us something about the evolutionary origins of our storytelling abilities and the particular pleasure and attraction their rhetoric elaborations produce in us.

NARRATIVE COGNITION AS A MULTI-FUNCTIONAL CAPACITY PRIOR TO LANGUAGE

There are some traits that have one clear concrete function: our eyes allow us to see, our hearts serve to pump blood, and our phobias help us avoid certain dangers. However, some other traits have several or very general functions, which makes it more difficult to account for their exact adaptive rationale. Hands, consciousness, and language are examples of this. Like them, our capacity to cognize stories might be a macro-component useful for many of the purposes stated in the abovementioned hypotheses, in an intertwined way. But, if so, which one came first, and how did the other ones emerge? The task of evolutionary psychologists would be, at this point, either to argue the priority of one of these hypotheses, or to reconstruct the –expectedly complex– evolutionary history that progressively revealed the adaptive rationale of the narrative cognition at these diverse levels.

My hypothesis, on this regard, is the following one –necessarily expressed as a story. Narrative cognition could have appeared first as an extension of our capacity of causal reasoning. Increasing our capacity to make inferences and predictions about the world, this trait would've been already useful enough to be naturally selected and even fostered as an instinct. At this stage, narrative cognition would not have even required language in order to reveal its usefulness and be selected.[25] This means that language would have not suffered modifications so as to adjust it to express narrative, but it must have been already from the beginning designed for facilitating the expression of narrative –as it was already from the beginning designed for expressing other prior mental structures, such as our innate intuitions of physics, biology, psychology, etc. (Pinker, 2007: 220-1). Still, once human languages were developed and narrative cognition got to be expressed through them, new advantages of narrative cognition beyond causal reasoning would have been revealed, such as its usefulness for safely simulating potential future life problems, for creating social identities, and even for training many other cognitive skills. Moreover, until humans acquired language, narrative cognition would not have been an observable trait, but when we did –when we could eventually *tell* stories to each other–, narrative cognition would have become observable enough to be sexually selected as a fitness signal, which would have created the conditions for the emergence of the literary art of storytelling.

It is certainly not easy to account for how our mind works when processing stories and what the evolutionary rationale might be. But the inquiry of evolutionary psychology –along with cognitive psychology and neurocognitive science– into the phenomenon of story cognition has undoubtedly opened a series of paths of research from which a novel and exciting understanding of language and literature is emerging.

---

[25] The idea that the human capacity to cognize narrative is evolutionary prior to the development of language has also been formulated by Mark Turner (1998) and Ellen Dissanayake (2011), among others.

CHAPTER 5

CHARACTERS

HOW WE JUDGE OTHERS THROUGH LANGUAGE

THE MINDREADING INSTINCT: YOU KNOW THAT I KNOW WHAT HE KNOWS

> He was an old man who fished alone in a skiff in the Gulf Stream and he had gone eighty-four days now without taking a fish. (Hemingway, *The Old Man and The Sea*, 1952)

As we have been mentioning along this book, we have a cognitive faculty called *mindreading*. This faculty is what allows us to make sense of others: to imagine their mental states and, thereafter, to form an opinion of their personalities, attitudes, intentions, values, etc. And we use this capacity in every communicative interaction. Reading also requires mindreading: When cognizing written language, we also assume (as when we hear someone speaking) that the words we encounter express the intention of a thinking and sentient author (called in literary theory *implicit author*), and we also assume that that expression has been addressed to a certain thinking and sentient recipient (*implicit reader*) –even if this reader is non-explicit, abstract, or fictitious. Every act of reading requires us to create a mental representation of the minds of these hypothetical others. In the last decades, this model has started to be applied to the study of literature (s. Zunshine, 2006). And what it was discovered is that literature is one of the fields in which our mindreading capacity works most intensively. Let us see how, why, and what effects this produces.

The aforementioned opening line of *The Old Man and the Sea*, simple as it looks, demands from us a third-level of mindreading. So as to make sense of that line, our minds have to picture at least three other minds: that of the implicit author, that of the implicit reader, and that of the character of the story (the referred *He*). Literature regularly works on this third-level of mindreading, making us process a network of hypotheses about the mental states of several entities at the same time, with their complex relationships (in these case, with three nodes).

If we are trained readers, we might also be capable of making a further distinction: to cognize the mind of the author and the mind of the narrator as different representations. We are, indeed, capable of mindreading at much higher levels. For instance, each time an extra character is added to the equation, our mindreading-network naturally grows one node:

> Sing, O goddess, the anger of Achilles… (Homer, *The Iliad*, c. 1260BC-1180BC)

The classic opening line of *The Iliad* requires a fourth-level of mindreading: making us cognize at least four mental states at the same time: the author’s, the reader’s, the goddess’, and Achilles’s.

How far can we go? Here is limit example:

> Many years later, as he faced the firing squad, Colonel Aureliano Buendía was to remember that distant afternoon when his father took him to discover ice. (García Márquez, *A Hundred Years of Solitude*, 1967)

The beginning of *A Hundred Years of Solitude* requires us to think of the minds of the author, the reader, the Colonel Buendía, the Colonel's father, and also of the whole firing squad –as numerous as we want to imagine it.

And even when no characters are mentioned at all, literature still recruits our mindreading skills, since we still must imagine the minds of the implicit author and the implicit reader so as to make sense of the text, as in the opening of the classic Japanese work *The Pillow Book*:

> In spring, the dawn — when the slowly paling mountain rim is tinged with red, and wisps of faintly crimson-purple cloud float in the sky… (Shōnagon, *The Pillow Book*, 990-1000).

As cognitive literary scholar Lisa Zunshine explains in her book *Why We Read Fiction*:

> Fiction exploits the fact that, on some level, we do not differentiate between mental states of real people and of fictitious entities: we interpret the behavior of fictional characters (and their creators) using the same cognitive adaptations for mindreading that use when we interpret (and misinterpret) the behavior of people around us. Fiction builds on and experiments with our theory of mind, intensifying and transforming mindreading patterns present in our daily social life. (Zunshine, 2006).

Through neuroimaging techniques, it has been discovered that an area that is always active when we perceive or think of an individual is the medial prefrontal cortex (mPFC) (Mahy et al., 2014). It gets activated every time we try to guess the *intentions* of others, their *beliefs*, and even when we engage in *joint attention* (Call&Tomasello, 1998; Wimmer&Perner, 1983; Baron-Cohen, 1991); that is, in every social interaction (Tomasello, 2010). And the mPFC does not get just any kind of activation when thinking of any kind of person. But, actually, it seems that each person we meet coins a particular pattern in our mPFCs.

A study performed by Nathan Spreng and his team proved exactly this last point (Spreng et al., 2014). They recruited 19 people and described to them the profiles of 4 different characters. Each character was described as having a different and distinct personality – namely, in terms of agreeableness and extraversion. The brain activations of the participants as they thought of each of the characters were measured with a brain scan (fMRI). And the results showed that each character produced a singular pattern in the mPFC of each participant. These patterns were so distinct that, whenever each participant thought of one of the characters, a computer could predict of which character the patient was thinking, only by checking the patterns in the fMRI.

The results of this experiment suggested that our brains have distinct regions for coding personality traits, that we integrated this information in the mPFC, and that we use this "personality model" to predict the behavior of others in novel situations (Ibid.).

The mPFC is also the region of the brain responsible for planning and decision-making. This brought some neurocognitive scientists to think of our mindreading capacity as a particular instantiation of our predicting coding (the constant unconscious hypotheses that our minds perform when interacting with the world –s. Howhy, 2013), but oriented towards other individuals.

Moreover, our mindreading capacity is considered to be universal and innate. Some of the firsts experiments that measured it were focused on what is called *false belief*. The most famous one is called Sally-Anne's Test (Wimmer&Perner, 1983). A participant is told the following story:

> Sally and Anne are in a room. They have a basket, a box, and a marble. Sally puts the marble into the basket and leaves the room. While Sally is out, Anna takes the marble out of the basket and puts it into the box. When Sally comes back: Where will she look for the marble?

We intuitively know that Sally would look for the marble in the basket, where she put it. And we know this while knowing, simultaneously, that that is not where the marble actually is (Anna put it in the box, instead). This means that our minds are capable of holding a notion of *what is the case* (the marble is in the box) and of *what people think is the case*. And we can also assess if they are right or wrong. If they hold a true or a false believe. And all this complex reasoning is already performed by the minds of children, who proved to be capable of passing the Sally-Anne's Test since 2 to 3 years old (the researchers displayed the story to children using dolls: Baron-Cohen et al., 1985).

Other experiments have shown different consistent stages in the development of the mindreading capacity in children. Our mindreading capacity has been tested not only in the recognition of false beliefs, but also in terms of our capacity for joint attention (looking together at the same point) (Baron-Cohen, 1991) and to predict others' intentions (Call&Tomasello, 1998) – indeed, *intentionality* had already been considered by philosophers of mind as a fundamental feature of mental states (s. Searle, 1983; Dennet, 2017).

Now, our mindreading capacity does not only allow us to recognize whether people are people but, when doing this, we also judge them: we make up an impression of others that is pertinent to ourselves and we respond thereafter consequently. And, as we have seen in previous chapters, this is the endowment that allows us to build up our social reality and to cope with it. Let us consider now some of the criteria our minds use for judging others, how that affects our responses and behaviors towards them, and how that applies to literary characters.

WARMTH AND COMPETENCE

What criteria do we use when judging others? How is the mental software that allows us to do it? Social-cognition scientist Susan Fiske and her team asked these very questions at its most fundamental level (s. Fiske et al., 2002). They reasoned that, if the perception of others' mind is manifested by a recognition of their *intentions*, then it would make sense that our mind be programmed for further inquire: *Which are the intention of this individual?* And, secondly: *Is this individual capable of enacting his or her intentions?* Fiske called these criteria: judgment of *warmth* and judgment of *competence*, respectively. And they verified them empirically.

The experiments and surveys carried out by Fiske and her team showed that people make universal. Whenever we encounter others, all us perform a warmth-judgment (*Are they friends or foes?*) in a fraction of a second, and a competence-judgment (*What can they do about it?*) a fraction later (s. Willis&Todorov, 2006).

This also makes sense from an evolutionary point of view. If it is useful to evaluate on the spot whether the individuals we face are allies or enemies, it would be useful to further ask ourselves what power thse individuals have to enact their intentions (v.g. if it's an angry lion, you better flee; but if it's an angry 3-year old child, you don't need to). In this sense, these instinctive categorizations would be useful for assessing potential risk as well as potential benefits. Being able to perform these categorizations is also clearly advantageous for social behavior. Having the instinct to judge people according to this criteria –as part of our intuitive systems– allows us to sort who is part of our group and who is not, who shares our goals and who doesn't, and towards whom should we direct our cooperation or our competition (Haidt, 2012).

STEREOTYPE CONTENT MODEL

On the basis of the distinction between warmth and competence (W&C), Fiske and her team inferred a schema of four possible complex spaces that would represent typified stereotypes in accordance to which we would organize our judgment of individuals (Fiske, et al., 2002):

[+Warm] [+Competent]: Individuals perceived as being *good* and *smart*

[-Warm] [-Competent]: Individuals perceived as being *bad* and *stupid*

[+Warm] [-Competent]: Individuals perceived as being *good* and *stupid*

[-Warm] [+Competent]: Individuals perceived as being *bad* and *smart*

Each space in this schema represents how people perceive different social groups of their society. If the schema has a psychological reality, that is, if categorizations of W&C are actually structure our judgment of other individuals, then they should predict and explain structural differences across stereotypes. Fiske et al. called this the Stereotype Content Model (SCM)

Fiske and her team did many polls across USA, where people were asked to name the stereotypic groups of their society ("As considered by your society, not necessarily by yourself") (Ibid.). Afterwards, the most frequent stereotypes were included in short stories, and the participants had to qualify these stereotypical characters in different contexts in terms correlated with warmth and competence (such as moral quality and intelligence). When the results where statistically analyzed, they showed that people systematically placed the same groups in the same spaces of the Stereotype Content Model. The four spaces formed clearly defined clusters, which means that the SCM probably maps a real mental pattern.

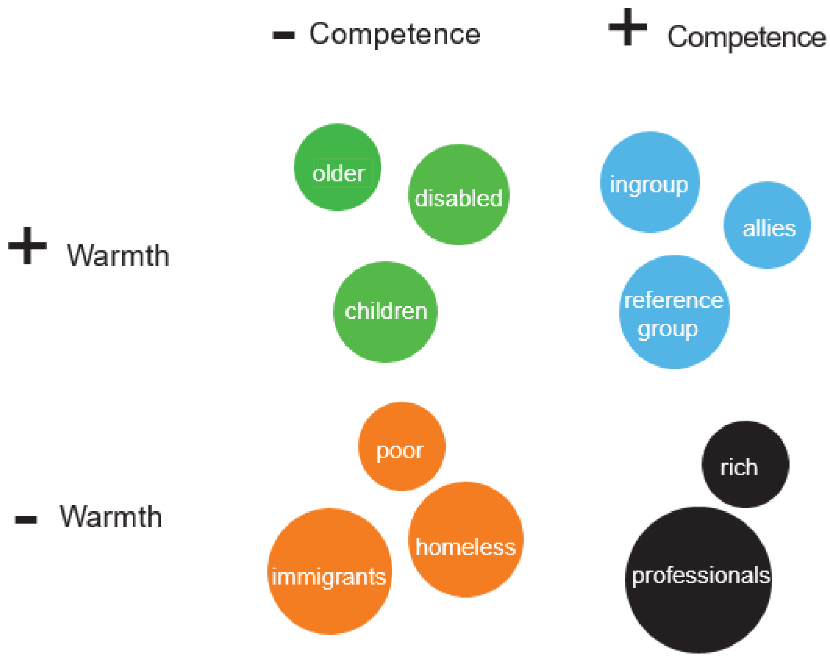


Figure 1. Map of perceived Warmth & Competence (Fiske et al., 2002)

The results of the first two groups where the following:

+W, +C: In-groups, allies, reference groups.

-W, -C: Poor, homeless, undesired immigrants.

These polar spaces (++ and --) showed results that match those of previous models of social psychology, which typically analyze social stereotypes on the basis of binary values: positive and negative perception. What those models couldn't account for is the other two ambivalent spaces, the spaces with mixed evaluations, which are a bit more complex:

+W,-C: Older, disabled (physical/mental).

-W, +C: Rich, professionals.

STEREOTYPES PREDICT EMOTIONS

After verifying that the Stereotype Content Model was actually explicative of the social structure of stereotypes, Fiske et al. decided to test if it also predicted emotions (2007). And they discovered that it does.

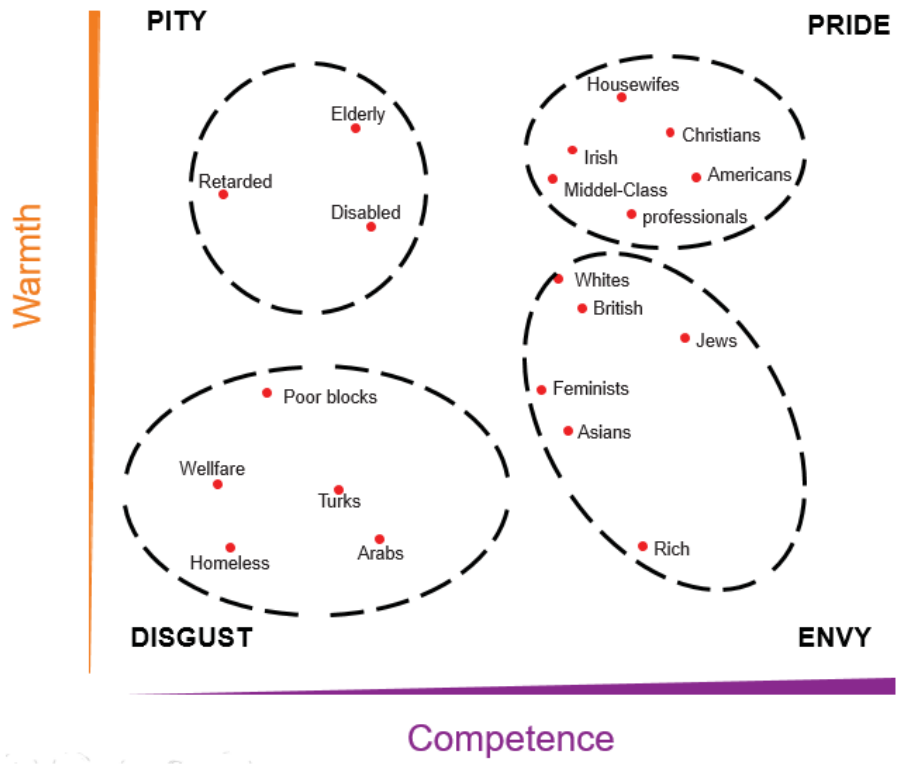


Figure 2. Stereotype Content Model (stereotypes predict emotions) (Fiske et al., 2007)

Once again, the polar spaces are more expectable. In the +W+C group, American test subjects included concretely housewives, Christians, Americans, Irish, middle-class, and professionals. And when presented with images of these stereotypes and asked about their emotions, test subjects reported experiencing pride and admiration. Whereas, in the –W–C group, Americans included "poor blocks" (again, in their own terms), wellfare, Turks, Arabs, addicts, and homeless (the most –W–C, by far). And when asked about them, people reported experiencing disgust and contempt.

But what do these emotions really mean? Further experiments provided a clue on the nature of these emotions. When confronted with the photos of a –W–C stereotype, like drug addicts injecting or a homeless begging, patients presented hardly or even no activation of the medial prefrontal cortex. Considering that the mPFC is the part of the brain that we use for categorizing an individual as such, these results seem like a good candidate for a neural correlate of what we usually refer to as *dehumanization*. Questionnaires confirmed this. Participants were furthered asked questions intended to foster them to empathize with the individuals in the photos, such as: "What do you think this person is thinking?," or "How must a day in his life be like?" And the answers showed precisely what one would expect in cases of dehumanization: "I cannot get into his head," "I cannot even start to imagine what it would be like," "I don't know what he does," "I don't want to think about it," "it's gross, disgusting," etc. And if they attempted a description of a day in their lives, the result was much more superficial and with much less mental content than when referring to other groups: participants would focus in describing only external descriptions of movements and actions, more than mental states.

The mixed groups showed, expectedly, mixed emotions. In the +W-C stereotype, American test subjects included elderly, retarded, disabled, and children. And, when thinking about them, people reported feeling sympathy and pity. Pity is indeed a mixed emotion: it may feel good for the provider, but it doesn't feel good for the receiver. People in this group were basically perceived as being well-intentioned, but at the same time as not capable of carrying out whatever intentions they might have.

The –W+C stereotype included: whites (though quite high on warmth), British, Jews (second most competent after Americans), feminists (the less competent of this group), Asians, and rich (the coldest of all, at the same level of homeless). With regards to this stereotype, people reported experiencing envy, which is also mixed: it feels bad for the provider but it may feel good for the receiver. But, with regards to this group, test subjects also systematically reported a second mixed emotion, *Schadenfreude*: the feeling of joy about someone-else's suffering. That is: In relation to cold-competent stereotypes we become somewhat sadistic. And a very interesting experiment confirmed this. Patients were wired with sensors around their mouths so as to track the electric activation of their smiling muscles (zygomaticus muscles). When showed with images of people having a positive experience (like somebody celebrating a birthday), patients usually react by activating smiling muscles, and when they see a person having a bad experience, smiling muscles don't get activated. The same pattern occurs with all the stereotypes: elderly, women, homeless. With all groups, except one: the cold and competent. With regards to this envied group, the exact opposite occurs: test subjects smile when they see an portentously rich young man, for example, stepping into chewing-gum or falling into water and they do not when they see this man celebrating with a trophy in his hands. We enjoy the tragedies and dislike the victories of people that we perceive as cold and competent.

EMOTIONS PREDICT BEHAVIOR

The same logic was further applied in behavioral experiments to test if a person categorizing somebody in accordance to a certain stereotype, and experiencing a certain emotion, was a good basis to predict how the person will conduct herself. And, once again, experiments verified this hypothesis. The results where integrated into the so-called BIAS map (Behavior from Intergroup Affect and Stereotypes) (s. Fiske et al., 2007), which is a model to map the "attitude schema:" the relationship between social stereotypes, emotions, and behavior.

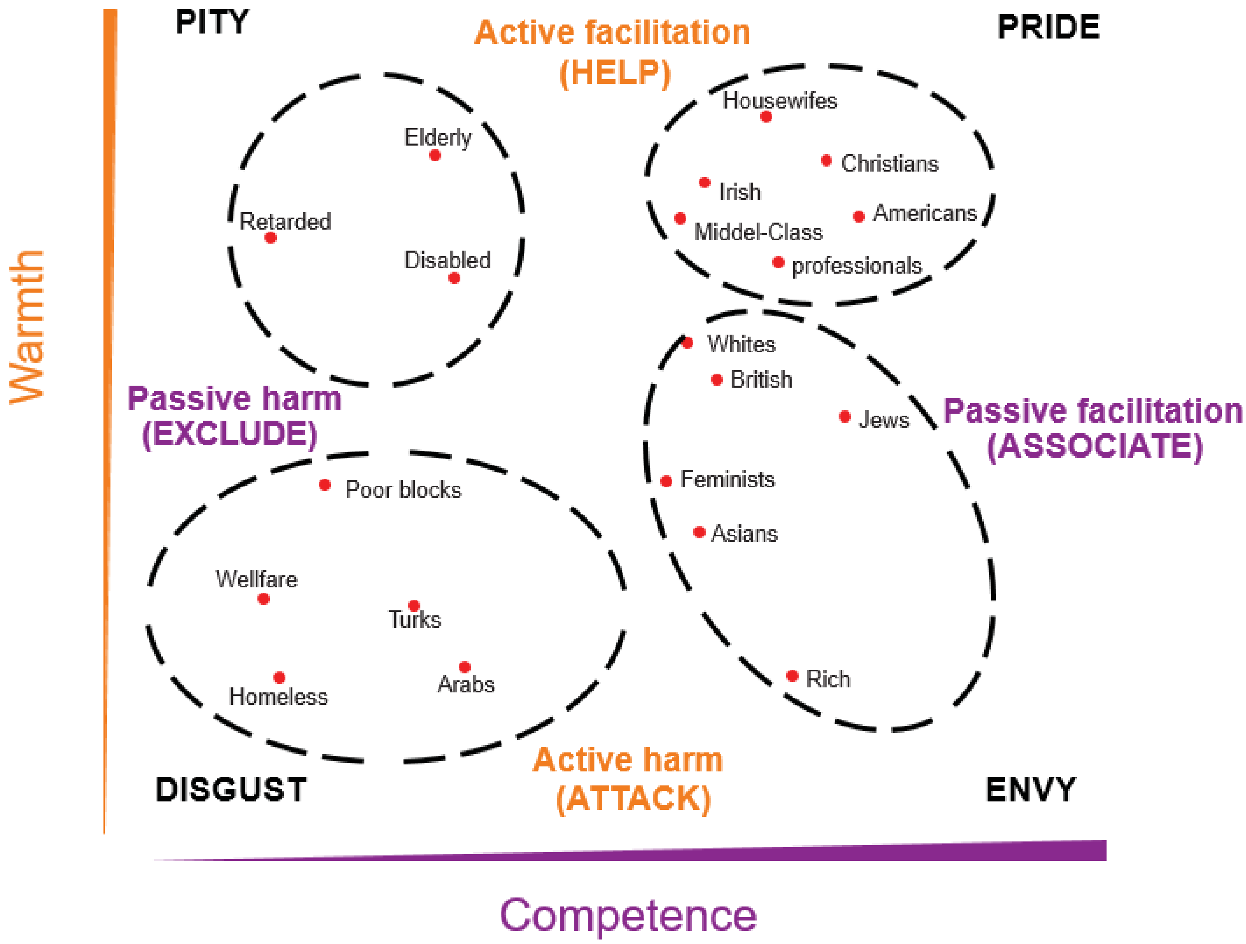


Figure 3. Behavior from Intergroup Affect and Stereotypes (BIAS Map) (Emotions predict behavior) (Fiske et al., 2007)

What the results show is that the perception of warmth determines our *active* behavior: towards individuals perceived as warm we manifest an attitude of active facilitation (help, protect, etc.) and towards cold people we assume attitudes of active harm (attack, fight, harass, etc.). The perception of competence determines instead our *passive* behavior: towards competent people we manifest passive facilitation (associate, go along) and towards incompetent people, passive harm (exclude, demean).

This basically means that people help and go along with their friends. Homeless are either ignored or, when noticed, attacked. Elderly people are protected but excluded (v.g. from driving, voting, partying, technology, fashion, sex, etc.). And with regards to the envied ones, people basically go along with them when they need them but attack them when they do not. This last attitude has been frequently registered, for example, in many cases of Asian students in American universities, and Asian employees in American companies. And it is also worthy to mention that this particular group (not Asians in particular, but -W+C in general) has often in history been object of genocides. Genocides have actually been more frequently directed against groups that elicited *Schadenfreude* and envy (-W+C) in the aggressors, than against groups that elicited disgust (-W-C).

WHAT CUES DO WE USE FOR JUDGING PEOPLE?

What we have until now is that stereotypes predict emotions and this predicts behavior. But how can we predict stereotypes in the first place? How can we know which people is likely to be categorized in terms of W&C? A quite large record of experiments has shown several physical and social cues to be very effective predictors of stereotypes.

The first predictor of stereotypical categorization is precisely temperature. The use of the term "warmth" for the judgment of intentionality is not arbitrary. It responds to a conceptual metaphor that has proven to be universal in our species: *Affection is heat* (S. Lakoff and Johnson). According to this, the idea that somebody can be "a warm-hearted woman" or a "cold-blooded man" would not only be a manner of speech, but primarily a manner of thought, where our sensorial experience of temperature would function as a source domain by means of which our minds conceptualize the target domain of affection. These cases are called *conceptual metaphors* (Lakoff & Johnson, 1980). Now, if this model describes correctly an actual mental structure, then not only our language but also our behavior should evidence correlations between physical warmth and interpersonal warmth. And, indeed, this is what behavioral experiments have consistently found. The correlations between warmth and affection do not only manifest in language, but in people's self-perception: for example, in warmer days, people perceive themselves as more agreeable and sociable (Fetterman et al. 2017). It also manifests in people's judgments: Asked to describe imaginary characters, people do it in more personal and favorable manners when they have been previously given hot coffees, but they do it in the opposite manner if the coffees were cold (Williams & Bargh, 2008). And asked to describe a moment from their own past, people recall memories of inclusion when the room is warmer and of exclusion when it is colder, as if their moods could be guided by the knob of the thermostat (Zhong & Leonardelli, 2008). Recently, neurologists discovered the neural circuitry that produces this phenomenon: People indeed use the brain areas that process temperature perception (in the somatosensory cortex) when cognizing affection (Lakoff, 2014). Claims of the possible universality of this metaphor in humans has made psychologists propose hypotheses to explain it in evolutionary terms: as an adaptive associative mapping generated by the embodied experiences of human evolution –such as newborns being held in their mothers' arms (Hukkinen, 2012). Evolutionary hypotheses about the human mind, like this one, are supported by findings in comparative psychology (across species): In recent years, evidence was found of metaphorical thinking in our closest relatives –e.g. chimpanzees cognize social rank in terms of spatial position (high-low) (Dahl, 2013). The metaphor *affection is warmth*, in particular, has not been extensively tested in non-human primates, but some behavioral evidence suggests its plausibility: an early experiment performed by Harry Harlow (in 1958, prior to the theory of conceptual metaphors) showed that, when offered with artificial (i.e. man-crafted) surrogate mothers –in normal ambient temperatures–, infant monkeys preferred the warm ones rather than the cold ones, even though the cold ones provided them with food.

At the same time, certain face traits also proved to be predictive of how a person is likely to be categorized in terms of W&C. People with slightly surprised, happy faces, and baby-faced people are perceived as warm. People frowning, with angry faces, are perceived as cold. People with strong, dominant faces are perceived as competent. And people who look weak, sick, and submissive are perceived as incompetent.

Finally, many bodily traits are particularly predictive of judgments of competence. Taller candidates routinely win presidential elections. People who are literally higher up seem like leaders. People who take up space dominate people who take up less space, as do people making fists. Standing strong –"power posing"– actually raises testosterone, lowers the stress hormone cortisol, and emboldens decision making. All these physical stances have effectively been proven to convey dominance, status, and competence to others. (Fiske & Malone, 2013: 24 and Fiske et al., 2011).

At the same time, Fiske and her team showed that judgments of W&C can also be predicted with relatively high accuracy by social structures (Fiske, et al., 2002). As suggested above, warmth is predictable by *goal interference*. People obstructing our goals, indifferent to our goals, or contributing to our goals are systematically perceived with distinguishably different degrees of warmth. This has to do with the fact that we ground our judgments of warmth on conjectures about intentions. And it explains why in-group people are typically perceived as warm (they are part of us, and therefore share our goals). Competence, on the other hand, is predictable by *status*. That is, people with high status (whatever might be considered to be a provider of a status in any given society: having more money, a better job, better looks, etc.) are systematically perceived as competent, and people with low status as incompetent. This shows that –even if undeclared– most people believe their societies to be ultimately meritocratic –despite the overwhelming evidence that exists against it. However, differently from warmth, status seems to functions relationally: people with high status are perceived as having more competence by people that have lower status in relation to them.

This sensibility to competition and status seems to be quite primitive and it even takes place among most other social animal species. A simple experiment proved this with regards to dogs. Two dogs were kept together, separated from each other only by a grill. In a first trial, they were both fed with high quality pellets for some days; and, when the dogs were used to it, their diet was changed and they started to receive low quality pellets. The dogs stopped eating for some time, but they didn't revolt and soon ended up accepting the low quality pellets. In a second trial, another couple of dogs were also fed with quality pellets in the beginning, but then only the diet of one of them was changed for low quality pellets, whereas the other dog continued to receive the better food. The dog that received the low quality pellets didn't adapt passively to the diet, but showed instead aggressive behavior: it would turn around its plate, bark to the other dog and to the assistant, etc. It seems that it is common among social species to have some kind of awareness of packing orders, of who is getting more and who is getting less, and even a feeling of deservedness, which among humans is called "social justice."

OVERAL PREDICTIVE MODEL

The bottom line of all this is that the way we categorize other individuals, despite being extremely varied, is certainly not random. Of course, there are countless ways of judging people, but all these ways seem to respond to constant principles that can be recognized in relation to typified stereotypes, emotions, and behaviors.

The overall predictive model that integrates Fiske's studies has the following form: Social structure (competition and status) and bodily cues (temperature, face, posture, etc.) predict stereotypes (warmth and competence); stereotypes predict emotions (disgust, pity, envy, pride); emotions predict behavior (active, passive, help, harm); and behavior, inasmuch as it is actualized in social roles and bodily manifestations, can in its turn predict also stereotypes.

Figure 4: Overall predictive model

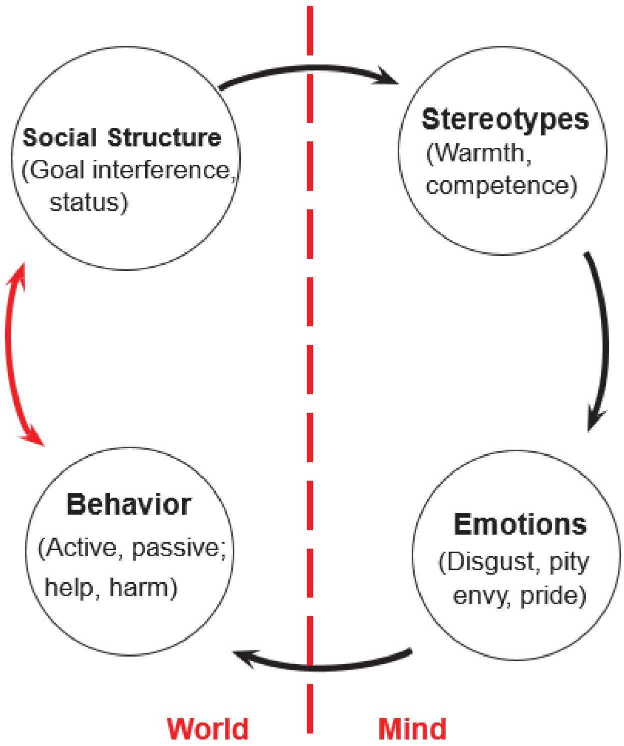


Figure 4. Overall predictive model. (Fiske et al., 2007)

By some measures, the judgments of W&C influence more than 80% of all human social behavior (Fiske, 2013: 22). We use W&C to assess not only people, but everything in our lives that acts or seems to act of its own free will: people, groups of people, pets, animal species, teams, companies, brands, and nations. (Ibid.). The predictive power of this model has been tested, for example, regarding political candidates. Research participants shown with photos of unfamiliar out-of-state political candidates were able to pick up the winners on the basis of assumed competence and warmth two-thirds of the time. (Todorov, et al., 2005).

This model has been applied to several different societies and subgroups (more than 30 researches in several countries), and its explicative and predictive power seems to function everywhere. In a superficial sense, the results are of course varied, because each society sets its own stereotypes: French and English may have different stereotypes of Germans; the same stereotype can be considered warm for a group but cold for another; etc. But the point is that all the societies analyzed through the lens of W&C, show that, whatever stereotypes people might report, these stereotypes never appeared randomly distributed, but they always ended up clearly grouped in distinct clusters from which you can quite accurately infer who likes whom and who hates whom.

For the moment, this model has been proven universal regarding geographical contexts and to each country’s own groups (Cuddy, et al., 2009; Durante et al., 2012), to different historical moments (Italian fascists: Durante, Volpato, & Fiske, 2008; American students: Bergsieker, 2012), to different subgroups (subtypes of women and men: Eckes, 2002; immigrants: Lee & Fiske, 2006; gay men: Claussel & Fiske, 2005; African Americans: Fiske et al., 2009); to different entities (brands: Fiske & Malone, 2002); and even regarding our perception of other animal species.

JUDGING LITERARY CHARACTERS

Many of the abovementioned experiments suggest that judgments of W&C occur independently of whether the observed individual is real of fictional. Still, a specific experiment about our perception of popular fictional characters, namely from literature, could be carried out. And, if the W&C model is correct, its predictions should be also verified in literary cases. Let's make a simulation of such a case.

First, the results shouldn't show characters randomly distributed across the W&C graph, but they should show the characters grouped in distinct clusters that correspond to the four spaces of W&C. For instance, let us imagine the results that could be expected from this model in relation to J.R.R. Tolkien's *The Lord of the Rings* (1954-55) (henceforth, *LotR*):

+W+C: Frodo, Gandalf, Elrond, Aragorn
-W-C: Gollum, Orcs
+W-C: Sam, Merry, Pippin
-W+C: Sauron, Saruman

Secondly, these stereotypes should predict emotions of the readers:

+W+C: Towards Frodo et al., people tend to feel pride and admiration.
-W-C: Towards Gollum et al., disgust and contempt.
+W-C: Towards Sam et al., sympathy and pity.
-W+C: Towards Sauron et al, envy and Schadenfreude.

Thirdly, these emotions should predict attitudes (potential behavior) of the reader:

+W+C: We welcome the help to and association with Frodo et al.
-W-C: Reject and attack Gollum et al.
+W-C: Protect Sam et al., but restrain their vote in important decisions.
-W+C: Attack Sauron but would accept to try to go along with him in a situation of extreme danger and necessity.

Fourthly, bodily signals of the characters and social-structure cues of their world should predict the stereotype categorization made by people:

+W+C: Gandalf contributes to the mission of the protagonist (destroy the ring), which predicts perception of warmth, and he is powerful and tall, which predicts competence.

-W-C: Gollum lives naked in a dark cave, which predicts lack of warmth, and he is short, weak, and stopped over, which predicts incompetence.

+W-C: Sam is friends of the protagonists (warmth), but he is also shorter and chubbier (incompetence).

-W+C: Sauron lives in a dark cold place (cold), but is very tall and powerful (competence).

These responses to the characters of *LotR* are hypothetical. But the point is that, having a concrete model (such as the one of W&C) so-formulated, allow us to test it in novel cases, so as to discover the extent of its predictive and explanatory power in relation to the cultural phenomena we are interested in describing.

In the case study of the third part of this book, I have applied the model of W&C for predicting how readers would judge the characters of a text, and what emotions they would feel towards these characters. And the predictions of this model proved to be highly accurate also in these literary cases. Increasing evidence suggests that we judge characters with the socio-cognitive systems that we use for judging people. And the systems we use for judging people are increasingly showing consistent biased that can be measured, described, explained, and predicted.

Now, there is a cue that we systematically take into account for judging people (to form a stereotype of them) that is crucial for literature, and that should be considered in detail: the voice.

THE PERSONALITY BEHIND THE VOICE

As we have seen in previous chapters, organisms do not only evolve to cope with their natural environment (natural selection) but also to cope with other organisms (sexual selection). For that reason, organism develop systems for sorting other organisms and guiding their attitudes towards them. Some of these mechanisms have to do with the voice. For instance, male mice seeking to impress mates sing unique high-pitched songs, vocalizing in the ultrasonic range. They produce these whistling sounds by creating a type of feedback loop of airflow in the windpipe and larynx. And female mice are picky about which songs they like: they prefer tunes that differ from those sung by their own relatives –which expectedly helps them to avoid incest and thus increase the genetic diversity of the offspring, which is biologically advantageous. (s. Pinker, 2007). As it can be observed, the voice can be useful for signaling characterizes of the organism that are evolutionary pertinent –such as recognizing kin.

Voice is also highly pertinent in humans. Not only as a means for communicating ideas, but also as a system for signaling –already in the physical dimension of the sound– other characteristics of our personality. Our minds are designed to pick up a great range of complex information from vocal cues: from a speaker’s tone, pitch, cadence, volume, etc. And the way we do it is, in many ways, instinctive (automatic and effortless), befitting to the interests of our species, and predictably biased. (S. Mladinow, 2012).

An experiment by evolutionary psychologist David Andrew Puts and his team showed that men adjust the pitch of their voices higher or lower in accordance to their assessment of where they stand on the dominance hierarchy (Puts et al., 2006). The experiment placed participants in blind rooms –they could only hear each other. More than 200 men partook a fake contest for a date with an attractive woman, in which each man had to orally state the reasons for which he should be admired by others. Later, the participants had to fill a questionnaire about themselves (so that, afterwards, their self-perception could be compared). The oral statements of the participants were recorded and then the fluctuation in the sound properties of their voices was analyzed. The researchers found that men systematically lowered their pitch whenever they perceived themselves as physically dominant with regards to a competitor, and they raised their pitch when they believed they were less dominant. And they did not seem to realize that they were doing this.

This discovery is complementary with discoveries that have been done with regards to female voices and judgments. Empirical evidence has shown that women consistently rate men with lower voices as more attractive (Collins, 2000). And this preference is even more pronounced in their fertile phase of their ovulatory cycles (Puts, 2005). Moreover, also women’s voices consistently vary with the phases of their reproductive cycles –in both pitch and smoothness–, and the greater a woman's risk of conception the sexier men find her voice (Nathal Pepitone et al., 2008).

We judge people by their voices and we seem to do it in an intuitive manner. A way to prove this is by showing how predictable the biases of our voice-judgments are and how easily they can be tricked. At the Stanford University, communications professor Clifford Nass ran a series of experiments that had participants sitting in front of computers that would talk to them with prerecorded voices (Nass et al., 1994, 1997; and Nass&Lee, 2000). All the participants knew the voice did not belong to a human but to a computer and all the computers would provide the exact same information about a series of subjects. However, some computers would talk with a female voice and others with a male voice.

In parallel, the scientists counted with a statistical research of gender stereotypes correlated with different topics. For example, the stats showed that people tend to think that women have more knowledge about relationships and intimacy. And, as a consequence of this stereotype, if a man and a woman show the same knowledge in this domain, the woman is still more likely to be perceived as more competent in it.

When the results of the computer-interaction experiment were coupled with the gender-stereotypes survey, they discovered that the participants consistently projected all their gender-stereotypes to the computers. Even if all the computers provided the same information, when the information was about gender-biased topics (such as "love and relationships"), people judged differently the information provided by the computers, biased by the gender-cues of the voice: the participants judged the information about relationships as more qualitative and sensible than when it came from a female computer-voice than when it came from a female computer-voice, even though all the computers provided the exact same information. Consistently with this, when the topic in question was gender-neutral (the one tested was "mass media"), the participants gave similar scores to the computer voices, independently of the gender-cue. This shows that we are biased by our very intuitions to pick up personality cues from voices. All the participants were rationally aware that they were hearing a computer, not a real person. Nevertheless, that did not prevent their minds from making intuitive personality judgments –with their predictable biases– of the voices they heard.

Moreover, we do not only make personality judgments from the kind of voice, but even from the kind sounds we use.

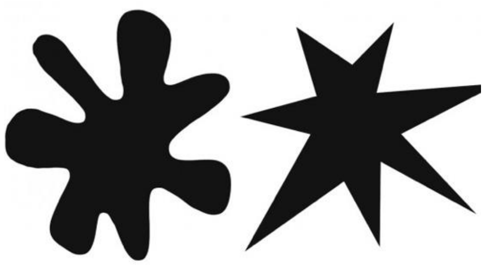


Figure 5. Shapes (Köhler, 1929)

Imagine that these two shapes are characters. One is called Kiki and one is called Bouba. Could you guess which is which?

Psychologist Wolfgang Köhler asked this same question to different groups of Spanish speakers, in an experiment in 1929. With an unexpectedly great consistency, most people judged that the round shape (left) was Bouba and that the sharp one (right) was Kiki (s. Köhler, 1929 and 1947). In 2001, psychologists Vilanyanur Ramachandran and Edward Hubbard repeated this experiment with English and Tamil speakers (in India): between 95% and 98% of both groups gave the exact same answers (s. Ramachandran&Hubbard, 2001). Ever since, the so-called *bouba/kiki effect* has been verified across several languages, it has been proven to occur already in infants as young as two and a half years old already (Maurer et al., 2006).

The bouba/kiki Effect suggests that our mental mappings between sounds and concepts might not be completely arbitrary (s. Ramachandran&Hubbard, 2001). Our minds seem to match –using mental metaphorical systems– certain sound features with certain visual features, creating an actual *synesthesia*-like mapping. These mappings are embodied: our minds link the round shape of our mouth (bou-ba) with visually round shapes, and edgy sounds (*ki-ki*) with edgy shapes. Similar effects have been discovered across different kind of linguistic sounds: we seem to be biased not only to attribute embodied qualities to consonants but also to vowels, to their quality and their quantity (long vowels evoke long objects and short vowels short objects, for instance –Bross, 2018). It seems that Shakespeare was wrong about this: a rose by any other name might *not* smell as sweet.

These biases are also connected with social judgments. People already start forming judgments of others even from the sound of their names (Sidhu&Pexman, 2015). When people read names such as "Molly," they imagine an individual with a round silhouette, and they associate this name with personality traits such as easygoingness. Whereas, when people read names such as "Kate," they imagine an individual with a sharp silhouette, and associate it with personality traits such as determination.

Vocal sounds are an ancestral code we use to communicate our personality. And in the psychological functioning of this phenomenon might lay a key to understand literary voices.

When we read, our minds form impressions and judgments of the personality that we imagine as having authored the text. This personality is communicated to us not only by the content that an author conveys, but also by its voice. As mentioned in the second chapter (Words), also our hearing systems are active when reading. And our mind also applies in these cases the cognitive criteria we use for guessing personalities behind vocal sounds. Some texts give us the impression that we are before a feminine or a masculine voice, a young or an old voice, a soft or an aggressive tone (s. Mladinow, 2012).

The intuitive systems that we use for judging people from vocal cues are highly pertinent for literature, because they model our mental representations of the personalities suggested in a text. As a corollary of this, I think that the notion of *style* in literature can be reconsidered from this cognitive perspective. In terms of literary production, the style is typically defined as the synthesis of the rhetoric and formal habits of a writer; but, in terms of reception, the style could be described as the personality that an author elicits in the readers' minds through the quality of its voice, and of the voices of his or her characters.

CHAPTER 6

EMOTIONS

HOW TO MOVE PEOPLE WITH WORDS

INDUCING CHEMICAL CHANGES IN THE READERS

We all know that stories affect us emotionally. A significant part of the value we attribute to stories is based precisely on the fact that stories can make us laugh, cry, feel horror, intrigue, excitement, despair, hope, and many other emotions. We also know that emotions are grounded in chemical changes in our bodies and brains. As physiological measuring technologies and techniques progress, it is becoming increasingly accessible to correlate these chemical changes with different kinds of stimuli in controlled experiments. Thereafter, the ways in which particular story-features manipulate our emotions is starting to be mapped at its chemical level and, thereby, integrated into neurocognitive accounts.[26]

DOPAMINE

Please, read carefully the following text:

> It was one of the outer rooms of the first floor. I stumbled on something –I think it was a footstool– and I almost went down. I banged into a table to hold myself up.
> “That’s right,” said Harry, “wake up the whole fucking household.”
> “Look,” I said, “what are we going to get here?”
> “Keep your fucking voice down!”
> “Harry, do you have to keep saying fucking?”
> “What are you, a fucking linguist? We’re here for cash and jewels.”
> I didn’t like it. It seemed like total insanity. Harry was crazy; he’d been in and out of madhouses. Between that and doing time he’d spent three-quarters of his adult life in lockup. He’d talked me into the thing. I didn’t have much resistance.
> “This damn country,” he said. “There are too many rich pricks having it too easy.” Then Harry banged into something. “Shit!” he said.
> “Hello? What is it?” We heard a man’s voice coming from upstairs.
> “We’re in trouble,” I said. I could feel the sweat dripping down from my armpits.
> “No,” said Harry, “he’s in trouble.”
> “Hello,” said the man upstairs.
> “Who’s down there?”
> “Come on,” Harry told me.
> He began walking up the stairway. I followed him. There was a hallway, and there was a light coming from one of the rooms. Harry moved quickly and silently. Then he ran into the room. I was behind him. It was a bedroom. A man and a woman were in separate beds.

[26] Disclaimer: Reading the following sections is likely to increase your levels of dopamine, cortisol, oxytocin, and endorphins.

> Harry pointed his .38 Magnum at the man and said: "All right, buddy, if you don't want your balls blown off, you will…"
> (Bukowski, "Break-In," 1979)

Let me stop the story at this point. This excerpt is the beginning of one the most intense stories by Charles Bukowski: "Break-In." Do you feel some kind of annoyance after the interruption of the last line? Do you feel an unsatisfied curiosity? If you do, this means that your brain has effectively received a rush of dopamine.

Our brain has a system for making constant predictions about the world –which has been described as *the predictive mind* (Hohwy, 2013). We insistently try to spot patterns that allow us to calculate the consequence of our actions. When our predictions fail, our brain makes an automatic correction –this mechanism is called *negative mismatching*. And, when they succeed, our limbic system releases a chemical that makes us feel good. That chemical is precisely dopamine, a neurotransmitter –which, unlike hormones, travel through the nervous system, not through the blood. Dopamine is the substance that regulates the system of rewards that moves our brain into spotting patterns, correct them, and complete them (s. Previc, 2011). It makes us feel eager and ready to receive and process pertinent information about the eliciting condition: it makes us focused, motivated, and it increases our memory. It is specially released in activities that require a sense of alert and reaction readiness (such as sports, hunting, fighting, etc.). Dopamine keeps us awake, with a feeling of control, motivation, and confidence; it is therefore also related with self-esteem –depressed people show lower levels of dopamine– and also with addictions –It is precisely dopamine what cocaine elicits in our brain.

Dopamine has also been found to be elicited by stories whenever they suggest a pattern to be discovered (such as detective and mystery stories), in the *aha!*-moments where plot-pertinent information is revealed, in turning-points (*Wendepunkte*) of stories, and specially in suspenseful stories (s. Lehne, 2015). Let us take another look at Bukowski's excerpt. The sense of danger depicted in the scene, the intrigue about what the characters are doing there, the tension between the hidden attackers and the ignorant attacked, the fear of being discovered, the surprise of the characters, the rapid and violent development of events, all that contributes to make our brains become progressively alert and motivated, which concretely means an increase in our levels of dopamine. All the features that make a story more dramatic are dopamine-booster rhetorical structures.

An experiment was performed by Moritz Lehne at the Freie Universität Berlin in 2015 for measuring neurological traces of story-induced suspense. In it, a group of people was exposed to ETA Hoffmann's short story "Sandman." This story that has been traditionally been regarded as one of the best examples of the suspenseful style that was characteristic of Hoffmann's literature (s. Krüger, 2010). But, does suspense produce a recognizable neural trace? In the experiment, the participants had to read the story segment by segment. And they had to rate each segment in terms of the perceived suspense. While doing so, their neural activations were measured with fMRI. What Lehne found was that the most suspenseful segments where associated with activation of the medial and dorsolateral prefrontal cortex, in particular the inferior frontal sulcus, the inferior frontal gyrus, and the precentral gyrus (lateral premotor cortex), as well as posterior temporal areas extending into the TPJ. These brain regions are precisely the ones responsible for social cognition and predictive inferences. This suggest that when reading a suspenseful story we are trying to make as many predictions as possible: about the situation and about the mental states of the characters (What is their intent? What is the risk? What will happen next?). We become alert and eager of information, as if the events depicted were pertinent for ourselves.

Now, if suspenseful stories boost our dopamine levels, which makes us so eager for pertinent information, it would be expectable that, whenever the flux of information be abruptly interrupted –as in my cut of Bukowski's story–, readers will feel some form of displeasure. Indeed, this may be related with the role of dopamine in addictions. Inducing a rush of dopamine in a reader, by engaging him or her in spotting a pattern, and then omitting the necessary information to fill it, would be a direct way of making the reader feel a sudden dopamine abstinence. Interrupting the dopamine stimulus in its peak would bring readers to feel like story-addicts in need of an extra dose. This is what makes cliffhangers so useful –as frequently seen in serial novels and TV shows– to keep the audience desirous of following installments. And this can also explain why closed endings are felt as more comforting (our limbic systems reward pattern closure releasing more dopamine) and, in consequence, are massively more frequent than open endings, which cut the dopamine flux, leaving us experiencing a form of unsatisfied curiosity.

OXYTOCIN

The Yaghans are an indigenous people in the southernmost area of the Southern Cone: the last extreme of the Chilean and Argentine Patagonia. The end of the world. They were nomad hunter-gatherers who counted among the first historical settlers of the region: their ancestors lived there for more than 8,000 years, spread across the islands of the archipelago of Tierra del Fuego (Chapman et al., 1995). A large percentage of them died when the Europeans arrived to the Americas, due to new diseases that came with them, for which the indigenous people lacked antibodies. They used to cover their bodies with seal fat so as to stay dry and warm when fishing in the icy rivers of the region. Their nakedness was perceived as indecent by the colonizers, so they forced the Yaghans to wear clothing. But clothing gets wet. As a result, many died of hypothermia. Many other died in slavery. Their population decreased progressively in the last centuries. The last available census (from 2002) registers 1,612 direct and indirect descendants of the Yaghan people, disperse across the two countries. But their language is virtually extinct. Until recently, there were only two Yaghan native speakers left in the world. Two sisters: Ursula and Cristina. In 2013, Ursula died. Cristina was then 84 years old. Here is a translated transcription of her voice, speaking in her own language, as registered soon after her sister passed away:

> I live in Ikika, next to the river. I did not meet my father; he died when I was very little. My mother also died, when I was five years old. I grew up in Mejillones. I'm speaking my mother tongue. Yaghan is the first language I learnt. It is the one we spoke in my family, with my mother. We used to sail in the river, all the way down. We would go there to hunt otters. I liked it… I had to do it, because my cousin would bring me… I do not have with whom to talk Yaghan anymore. I used to have long conversations with my cousin, my sister, with my brother in law, my aunt… All of them spoke Yaghan…[27]

But Cristina is the last native Yaghan speaker. In 2009, she was recognized as Living Human Treasure by the Unesco. She has today 89 years, and when she dies, her language will die with her.[28]

---

[27] Visualarte, 2013

[28] The story described here is based on registers of the life of Cristina Calderon: Hitt, 2004; Azúa, 2011; and Infobae, 2018. Facts of the story have been modified for the purposes of this book.

Stories on the topic of social relationships, like this one, typically make us feel a mix of empathy (regarding social integration) and distress (regarding social disintegration). In this case, one is drawn to think of this woman, to pay attention to her voice, to imagine her language. Languages are social systems… What would it feel like to be the last member of a millenary social network? The story suggests a plurality of social bonds: the ancestors, the family, the friends, that converge in this individual woman. And it also suggests despair: about all the stories, all the ways of saying and thinking that are registered in a language, that abstract document of the culture of a whole people, so intimately embodied in her mind, and that will find an endpoint in her. We feel bond and grief. And these feelings have a very specific chemical ground: oxytocin and cortisol.

Neurologist Paul Zack and his team studied in the lab the effects of social-bonding stories in a series of experiments (Zack, 2012). They exposed participants to a moving story and to a control one. The moving story was about a father and his son, Ben, who died of cancer. The control story depicted Ben and his father carelessly walking in the zoo, where nothing happened. The participants expectedly reported different feelings associated to these stories: whereas the latter produced boredom and made the participants rapidly disengaged from it, the moving one made them experience feelings of empathy and distress, and the participants appeared more interested in the story. At the same time, the scientists took blood tests of the participants before and after exposing them to the stories. And they discovered strong correlations between the differences in the stories and changes in the levels of two hormones: cortisol (correlated with distress) and oxytocin (correlated with empathy).

Oxytocin is a very singular hormone. It plays a crucial role in social bonding and sexual reproduction (Yang, et al 2013). The hypothalamus in our brains generate oxytocin in many human activities, including childbirth, social recognition, pair bonding, anxiety, and orgasm (Lee et al., 2009). Other activities that have proven to rise our oxytocin levels are: hugging, massaging, partner-dancing, holding hands, and praying. Due to its crucial role in social behavior, oxytocin is popularly known as the *love hormone*, the hormone that keeps us together. It is considered to be very primitive: it is actually present, with a similar role, in all social mammals (Anaker&Beery, 2013). However, there is a particularly potent trigger of the love hormone that affects only humans: social-bonding stories, such as the stories of Ben and Cristina. When we are brought to imagine characters that experience intense personal connections, our brains get flooded with oxytocin, which makes us feel a form of attachment to individuals that may not even exist.

In a second version of their study, Zack and his team exposed the participants to the stories and scanned which brain regions got activated when cognizing them (Zack, 2012). The results showed that the story of Ben's death activated not only the brain region that produces oxytocin but also the medial prefrontal cortex, which is responsible for mindreading. Mindreading, as we've seen, is precisely our capacity to imagine others' mental states. The feeling of empathy produced by a chemical change in the brain, induced by a story of social bonding, brings us to think of what others might be thinking.

In the last version of this experiment, Paul Zak and his team measured other physiological markers of distress and empathy, beyond the blood levels of oxytocin and cortisol. They measured also things like heart-rate, skin conductance, and respiration, before and after people would read the story. And, at the same time, they tested whether it was possible to predict people's behavior with this information. So, before exposing the participants to the stories, they gave $20 to each of them. After the story, the participants had the opportunity to make a donation for a children's organization. The incredibly discovery was that the physiological measurements predicted with 80% of accuracy whether a participant was likely to donate and even how much. The increase of oxytocin produced by social-bonding stories affects not only how we feel, but also how we behave.

ENDORPHIN

> A woman gets on a bus with her baby. The bus driver says: 'Ugh, that's the ugliest baby I've ever seen!' The woman walks to the rear of the bus and sits down, fuming. She says to a man next to her: 'The driver just insulted me!' The man says: 'You go up there and tell him off. Go on, I'll hold your monkey for you.'

This joke, written by comedian Tommy Cooper, was voted in 2010 as the "best joke of all times," in a survey partaken by 36,000 participants –The voters had to chose among 50 jokes that had been selected by the researchers after scrutinizing more than 1,000 jokes in the web in terms of popularity (Hutchison, 2010).

Humor and laughter are very singular universal human traits. Humor is a deceptively simple emotion. It has been an intriguing puzzle all along history: From the ancient Greeks, who were interested in the political uses of humor, to Freud, who attempted to discover our repressed desires and fears in the kinds of things that we find funny. But, only recently, the biological nature of humor has begun to be explored in controlled experiments, especially in cognitive, neurological, and evolutionary ways. And what was discovered is that humor really *moves* us: At a chemical level, laughing makes us release endorphins, which are hormones that make us feel a particular form of pleasure, sometimes even excitement, and inhibit the communication of pain signals.

Evolutionary psychologist Robin Dunbar performed a study in 2011 so as to identify correlations between social laughter and endorphin levels. In a series of five experiments, he exposed a group of people to humorous and non-humorous videos. The humorous ones included excerpts from *The Simpsons*, *Friends*, *South Park*, and various stand-up comedy performances. The non-humorous ones included things such as animal videos, cooking lessons, non-comedic series, etc. As said, endorphins make us more resistant to pain. Since the presence of endorphins in the blood is difficult to measure at the chemical level, pain-resistance is typically used as an indicator of endorphin levels. So, before and after having the participants exposed to these videos, Dunbar tested them in tasks related to pain resistance. In particular, the participants received increasing degrees of temperature or pressure in their arms (with thermic and pressure sleeves), and were asked to say when the pain had reached the point where they could not stand it. What Dunbar found was that the greatest levels of pain resistance in the participants were correlated with the humorous videos, in particular the ones that made people laugh the most.

In cognitive terms, humor has been described as operating a frame-shift. Steven Pinker credits author Arthur Koestler with the clearest formulation of this principle (1964), and he summarizes it in the following way:

> Humor begins with a train of thought in one frame of reference that bumps up against an anomaly: an event or statement that makes no sense in the context of what has come before. The anomaly can be resolved by shifting to a different frame of reference, one in which the event does makes sense. And within that frame, someone's dignity has been downgraded. (Pinker, 1997)

We can see that effect clearly illustrated in the joke at the beginning of this section. In the woman's frame, her son had been unfairly judged as ugly. But, in order for her to make sense of the fact that a well-intentioned man (who is encouraging her to defend her dignity) thinks her baby looks like a monkey, she must shift her original frame and infer that her baby *is indeed* ugly.

This *dignity downgrade* produced by humor has been identified and described by many philosophers and literary critics in the past (e.g. Bergson, 1900; Bakhtin, 1940). But the crucial aspect to undernline here is that humor performs this downgrade *undercover*. In other words, a humorous attack is an indirect attack: The frame-shift is not explicitly explained, but it occurs in the minds of the listeners. In this sense, our capacity for humor could have effectively evolved among humans as a tactic for undermining undeserved authority in a way that somehow protects the speaker (Provine, 1996; Pinker, 1997). The social importance of this tactic is such that it has been personalized in many cultures across history: such as in the figure of the jester in medieval courts and all the variety of comedians that we know nowadays. We also see this role represented by many characters in literature and cinema, every time a powerless individual is shown dismantling illegitimate authorities or bullies by ridiculing them with a humorous attack. Here is one of my favorite literary examples of this:

> A linguistics professor was lecturing his class the other day. "In English," he said, "a double negative forms a positive. However, in some languages, such as Russian, a double negative remains a negative. But there isn't a single language, not one, in which a double positive can express a negative."
> A voice from the back of the room piped up, "Yeah, right."

The professor is in a position of authority with regards to the student. The student thinks the professor is wrong. But, if she overtly attacks the professor, she risks suffering consequences. With this joke, the student wins the argument before it even starts. Because, if the professor event attempts to respond as if the student was contradicting him, he is already admitting that the student is right (that "yeah, right," two positives, do make a negative, even in his own language). This way, the attack is overt but safe. It reaches the goal and protects the speaker. That is why it works, and that is why we find it funny. Moreover, it operates at a social level: every person who understands the double-positive negative of the student is already on her side, by the very act of understanding it. The joke leads us to imagine the whole class laughing –laughter is also contagious, which is also an evidence of its social nature (Dunbar, 2011).

Another evidence of the fact that humor evolved as a social tool for reducing social risk, for "playing safe," is precisely that one of the first manifestations of laughter is related to tickling (s. Provine, 1996). Just like in other primates, tickling has an evolutionary function that is actualized in children play. Infants from all primate species play simulated fights, which is a form of training of an adaptive skill. In this context is that tickling makes sense. Tickling triggers laughter, which works as an ostensible marker of the fact that we are not *really* attacking each other, but only *simulating* a fight, "playing safe" –Monkeys actually produce in these cases also a kind of laughter, which is called *panting*. For this reason, we are sensible to tickling precisely in the parts of our bodies that are most vulnerable. And, also for this reason, it is so difficult for us to tickle ourselves –you can try: it does not work… laughter is social, and there is no need to protect yourself from your own body.[29]

[29] The distinction between one's body and others' bodies is so pertinent regarding tickling that being able of tickle oneself is considered to be a possible sign of schizophrenia –schizophrenic patients are, indeed, more capable of self-tickling (Lamaitre et al., 2016).

WHAT ARE EMOTIONS? THE ELEPHANT AND THE RIDER

Humans have long debated about what our emotions are and particularly how they relate to reason. Through philosophy, religion, art, and literature humans have conceptualized the nature and relationship of these dimensions of the human mind in diverse ways. Let us consider what cognitive science has to say about it.

Emotions are processed by the most primitive parts of our brains: the limbic system (which includes the amygdala and the hypothalamus, among others). They are part of our intuitive software, which –as seen in previous chapters– is innate and gives us a set of instinctive and automatic ways of processing information. We can think, in this sense, of emotions as programs of predefined responses that have been tested over evolutionary time to be effective for confronting predefined situations (s. Barkow et al., 1992).

A clear example of this is *fear*: If a predator is attacking us, it might be counterproductive to try to calculate rationally all the pertinent variables and evaluate the best response. The variables are always too many. And rational calculations are costly. Instead, it would be more advantageous for any organism to count with a predefined program that is triggered by "perception of threat," interrupts other conscious cognitive processes, and automatically moves the organism to act in a plausibly befitting way (i.e. *fight or flea*) (s. Oatley, 1992). Such a program would have been obviously useful, and therefore it must have evolved as an adaptation. Indeed, most animals have this program, and also do we. We call it *fear*. That is why, when fear is triggered in our bodies, all of us have a similar response: we all experience a similar symptomatology (our bodily temperature changes, our heart rate increases, we breath more rapidly, we produce adrenaline), we make similar bodily expressions (we *look* afraid), and we behave in similar predefined ways (fight-or-flea response). Humans actually share a set of basic universal emotions, as part of our biological endowment, that we all process similarly and can easily recognize. As first conceptualized by psychologist Paul Ekman, these include: anger, disgust, fear, happiness, sadness, and surprise (Ekman, 2007). Different cultures might make attribute different meanings to emotions, might judge differently what counts as funny, for example, but we all respond to perceived funniness in the same way programmed in our biology: by releasing the same chemicals, feeling the same joy, and making the same facial expressions.

What we call *reason* is a series of cognitive capacities that are performed mostly by the brain cortex –which evolved later and is excessively developed in humans in comparison with other animals. These capacities that constitute our reason are also adaptive, that is: naturally selected for being advantageous for survival or reproduction. Among other things, our reason allows us to learn new skills (that are not evolutionary determined, such as playing the guitar or reading), to make better inferences about the past and hypotheses about the future, and also to justify what the elephant does intuitively (we will come back to this last function -s. Gazzaniga, 2000; Haidt, 2012).

Psychologists and philosophers, for a long time, thought of emotions as a system parallel and independent from reason. As if, for every decision, our brains could choose with which protocol to process it: Do you chose *emotionally* or *rationally* what to order in a restaurant? In the 1990s, neurologist Antonio Damasio found empirical evidence to show how much more complex the interrelation between these two processes is. The previous conception –which considers emotions and reason as separate– has been historically credited to philosopher René Descartes. In 1994, Damasio published his results in a book provocatively entitled: *Descartes' Error*.

Damasio treated a series of patients that had suffered brain lesions in the areas that process emotions. These patients were still perfectly capable of performing normally in IQ tests, but they had become unable to experience emotions: they could hardly feel joy, sadness, amusement, or fear, for instance. The curious thing is that, as a consequence of this condition, the patients started to show also difficulties for performing many activities that people typically think of as rationally-oriented. For instance, many people think that, in business, emotions are more an obstacle than an advantage. However, the emotionally-hindered patients did not become efficient heartless business men; instead, they also became unable to take rational decisions, such as when to buy, when to sell, or even whether to attend a meeting. Here is a limit example. One of Damasio's patients reported a problem he had when going to the supermarket to buy cereal. He stood in front of the shelve of cereal boxes and started to rationally calculate: "A is cheaper than B and C. But B and C have nuts. B has also chocolate, but C has fruits… etc." After one hour of rationally calculating advantages and disadvantages among cereals, the patient still could not decide which cereal he preferred. Damasio realized then to what extent reason and emotion are intertwined in our minds. Perhaps we can make abstract rationalizations (like mathematical operations) without recruiting our emotions; but most rational decisions we take in our lives are about concrete objects, subjects, and situations. And, in order to make rational decisions about these things, they must *matter* more or less to us. Calculating the best option when buying a box of cereal, in a purely rational way, would be both costly and inefficient: there are infinite variables that could be pertinent (e.g. content, price, size, dietary consequences, cost of production, demand, etc.). The way our minds normally deal with these situations is by intuitively project value to the things we perceive –Damasio called these projections *emotional markers* (Ibid.). This way, our minds form an emotional map of the world, and it is on the basis of these map (of how we intuitively *feel* about the things that surround us) that our reason can operate to take a decision. In short, without emotions, reason is groundless.

Damasio showed us that emotions and reason are deeply intertwined systems. But something else should also be added to this explanation: the relationship between the two is not really symmetric. What psychologists have been repeatedly stressing in the last decades is that emotion (more than reason) is in command. Social psychologist Jonathan Haidt found empirical evidence of this by analyzing how people perform moral judgments (Haidt, 2006 & 2012). The participants of these experiments were asked to read stories that portrayed taboos, such as the following one:

> Julie and Mark, who are sister and brother, are traveling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if the tried making love. At the very least it would be a new experience for each of them. Julie is already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other. (Haidt, 2012).

Afterwards, then the participants were asked: "So what do you think about this? Was it wrong what they did?" 80% of the participants said that it was wrong for Julie and Mark to have sex, and 20% said that it was OK (Ibid.). But the most interesting thing happened when the participants were asked to explain the reasons of their moral judgments. Here is a sample that illustrates the typical structure of the responses:

> Experimenter: So what do you think about this, was it wrong for Julie and Mark to have sex?

> Subject: Yeah, I think it's totally wrong to have sex. You now, because I'm pretty religious and I just think incest is wrong anyway. But, I don't know.
> Experimenter: What's wrong with incest, would you say?
> Subject: Um, the whole idea of, well, I've heard –I don't even know if this is true, but in the case, if the girl did get pregnant, the kids become deformed, most of the time, in cases like that.
> Experimenter: But they used a condom and birth control pills–
> Subject: Oh, OK. Yeah, you did say that.
> Experimenter: –so, there's no way they're going to have a kid.
> Subject: Well, I guess the safest sex is abstinence, but, um, uh… um, I don't know, I just think that's wrong. I don't know, what did you ask me?
> Experimenter: Was it wrong for them to have sex?
> Subject: Yeah, I think it's wrong.
> Experimenter: And I'm trying to find out why, what you think is wrong with it.
> Subject: OK, um… well… let's see, let me think about this. Um–how old were they?
> Experimenter: They were college age, around 20 or so.
> Subject: Oh, oh (looks disappointed). I don't know, I just… it's just not something you're brought up to do. It's just not–well, I mean I wasn't. I assume most people aren't (laughs). I just think that you shouldn't–I don't–I guess my reason is, um… just that, um… you are not brought up to it. You don't see it. It's not, um–I don't think it's accepted. That's pretty much it.
> Experimenter: You wouldn't say anything you're not brought up to see is wrong, would you? For example, if you're not brought up to see women working outside the home, would you say that makes it wrong for women to work?
> Subject: Um… well… oh, gosh. That is hard. I really–um, I mean, there's just no way I could change my mind but I just don't know how to–how to show what I'm feeling, what I feel about it. It's crazy!
>
> (Ibid.: 46-47)

What Jonathan Haidt discovered with this and other experiments (which were repeated across cultures –s. Ibid.) is that we seem to make moral judgments intuitively, on the basis of our emotions. And reason seems to operate as a subsequent elaboration for justifying these prior intuitive judgments. This perspective is known in psychology as the Intuitionist Model. And Haidt summarizes with the following principle: "intuitions come first, strategic reasoning second" (Ibid).

Haidt illustrates these dynamics between intuitions and reason with the metaphor of the elephant and the rider. The elephant represents our intuitive systems: our immediate phenomenological apprehension of the world and our instinctive emotions. The rider represents abstract reason, disembodied logic. Now, what Haidt stresses is that –despite whatever our inner riders might think– the elephant is the one in command.

The elephant is bigger and evolutionary more ancient than the rider. The elephant corresponds mainly to the neurological systems of our most primitive brain areas: the limbic systems. The rider corresponds to the superficial cortex, which is a later development that is characteristic of the human brain. In consequence, we should see the rider as having evolved *to serve* its elephant (due to the economy of natural selection, new structures evolve in integration with the old ones, enhancing their fitness, not making them obsolete). We could say that, through evolution, the elephant grew a little skillful rider on top of its own back. This rider increased the chances of survival and reproduction of the elephant, by allowing it to make better hypotheses and inferences about the past and the future, to better elaborate abstract ideas, to learn new skills, etc. All of this is done in service of the elephant. And one of the central functions of the rider is precisely to defend and validate whatever the elephant intuitively does and feels, as a press secretary or a lawyer would do. This is precisely what

Haidt's experiments showed. The participants on these experiments intuitively made moral judgments on the spot. Only when they were asked to provide explanations, they would recruit their reason, but their reason would behave like an obsequious press secretary of their spontaneous and unyielding intuitions. Under the description we have done of the relationship between emotions and reason, people's behavior becomes immediately understandable: One cannot affect the boss' judgment by merely addressing his or her press secretary. Likewise, people rarely change their minds about their intuitive judgments only by being exposed to factual evidence and rational arguments, because reason is not the one *making* these judgments, but it is merely the one *defending* them –most political debates are examples of this.

The lesson that Jonathan Haidt leaves us with is that, so as to reach people's deepest convictions (the intuitive core that guides their judgment), we mustn't address the press-secretary rider, but the boss elephant. And how do you talk to the elephant? The answer is: through things like art, music, and literature. The stories we have presented in the first section of this chapter serve as an example of this: They do not talk abstractly about the concepts of danger, sorrow, or respect; instead, they display concrete life-like situations –two men breaking into a house, a woman dying, a woman offended by a bus driver, etc.–, situations that the readers can effectively simulate in their minds, so that their instinctive elephants (designed to cope with concrete real-life situations) feel actually appealed to respond in emotionally engaged ways. Thereby, stories can work like *adaptive fictions* that bring concepts to flesh, and this is crucial because (coupling Damasio's and Haidt's insights) without the elephant, the rider would fall.

EMOTIONAL BIASES AND COGNITIVE ILLUSIONS

In 2009, journalist Rob Walker lead an experimental project called *Significant Objects*. He wanted to measure to what extent stories affect our feelings about objects and, in turn, our behavior towards them. He bought 200 items from Ebay, with an average price of $1 each. He called 200 literary authors, and ask them to write fictional stories about these items. He then posted the items back in Ebay. The new advertisements were just like the original ones, except for the addition of the corresponding story for each object. All the items were sold. But, thanks to the addition of stories, instead of paying $1 per item (which was their original price), people paid now an average of $62 per item. How much value can a story add to an object? In Ebay, at least, the answer seems to be: around 6,200%.

As shown by this and previous examples, stories influence greatly our emotions, and our emotions influence, in turn, our behavior. But emotions go even deeper: psychological experiments –in which particular emotions are induced into test subjects with different kinds of stimuli– have systematically shown that emotions can bias the most diverse cognitive task, from our direct perception to our moral judgments or our interpretations of meaning. Here are some examples. Positive mood makes it easier for people to perceive global components in pictures and more difficult to perceive local components –negative mood produces the opposite (Schmitz et al., 2009). People are more severe in their moral judgments of others when making these judgments after hearing audios of angry screams and loud metallic noises or after tasting bitter substances (Prinz, 2004.) Judges give harder sentences and professors give lower notes when they are hungry (and grumpy) than after having lunch (when they feel more relaxed) (Mladinow, 2012). People's interpretation of what is depicted in an image can be altered by accompanying the image with different kinds of music –in one of the experiments that proved this, the participants would interpret the image of a lying woman as if she was *sleeping* when the image was shown along with soft harmonic music, but they would interpret it as *dead*, instead, when the accompanying music was dissonant (Prinz, 2004). Emotions really seem to bias our behavior, perception, and judgments without us noticing their influence. And, what is even more shocking, we are often unaware of the very *cause* of our emotions.

The clearest example of this kind of phenomenon is perhaps the placebo effect, which induces people to feel typically lower levels of pain and to misattribute the cause of their relief to a particular medicament or treatment that they believe they have received (Chaplin, 2006). These cases in which people are lead to misjudge their own mental states (or the cause of their mental states) are called *cognitive illusions*, and many of these cognitive illusions concern directly our emotions. The most famous early research on this subject was performed by Stanley Schacter and Jerome Singer (S&S), in the 1960s (s. Shachter, 1996). In one of their experiments, they gave shots to the participants of an experiment, telling them that it was a vitamin called "Suproxin," that would enhance their visual skills. But what the participants actually received was a shot of adrenaline, which makes people get excited (either for positive or negative emotions). The participants were divided into three groups:

- Informed: They were told about the effects of the shot, explaining them as "secondary effects" of the Suproxin.
- Ignorant: They were not told about the effects of the shot.
- Control: They only received an inert saline solution and were told nothing.

After administering the shots, the researchers asked each participant to wait in a room for 20 minutes. In the room, there was an actor. The actor would simulate that he was also a participant, and would act as being either very happy for the privilege of being part of the experiment or very angry about it, complaining about having to wait or about the way in which the experiment was conducted. Afterwards, the (real) participants were asked how they felt. Here are the results: The informed and the control groups did not feel any emotional change after the experiment; but the ignorant group did feel happier or angrier depending directly on whether they had encountered a happy or an angry actor.

These results confirmed exactly S&S's hypothesis, and here is the theory that explains the phenomenon (which has been replicated many times, and is today known as the Schacter-Singer Model). Each of our emotions is correlated with a particular symptomatology (e.g. fear increases our heart rate, disgust reduces it, anger affects our bodily temperature, sadness makes us cry, etc.). But the connection between these two dimensions is not immediate. When we are induced to experience these bodily changes, our brains search for cues that might explain why we feel what we feel. After processing the available information, our brain can *interpret* whether we are actually sad or we are just crying because the onions we are chopping are very strong, for instance. In S&S's experiment, after receiving the adrenaline shots, the participants naturally felt a sudden euphoria, and their brains had to explain it. The informed participants had a very clear handy explanation for their euphoria: they had been told that it was an expectable secondary effect of the shot. But the ignorant group did not have this knowledge, so their brains would have started searching for further information. As soon as they found a pertinent cue that indicated that the experiment was either a privilege or an annoyance (i.e. the actors' expressions), their brains took it as a plausible explanation of their sensations. Misattributing the cause of their sensations to particular emotions actually elicited these emotions in the participants. Expectedly, the two other groups did not experience any emotional change: they had no unexplained euphoria to account for, so they were less influenced by the actors.

It is important to notice that this process is unconscious: the participants were not aware of what had induced them to interpret their euphoria as happiness or anger. They just felt happy or angry. This shows that our emotions (and our judgments of them) are significantly modelled by processes that our brains perform under our awareness and that can be manipulated. For this reason, we are also inaccurate judges when we try to evaluate introspectively why a certain literary work makes us feel the way it does. The cause of our own emotions is not really transparent to ourselves.

S&S's experiment was repeated with other kinds of emotions, and some of these experiments involved stories. One of them had the participants perform physical exercises and, when they felt recovered, they were showed a French erotic film, and they had to rate how exciting they found it. Even though the participants felt physically recovered, their brains where still flooded with adrenaline when exposed to the film. The group that was informed about this (informed), gave the film 28 points out of 100, similarly to the control group –who had not made any physical exercise– (31 points). However, the ignorant group gave the film 52 points. Expectedly, they mistook their arousal as being of a sexual nature and as caused by the film, leading them to judge the film as more erotic than they would have judged it in other conditions.

The bottom-line is that we do not interpret and judge literary works (or any other cultural object) only with our reason. Our emotions and our body intervene actively in our processing of stories. And this is not an accidental fact, but possibly one of the fundamental features that make stories so important for humans.

WHY ARE WE MOVED BY FICTICIOUS STORIES?

We can derive from this model of emotions a hypothesis to explain a long riddle of literary theory: Why is it that readers are moved by fictional characters and events to begin with? As seen in the previous chapter, stories draw our interest for many reasons (causal reasoning, problem solving, mindreading training, mating, etc.). But why would it be that stories also make us feel particular emotions, even if the events told in stories are fictitious and, in consequence, do not concern us directly? Perhaps what leads us to feel happy with the success of the hero of a story, angry against his or her enemies, or sad with his or her defeat, is also the result of a cognitive illusion.

Daniel Schacter and Joseph LeDoux explored the connections between emotions and memory (s. LeDoux, 1996). They observed Alzheimer patients for a long time, and they frequently spotted a particular phenomenon, which can be exemplified by the following case. A nurse visited an Alzheimer patient and gave him a painful shot. The next day, the nurse came back, and as soon as the patient saw her, he felt afraid. He claimed that he didn't know who she was, that he had never seen her before, but her presence still made him feel afraid. It is not strange to find Alzheimer patients that cannot remember a particular person, object, or event, but do remember the pertinent emotions associated with them.

This way, Schacter and LeDoux discovered that we seem to have two different systems for encoding our memories: a representational one and an emotional one. The first one constitutes a conscious storage of representational information about the most diverse things (the words you've learnt, the faces you've seen, the names of people, dates, etc.). The second one stores information such as our skills and emotional memories (how to ride a bike, how you felt around your grandmother, etc.), and it is largely unconscious (S&L talk about an *implicit* memory system). Normally, when the representational memory of an experience is triggered, this triggers also the corresponding emotional memory associated with that experience. However, as shown by cases such as the ones of the Alzheimer patients, these two systems have some relative independence from one another: we can still have emotional responses even when ignoring what is that moves us in that particular situation. The two processes are actually performed by different regions of the brain (Schlacter, 1996: 171-2).

As Schacter and LeDoux observed, our emotional memory can get activated independently of our representational memory. Maybe this does not only happen with Alzheimer patients, but also with people that are engaged in reading literary stories. Let me illustrate this with a hypothetical scenario: Let us imagine that I have lost my family when I was very young. I do not feel particularly sad at the moment, but then I read the story of the last Yaghan speaker, who lost her family, and I start to cry –coming to judge the story, thereafter, as being very sad. Perhaps what happens here is that the fictitious story activates the emotional memory of my real loss. As soon as the correspondingly symptomatology is released (I start crying), my brain picks up the most immediate cue to explain my feeling: the story I'm reading. In consequence, my brain does not activate the representational memory (it does not consciously recall my personal loss), because it doesn't have to: It has already found a more immediate plausible instance to attribute the emotion to: the story I'm actively reading.

If this was the case, then stories could be described as a technique for creating a particular kind of cognitive illusion. We could say that stories make us re-enact our own emotional memories while, at the same time, making us mismatch them with the fate of others. They deviate our focus of attention from our own representational memories and trick us into misattributing real feelings to fictitious characters and events. But, for this very reason, this activity can be considered as a training in selflessness, in that it leads us to find analogies between our own experiences and the experience of others in the identity of our emotional structures. Perhaps it is this very mechanism the key that makes of stories such a crucial part of the emotional education, in forming our empathy and our social skills.

THE LIMITS OF EMPATHY: FOR HOW MANY PEOPLE CAN YOU CRY?

Imagine that you hear about an earthquake where 821 people died; some hours later, you discover that actually 823 died in that earthquake, instead. Now imagine a second scenario: First you hear about a man that died in a car accident; some hours later, you discover that also his family was in the car: his wife and daughter died with him, but not his baby son, who was in the back sit. In which of these scenarios do you think that hearing about two more casualties would increase more your distress about the tragedy in question?

This thought experiment is meant to illustrate the limitations of our capacity for empathy. Psychologists have shown that our empathy is not indefinite: we seem to be better at imagining what few people must be feeling than what hundreds, thousands, or millions of people must be feeling. This has to do with the usual sizes of the small communities of hunter gatherers in which our evolutionary endowment was formed. Also, our empathy for the suffering of others reduces in intensity over time –condition that psychologists actually call *compassion fatigue* (Jennifer&Anderson, 2011).

Stories take advantage of these limitations of our empathy by portraying social events embodied in particular characters going through intense experiences before our eyes. For example, many people died when the Titanic sunk; nevertheless, most people that cried when seeing the film *Titanic* (Cameron, 1997) already knew objectively about this event in advance and had never shed a tear about it. It was only when the story was portrayed from the perspective of particular characters (even if fictional) that millions of people around the world were brought to develop a deep sense of empathy towards the crew of this old ship and imagine more intensely their pain.

Our empathy is also biased in favor of individuals that we know, that we feel close to us, or related to us. We tend to suffer the death of a known neighbor more than that of an unknown individual. This bias has been considered as an evolutionary adaptation, since it favors in-group social behavior –which would have made groups more efficient in surviving, growing, and competing with others. Stories leverage this bias each time the identities and personalities of the characters are displayed before narrating the events that they will face: The audience suffered the death of Jack, in *Titanic*, more than that of all the other anonymous members of the crew –probably even more than that of the people that really died in the historical event. And, in the thought experiment of the beginning of this section, the mere mention of the family-relationship between the victims of the car accident makes the deaths so much more tragic that even the mention of a baby who did not die becomes heartbreaking.

This bias and the capacity of stories to manipulate it has also an obvious downside: it can lead our moral decisions in the wrong direction. In a series of experiments, evolutionary psychologist Daniel Batson showed participants different cases of patients that needed an organ transplant (s. Batson, 2011). Some of the patients needed it more and some less. But a second variable was introduced: some the patients were personally described with a little story that told who they were, what they did, how their lives were, etc. The astonishing discovery was that the participants of the experiment decided who disserved the transplants guided more strongly by the personal stories than by the actual needs of the patients. This suggest the extent to which the very sense that positively leads us to form personal connections can also lead us to take less moral decisions.

This bias is also profusely leveraged by movies and novels each time the audience is brought to moderate their moral judgment of criminal actions by portraying them from the perspective of the criminals themselves, by engaging the audience with their personal stories, backgrounds, and mind-sets (e.g. *Crime and Punishment, Marnie*, *Braking Bad*, *Narcos*, *Spider*, etc.).[30]

Despite the downsides of the fact that stories can manipulate our empathy and moral responses, the advantages of this power of stories must also be stressed. Stories are the language that our intuitions and emotions better understand. And this fact is not to be undervalued. It is precisely what makes of stories such an efficient tool for educating the sensibility of human societies: by making us care more for people non-directly related to ourselves and thereby allowing us to better address large-scale social problems. The numerous films that were made about the Holocaust probably helped raise awareness against racial crimes more massively than most of the theoretical analyses and objective but impersonal reports written about the subject.

People can feel more for the tragedy of a fictional character they have gotten to know than for the tragedy of countless people they have never seen. In a way, this can be felt as a discouraging realization, but it gives us a great insight into human nature: namely, it is very telling of why we are literary animals. Understanding this human characteristic can actually give us a great lesson for thinking of current issues and how to communicate them. Global warming, for example, is actually a clear case of a huge threat that, due to the nature of our minds, we are bond to find difficult to grasp and we are unlikely to take as seriously as it

[30] S., respectively, Dostoevsky, 1866; Hitchcock, 1964; Gillian, 2008-13; Brancato, 2015-2017; Cronenberg, 2002.

deserves. The potential risks of this threat are long-term and massive; whereas our empathic skills evolved for facing short-term and local problems, of communities of directly related members. Our reasonable riders may be capable of objectively calculating and anticipating the dangers of global warming, but our emotional elephants cannot *feel* these dangers solely on the basis of objective evidence and abstract arguments. Having this in mind, perhaps it would help to start creating stories that describe the dangers of global warming (and any other large-scale problem) through the experience of concrete personalized characters who face these dangers in their flesh.[31] In this sense, the psychology of emotions can enlighten one of the fundamental powers of literature: being the language that allow us to talk with our intuitive elephants. From this perspective, we can describe literary stories as *adaptive fictions*: not aimed at literarily stating a truth, but at conveying a truthful metaphor, a figurative way of speaking, sensible to the design of our minds, that allows us to better grasp (better embody, better intuit and feel) a certain truth.

Gilbert Keith Chesterton wrote in the first chapter of *Orthodoxy* (1908):

> I have often had a fancy for writing a romance about an English yachtsman who slightly miscalculated his course and discovered England under the impression that it was a new island in the South Seas. (…) I have a peculiar reason for mentioning the man in a yacht, who discovered England. For I am that man in a yacht. (…) I am the man who with the utmost daring discovered what had been discovered before.

In this book, Chesterton tells us how he abandoned, when young, the superstitious popular stories he was taught in his society, pursued the meaning of life through rational thinking by his own, and found it again in the old superstitious popular stories he already knew.

I feel that, in a certain way, we have gone down a similar path in this first part of the book. We began with the intention to make sense of literature. We have undertaken a scientific exploration of the nature of literature and the human mind. And this exploration has brought us to discover that the human mind, in its core, actually makes sense through literature.

[31] An example of this, regarding data-privacy security, can be found in the short film “Autocorrect”, by Martín Piroyansky (2018).

Part II

DATA SCIENCE

YOUR COMPUTER KNOWS YOU BETTER THAN YOU KNOW YOURSELF

CHAPTER 7

BIG DATA AND PSYCHOMETRICS

DIGITAL LENS FOR READING MINDS AT LARGE SCALES

THE INFORMATION AGE

You wake up in the morning and you read your favorite newspapers online, you check your emails, your friends' posts on social media, and the weather for the day –so as to decide what to wear. You go to the street and you localize the place you need to attend by using a digital-maps app, which also tells you how to get there. You work using specialized programs designed for the needs of your field, and you update them regularly. You handle your finances online: your bank transaction, your purchases and investments. You even found your job online. You search online any kind of information you need. And if you don't find it you may post a question about it in an online forum, so as to exchange opinions about it with other people around the world. In your free time, you read an e-book, watch a film, or listen to music online. When you travel, you buy your ticket online, you find accommodation online, and you can even organize your itinerary by following recommendations online. Later on, you go on a date with a person whose profile you matched in a dating app. You will choose a restaurant nearby by browsing their online ranking and reading some reviews. In some places, you can pay your meal with a cryptocurrency that you store in your digital wallet. All along the day, you talk through chat and videoconferences with people from around the world and, at the end of the day, you go to sleep and set up your alarm clock in your smartphone and browse your customized news app one last time before you fall asleep –and let the phone turn off in your hand.

None of this is news: we are in a digital era, and few technological inventions have changed so much and so rapidly our living conditions and habits. Digital media represent the peak of thousands of years of human development. They have fulfilled one of the most ancient dreams of humankind: the smallest device with the greatest power, a pocket-size machine that facilitates a great deal of the most fundamental needs of our lives. The computer is the ultimate human tool. There are, of course, side effects to this, which we experience every day: people seem to spend progressively more time looking at their digital screens than at the world and people around them. This creates not only a new kind of isolation, but also other problems associated with attention and sedentariness, for instance. But the advantages that the computer and internet have brought us are so plentiful and deep that people has readily accepted to sacrifice many aspects of their traditional lifestyles so as to embrace digitally-mediated kind of life. We have become *users* –just as the state had characterized us as *citizens* and the market as *customers*.

We are constantly moving around the digital space. We feel very familiarized with it. But there is a crucial aspect of the digital space that is often disregarded: by focusing in the massive amount of knowledge we can learn from the web by surfing it, we often fail to notice the massive amount of knowledge that the web can learn about ourselves by having us doing that. That is the backside of the fabric, which is not immediately obvious for every one: the small but distinguishable changes that each user produces in the digital space on his or her way through it, the marks and registers we leave (for anyone to read) with each movement and decision we make. In other words, the millions of digital footprints that form the so-called *Big Data*.

BIG DATA: EVERY STEP YOU MAKE

Each interaction we perform with digital media leaves footprints in the web: millions of bits of data are created, transferred, and registered in the world servers every second due to the personal interaction of millions of users. We constantly produce data about all aspects of our behavior: what we like, read, comment, purchase, where we are, how much we move, when, etc. We generate several megabytes of online data per day only by using a smart phone in a regular manner.

This way, we have produced and registered more information in the last five years than in the previous five thousand. That colossal and proliferating cloud of data is what has come to be called *Big Data*.

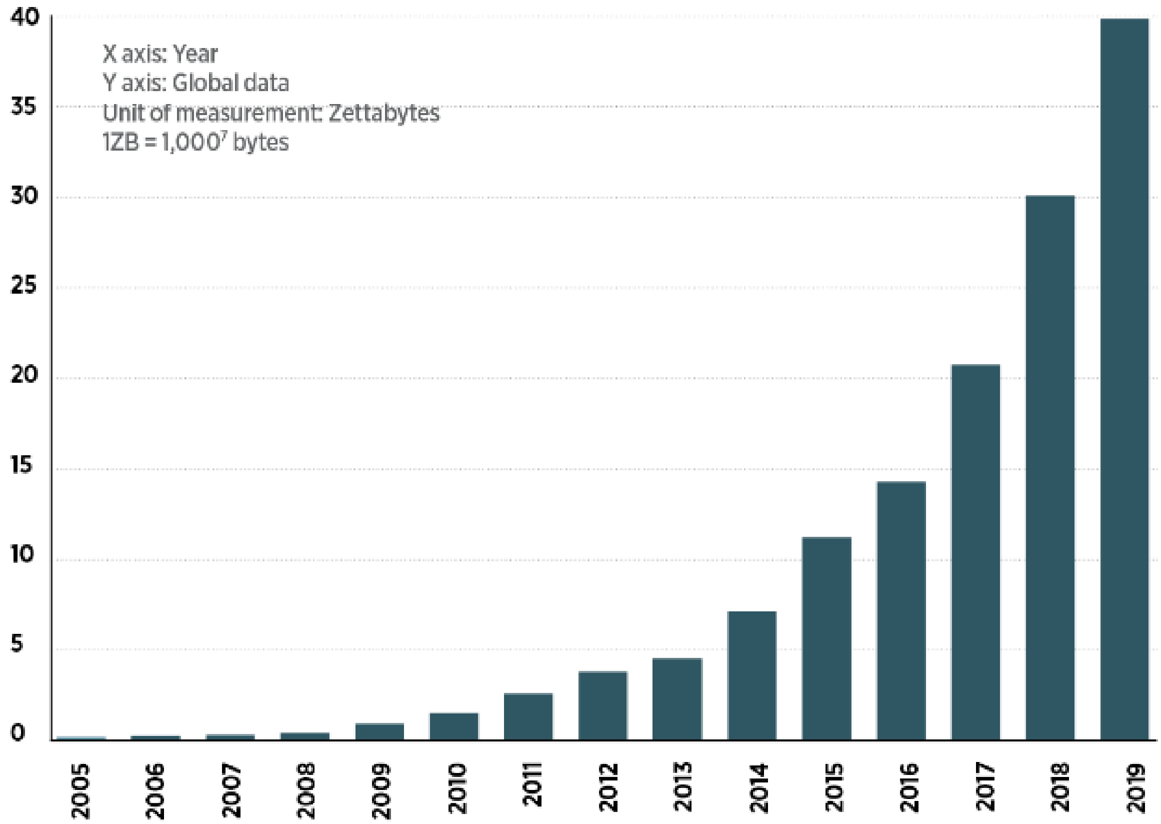


Figure 1: Data growth over time. (UNECE, 2016). (Note: Post-2016 figures are predicted).

Never in history have we had access to so much information about human beings, their behavior, and tendencies as today. In this sense, the big data constitutes an opportunity whose potential we still have not fully realized. But, at a fast speed, data science is developing digital tools for analyzing this data. And the results of this relatively recent discipline are already producing deep changes in virtually every field of research, including the social sciences.

PSYCHOMETRICS: THE QUANTITATIVE STUDY OF PEOPLE'S MINDS

In 2012, psychologist Michal Kosinski –then, a doctoral student at the Psychometrics Center of the Cambridge University– and his colleagues posted a quiz in Facebook (s. Kosinski et al., 2012). The quiz was a psychological test. It would ask a series of questions, out of which each user would have his or her personality measured in terms of the Five Factor Model (FFM) –the FFM is a standard taxonomy for personality traits which clusters them into five big categories: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (often represented under the acronym *OCEAN*). Thousands of people performed the test.

At this point, the researchers had the names of all these users, their answers to the test, and also the public information available in their their Facebook (FB) profiles. They decided, then, to search for correlations between these domains. And they found an enormous amount of patterns.

They began by assuming that each of our decisions and preferences expresses –directly or indirectly– some aspect of our psychology. Ultimately, what newspaper one reads, what music one likes, or which politician one votes are all indexes of one's personality. If this is actually the case, people's preferences would have distinguishable correlations with regards to people's personalities. So the researchers compared the FB *likes* of their sample of users with their psychological traits (as revealed by the test) and personal information.

This way, they discovered many individual correlations:

> The best predictors of high intelligence include "Thunderstorms," "The Colbert Report," "Science," and "Curly Fries," whereas low intelligence was indicated by "Sephora," "I Love Being A Mom," "Harley Davidson," and "Lady Antebellum." Good predictors of male homosexuality included "No H8 Campaign," "Mac Cosmetics," and "Wicked The Musical," whereas strong predictors of male heterosexuality included "Wu-Tang Clan," "Shaq," and "Being Confused After Waking Up From Naps." […] Each Like attracts users with a different average personality and demographic profile and, thus, can be used to predict those attributes. For example, users who liked the "Hello Kitty" brand tended to be high on Openness and low on "Conscientiousness," "Agreeableness," and "Emotional Stability." They were also more likely to have Democratic political views and to be of African-American origin, predominantly Christian, and slightly below average age. (Ibid.)

Each individual *like* gives in general very scarce information about all the other characteristics of a person. It is not easy (not for us nor for a computer) to predict whether a person is really open, conscientious, and agreeable by only knowing whether she likes "Hello Kitty." However, the computer has a crucial advantage over humans. A computer can aggregate the small predictions of hundreds, thousands, or millions of *likes*, and then the predictive power of the resulting model becomes much higher. If the computer counts in your profile many likes correlated with openness –and a very scarce relative proportion of non-openness likes– then the prediction of whether you are open and how much becomes increasingly accurate. Kosinski and his team trained his computer with a certain percentage of the results, and then tested it by making it predict on the remaining cases what answers would the user give in the psychological test. As a journalist reported:

> …before long, he was able to evaluate a person better than the average work colleague, merely on the basis of ten Facebook "likes." Seventy "likes" were enough to outdo what a person's friends knew, 150 what their parents knew, and 300 "likes" what their partner knew. More "likes" could even surpass what a person thought they knew about themselves. (Grassegger & Krogerus, 2017)

These results are shocking because they give us a glimpse at the enormous amount of knowledge treasured in Big Data. The Big Data is so massive that the simplest measurement becomes incredibly informative and insightful about our social and cultural behavior: just by looking at Facebook likes (such a seemingly trivial kind of data), the psychological profiles of a whole population become predictable.

The vertiginous field of possibilities opened by this field is as exciting as it is concerning. And the people invested in the market of information became immediately sensible to this. Indeed –as Grasser & Krogerus reported–: "On the day that Kosinski published these findings, he received two phone calls: the threat of a lawsuit and a job offer. Both from Facebook." (Ibid.)

CHALLENGES OF A COMPUTATIONAL SOCIAL SCIENCE

The power of these methods for describing and predicting people's personalities, preferences, and intentions became soon a motive of alarm. One of the cases that raised these alarms involved an English political-consulting company called Cambridge Analytica (CA). CA assumed a model similar to Kosinski and extended it. They performed a similar survey, but they also created an app that would track users' Facebook activity –which was installed when the people would take the quiz, but without their consent. They also added information from other digital platforms: where do people go, what websites they visit, what they buy, etc. Afterwards, they used this information for informing concrete electoral campaigns: namely, the Brexit referendum and the American presidential elections of 2016. Recently, the company has been sued for accessing private information illegally. But the social scandal emerged even before this was known: when a video was filtered that showed Alexander Nix, CEO of CA, explaining their method of research in front of an audience (s. Concordia, 2016).

The first information they would get out of people's data was their voting intention. For instance, according to the data, the larger the radio of movement of a person (tracked by our phone's localizers), the more likely was that person to vote for a Democrat in USA. Aggregating millions of data bits of this kind, their model enabled them to distinguish the voters who had already decided to vote for a particular party from the *swing voters*: the ones still undecided. This way, they would decide which was worth or not to be addressed for the electoral ends. Moreover, their model would give them information about the personality profile of each voter (again, in terms of openness, conscientiousness, agreeableness, etc.). And, thereafter, the company employed a technique now known as *microtargeting*, which consists in elaborating messages designed in accordance to the personalities of the addressees, on an individual basis.

This last strategy increased enormously the efficiency and impact of the political propaganda. And it is important to look at it in detail, because it is at the same time evidence of a crucial scientific discovery. There is a lot variety among humans: we are not all clones, we think differently, we have different personalities, and we can interpret differently the same message. For that reason, political-consulting companies have always manufactured different kinds of messages for different groups of any given society. However, they used to base their understanding of social groups in terms of traditional sociological models. Each of these models would assume a particular division as the pertinent one to be addressed for any given society: some would consider that the main subjective differences correspond to differences of social class; others consider age; others, gender, etc. So, the consulting companies would manufacture, for example, a kind of message for women and another kind for men, or a message for young people and a different one for older people, etc. But these top-down taxonomies –based on prior theoretical assumptions– are so general that systematically fail to capture individual nuances. From the viewpoint of most of these traditional sociological models, two sisters of a similar age that live together would necessarily fall in the same category of gender, class, age, and even household. And, in consequence, the consulting company would decide to issue the same message to both. Nevertheless, we know that these two sister could still have radically different political views and personalities, which would make them respond differently to the same propaganda. But in order to distinguish these individual differences, we would need a model that allows us to take a closer look at people's minds. This is the gap that the psychometric model used by CA gapped.

If instead of *assuming* what the pertinent personality-groups of a society might be, one measures the online activity of all of the members of the society, then a computer can discover by itself (in a bottom-up way) what these groups *actually are*. The computer would tell you if the minds of a society that might be more sensible to a political speech are better predicted by their class, age, gender, or perhaps by what products they buy online, what series they watch, or how many friends they have in their social media. Afterwards, the computer would cluster the results into distinctive groups, and it would produce, this way, a map of the pertinent subjective differences of the population of a whole country.

On this basis, CA could elaborate different propaganda for each of these groups (*microtargeting*). So, if two sisters, despite sharing gender, age, class, and household, has different personalities, the psychometric measurement would distinguish it, and they would receive different kinds of message. For instance, if one of them was measured as traditionalist, she would receive a spot where the political candidate is shown advocating for the legacy of the pertinent historical national figures; if the other one is measured as being more concerned with safety, she would receive a spot showing the political-candidate's plans for increasing security measures. In this way, a politician can address a mass, a whole population, with an efficiency much closer to that of a face-to-face interaction. It is a highly developed technology for large-scale communication. But the mindreading power of data science reveals in these kinds of technologies to be as great as it can be dangerous.

Michal Kosinski was, indeed, the first to alert about the risks of his own discoveries. As he declared in an interview: "Most of my studies have been intended as warnings. You can imagine applications that are for the good, but it's much easier to think of applications that manipulate people into decisions that are against their own interests." (Kosinski, 2017). And, indeed, his articles consistently insist in the legal and political aspects that this knowledge calls us to address: "The widespread availability of extensive records of individual behavior, together with the desire to learn more about customers and citizens, presents serious challenges related to privacy and data ownership." (Kosinski et al., 2012).

Every time a scientific discovery or technological invention proves to be useful and powerful, both well- and ill-intentioned people try to profit from it. The invention of the telescope became useful for exploring the space, but also for military ends; biochemical knowledge can be used for curing diseases but also for creating weapons. And, indeed, a great deal of the political and legal development of our societies emerges precisely as a consequence of the availability of increasingly powerful knowledge and technology. But a crucial aspect that must not be obliterated in this panorama is the importance that these discoveries and inventions have for human progress: inventions such as the telescope and the microscope also enabled the development of modern astronomy and biology. And while Einstein's theories enabled the creation of atomic bombs, they also reframed our understanding of physics –after which we found our current model of the universe- and allowed us thereafter to create nuclear energy, laser surgery, scanners, solar panels, digital cameras, GPS navigation, and many other technologies that are part of our current lifestyles.

Data science and digital technology confront us with an analogous situation: the same app that collects information of people's heartbeat when they jog can be used to study and prevent coronary diseases in large populations but it can also be used for military means – such as spotting fitter soldiers. Likewise, psychometrics and microtargeted advertisement can be used to influence people not to vote, and that would be a regrettable use, but these techniques can also be used to influence people to lead healthier lifestyles, not to smoke, to exercise, to recycle, and they can allow us to make faster, more accessible, and increasingly more accurate psychological diagnoses, for helping people match with jobs, friends, or partners, recommending readings one might like, or for facilitating the guidance of a personal education based on the individual interests and talents of each person.

We must foster the development of these techniques in one direction and to deter it in the other. And, indeed, the debate about the possible strategies and policies that could be used to regulate the use of Big Data are currently very intense (s. Broeders et al., 2017). But the first step necessary for properly addressing the potential advantages as well as perils of this new field is to get a deep understanding of it. Both prevention and profit can only come from research. In consequence, as much as data science requires us to assume an attitude of attentive precaution regarding its potential misuses, it also requires in us an attitude of passionate curiosity aimed at exploring the paths of knowledge and opportunities that it is opening.

DATA SCIENCE AS LENS FOR GAZING AT THE CULTURAL SPACE

The comparison of data science with telescopes and microscopes is not arbitrary. The dangers and alarms raised by psychometrics are indeed a measure of its amazing efficiency for describing ourselves, our minds, our intentions, desires, and believes. Data science is giving us the most sophisticated instruments of measurement for analyzing human culture and behavior at an unprecedented large scale. Inasmuch as the Big Data in the Cloud constitutes a space that reflects the cultural interests of the human kind of this era, the methods developed by data science constitute the lens that allow us to gaze at that space and discover the map of our collective minds. For this reason, data science has the potential to produce in the social sciences and the humanities revolutions comparable to the ones that the telescope and the microscope produced in astronomy and biology, respectively, at the beginning of the 17^{th} century, which marked the beginning of modern science.

The revolution that these methods are producing has given us a series of disciplines that are creating an empirically-based science of culture, driven by massive amounts of data about ourselves and our societies. I will refer to these disciplines in general as *digital humanities* (Burdick et al., 2016) –although this and other closely related fields have also been baptized with the names of *computational social science* (Alvarez, 2016), *cultural analytics* (Manovich, 2010), and *culturomics* (Michel&Liberman, 2010). The novelty exploited by these disciplines is that, for the first time in history, we have access to the largest representative samples of perhaps the most complex aspect of life (i.e. culture) and the technical means to analyze them.

In this second part of the book, we will explore some of the insights that data science provides us in relation to the particular cultural phenomenon that is the subject of this book: literature.

CHAPTER 8

QUANTITATIVE LITERARY STUDIES

HOW TO READ A THOUSAND BOOKS IN A SECOND

ARS LONGA, VITA BREVIS

This famous classic Latin aphorism ("Art is long, life is short") has been often interpreted as meaning that, whereas artists live and die like everyone else, their artworks can last forever. But, nowadays, it is becoming increasingly pertinent to interpret that aphorism in relation to another fact: our life is too short to hear all the music that was ever recorded, watch all the films that were ever filmed, and read all the books that were ever written.

A recent study by LiteraryHub compared life-expectancy in USA with the averages of books read by Americans per year (s. Temple, 2017). Here are the estimations of the amounts of books that you are likely to read before dying, by genre and age, in USA:

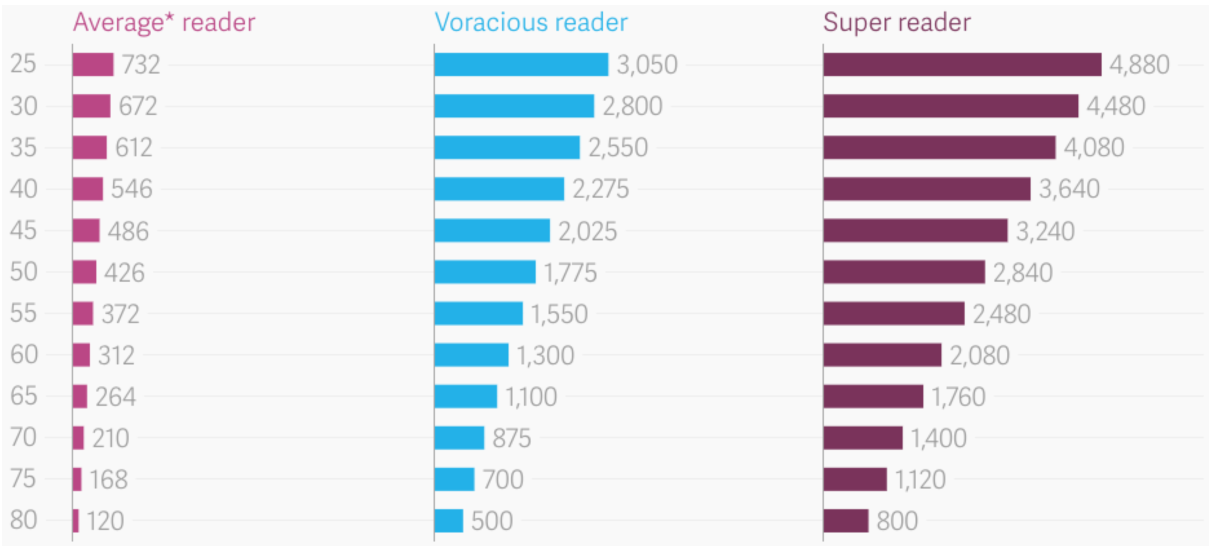


Figure 1. Books remaining before death *for women* by current age (Ibid.)

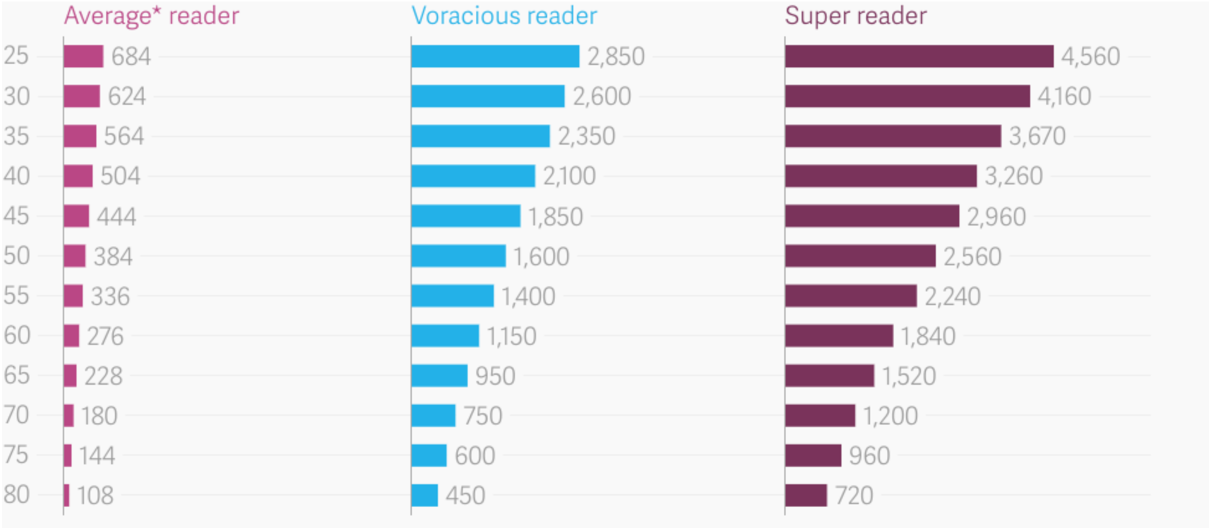


Figure 2. Books remaining before death *for women* by current age. (Ibid.)

According to this data, a 33 years-old man (such as myself) would have an average of 564 books left to read before dying. If such a man was a very diligent literary scholar, he might aspire to read even up to 4 thousand books –which is a really large amount for any human reader.

However, all these figures become immediately insignificant when we consider the staggering amount of books in the world and the speed at which they are being published. Only in a single year, an average of 2 million different titles are published across the world. China alone produces around half a million books per year. Here are the data of the most intensely publishing countries in 2015, as registered by the International Publishers Association:

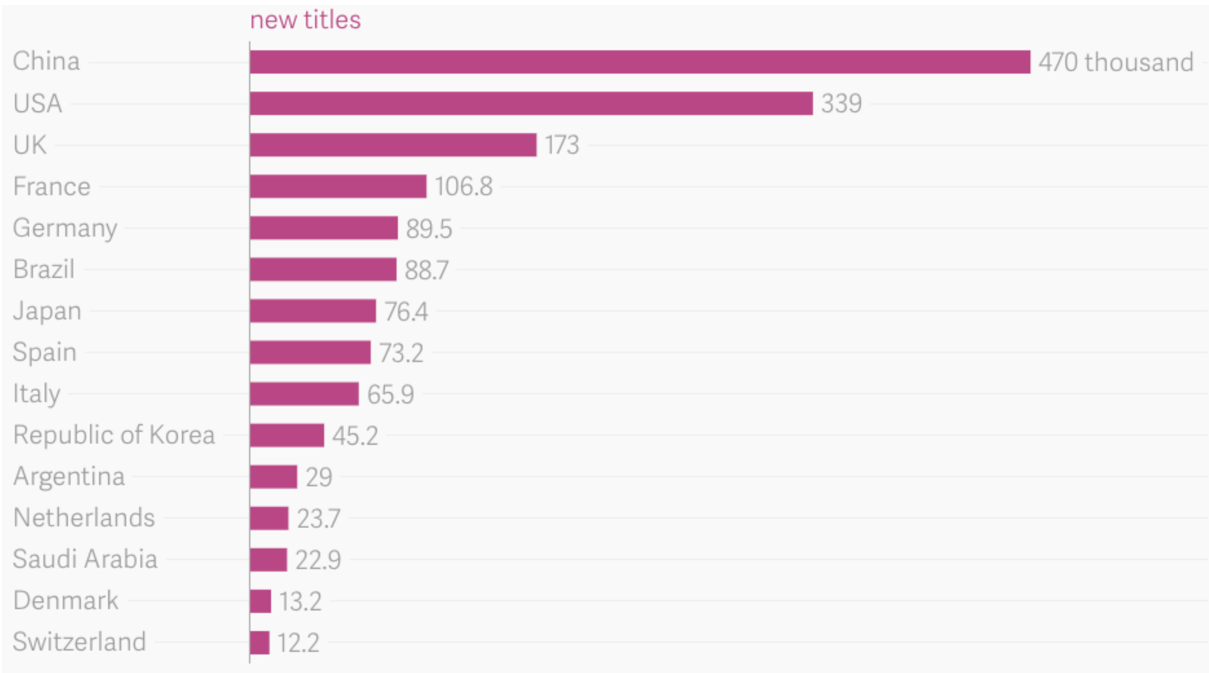


Figure 3. Book production by country in 2015 (Rubin, 2016)

Let us imagine that the aforementioned 33-year-old male reader is a cultural researcher working on a specific subject: *Books published in UK in 2015*. He is so seriously interested in the publications of that particular country in that particular year, that he devotes the rest of his life to do nothing but read them. Following the aforementioned figures, in an ideal scenario, he would have read at the end of his life around 4,000 titles. But, in relation to the total of 173,000 books published in UK in 2015, his readings would have covered only 2,31% of his corpus. This means that he still would not have enough data to consider what he actually read as a representative sample of the subject he intended to study in the first place.

This hypothetical scholar that we have imagined has a very specific object of study (one country, one year). Literary researchers frequently face much more difficult challenges. They often address subjects as ambitious as *the modern English novel* (Josipovici, 1976), *the Latin American short story* (Sayers-Peden, 1983), or *Chinese contemporary poetry* (Yeh, 1991). The corpora of any of these subjects is certainly even more immensurable for the reading capacities of any human than the one we have mentioned. It would be virtually impossible for anybody to read more than 2% of any of these corpora, even if devoting one's whole life to that endeavor.

Due to our biological limitations, we can accelerate the speed of our reading only to a certain extent. And, however fast you are, you're still bound to read one book at the time; so, our readings can only grow geometrically. In contrast, our publications grow exponentially, we publish more books each year, at an increasing speed. According to the last Google measure, our world has produced a total amount of more than 129 million works. How can we study all that? This "slow-reading vs. fast-publishing" paradox is one of the most pressing current difficulties of cultural research. One of the main motives that gave birth to the field of Quantitative Literary Studies (a.k.a. Digital Humanities) was precisely the attempt to work around this paradox. The solution proposed by these scholars was to integrate the quantitative tools of data science into the inquiry of cultural and literary research.[32]

7,000 TITLES: DISTANT READING

For a long time, libraries around the world have been digitalizing parts of their collections in electronic (often online) catalogues. These bibliographies contain plenty of information about different aspects of each book, such as: title, year of publication, language, genre, gender of the author, nationality, publishing house, etc. This kind of information is called *metadata* (s. Jockers, 2013). Superficial as this kind of data is, it can be extremely informative when analyzed properly. For this reason, the analysis of metadata constitutes one of the most elementary examples for illustrating the fundamental advantage that a quantitative approach can provide to literary studies.

As mentioned, literary studies do not typically take more than some tens or (if very ambitious) hundreds of works into consideration. In 2007, literary scholar Franco Moretti – one of the pioneers of quantitative literary studies, co-funder of the Literary Lab at the Stanford University– performed a study on a record of metadata that consisted of 7,000 titles of British novels from 1740 to 1850 (Moretti, 2013).

The first aspect showed by the statistical analysis of this corpus was a dramatic growth in the publication of novels during this period. Moretti concretely hypothesized that, if the production of novels increased, then it would be expectable that some aspects of them would

[32] The power of computer for analyzing large-scale corpora of texts is such, that from the enthusiasm with it the discipline so-called *World Literature Studies* has begun to be finally considered as a plausible endeavor (Moretti, 2013).

have gotten more standardized. He tested this hypothesis by comparing the variance in the length of titles –which is something that would be imperceptible from the perspective of a close-reading analysis–: at the beginning of the period (1740) the length of titles had an enormous variance, whereas in the end (1850) not only the average length of titles had decreased considerably, but also the variance was reduced to a minimal rate (s. Figure 4):

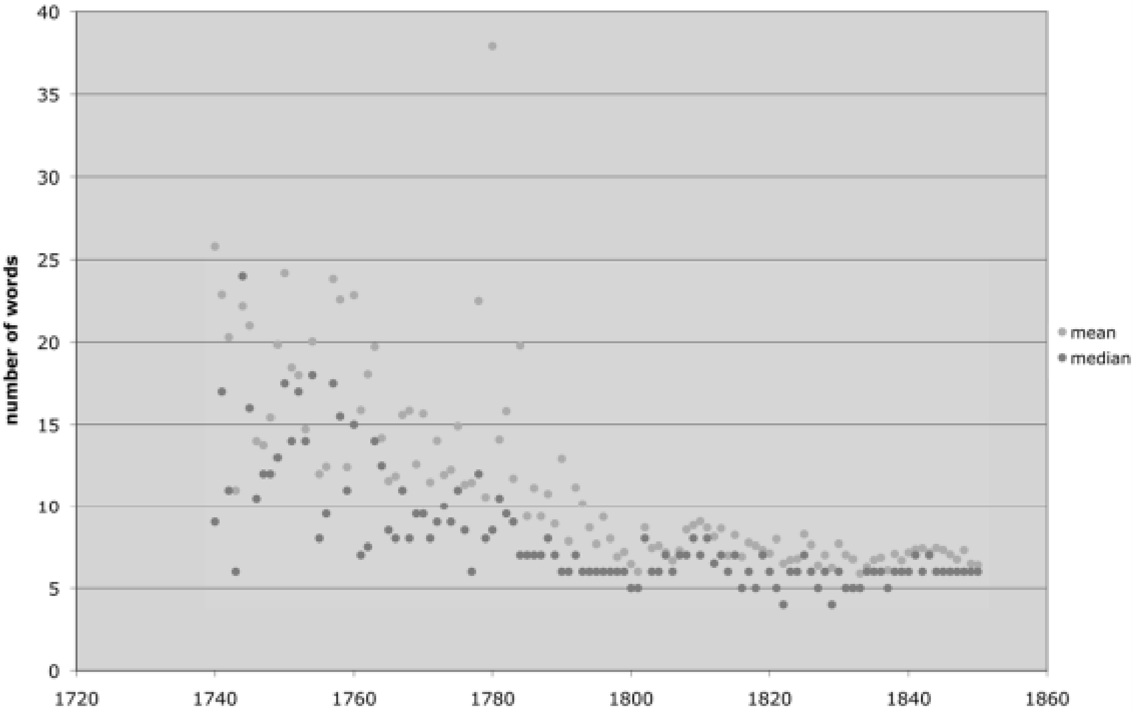


Figure 4. Length of titles per year in British novels (Moretti, 2013: 183)

As the graph clearly shows, with the increase of production, the length of titles effectively got more standardized. Once we are certain of these quantitative results, we can extend the original hypothesis. If a process of standardization actually took place, it would have expectably affected also other features beyond the length of the titles. Following this line, Moretti studied statistically other patterns of the titles in his corpus, and discovered that, effectively, several title-formulas got standardized along with the standardization of the length variance. For example: Titles with article-noun structures turned very frequent by 1850s and became the standard way of indicating an exotic-transgressor subject (*The Faker, The Vampire, The Pirate*); article-adjective-noun titles became standard for familiar-transgressor subjects (*The Unfashionable Wife, The Discarded Daughter, The False Friend*); and titles made of a single female name –without last name– (*Emily, Lucy, Georgina*) became the standard way of presenting unmarried heroines.

Moretti obtained, then, two kinds of quantitative verifications for his hypothesis: firstly, that his hypothesis correctly explained the known data he had; secondly, that it allowed him to predict unknown data. At this point, he advanced a qualitative interpretation: *Why* did these aspects (such as title-length variance and the title-formulas) got standardized in this period? Moretti interprets that long titles were used to describe the content of the novels but, as literary production grew, more and more magazines started to list the latest publications and write reviews and summaries, which turned long titles not only unnecessary but even inconvenient. This qualitative interpretation, in its turn, constitutes a new hypothesis to be furthered verified with quantitative data, inquiring: Was content-description effectively the main use of long titles? Did the publication of novels' summaries in magazines effectively grew at a rate proportional to the standardization of title-length and other features? Etc.

Moretti called this kind of quantitative approach to literature *distant reading*, in opposition to the traditional approach of criticism, known as *close reading*. A close-reading approach allows us to perceive the nuances in a poem or a story. A distant-reading approach allows us to perceive the nuances in hundreds, thousands, or millions of texts through the span of many years. Both are necessary and complementary: the combination of the two can show us the interaction and evolution of local objects in large contexts.

At the same time, quantitative analysis provides us with an epistemological advantage: it allows us to formulate data-driven hypotheses and to test them against representative data. This constitutes a significant epistemological upgrade for cultural research, because it facilitates the production of cumulative knowledge. And this advantage can be observed even by considering data as superficial as book titles –as long as it is analyzed with the adequate tools.

CORRECTING LITERARY HISTORY WITH QUANTITATIVE DATA

A particular field in which quantitative analysis has proven to be crucial is literary history. And, again, this can be also illustrated in its most elementary form by only considering metadata.

Charles Fanning wrote an impressive critical work called *The Irish Voice in America: 250 Years of Irish-American Fiction* (1999). Fanning was a literary scholar who explored the history and evolution of the Irish-American canon using a generational approach: he grouped the authors in successive generations and then studied each one in relation to key historical events of American history.

In the course of his research, Fanning discovered an apparent scarcity of Irish writers in USA from 1900 to 1930. He proposed thereby that that period represents a "lost generation," a period he defined as one of "wholesale cultural amnesia." And he presented a very strong hypothesis to explain this: that a variety of social forces led Irish Americans away from writing about the Irish experience. As quoted by Matthew Jockers, Fanning interpreted that "with the approach of World War I, Irish-Americans ethnic assertiveness became positively unsavory in the eyes of many non-Irish Americans. When the war began in 1914, anti-British feeling surfaced again strongly in Irish-American nationalist circles The War effort as England's ally, and the negative perception of Irish nationalism after the Easter Rising all contributed to a significant dampening of the fires of Irish-American ethnic self-assertion during these years" (Jockers, 2013).

Matthew Jockers –co-founder of the Stanford Literary Lab, along with Moretti– performed in 2013 a quantitative study of Irish-American literature (Ibid.). Whereas Fanning employed around 150 works in his research (a very significant amount for a human reader), Jockers counted with a bibliographic record of 758 works of Irish American prose literature spanning 250 years (a very humble amount for a computer). This corpus had been carefully curated and manually enriched with metadata indicating each author's gender, birthplace, age, and place of residence, as well as more content-oriented information, such as whether the setting of the text was primarily urban or rural. Jockers analyzed these variables in a quantitative way, and attempted thereby to test Fanning's hypotheses.

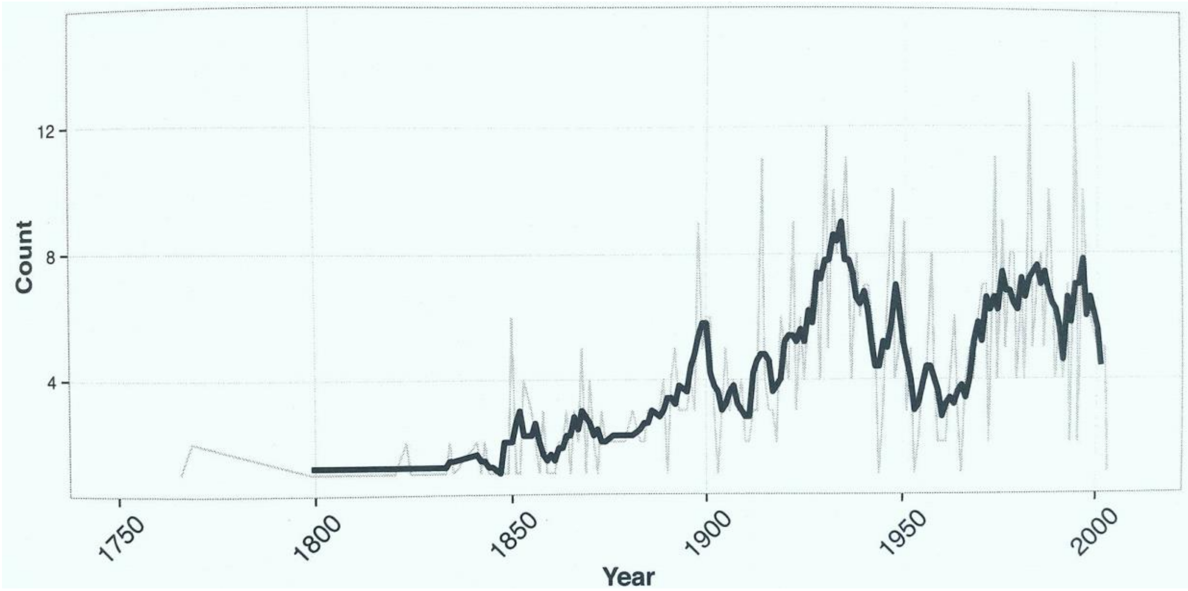


Figure 5. Chronological plotting of Irish American Fiction. (Jockers, 2013)

Only by observing a graph of the Irish American novels published over time, we can see that the literary depression, or "lost generation," that Fanning hypothesized appears to be much shorter than he imagined. A first peak in Irish American publication occurs just at the turn of the 20th Century. This is followed by a short period of decline, but only until 1910. Then the trend shifts immediately upwards, and the number of publications increases dramatically from 1920 to 1930 –growth that encompasses the second half of the exact period that Fanning identified as one when Irish Americans were supposed to have been silenced by cultural and social forces.

One simple graph is enough to refute the intuitive interpretation of Fanning: he had his basic objective data wrong. But a quantitative analysis of this corpus can do much more: it can show the progression of Irish American literature in a more nuanced way, and even discover in what particular way was Fanning wrong, and what his mistake entails.

Jockers analyzed many other variables of the same corpus, distinguishing also between eastern and western authors, males and females, and novels based on urban and rural settings. And he discovered that that Fanning's lost generation was in fact a very particular decrease: one of eastern, male, Irish-American writers with a preference for urban themes. A reader with a bias towards that particular subset of Irish-American writers would effectively find very little in the Irish-American corpus between 1900 and 1940. Indeed, only 5% of the texts published in that period are male, eastern, and urban. But, when the biases are completely reversed (female, western, and rural), we find almost twice as many books. (Ibid.: 46).

Like most studies based exclusively in close reading, Fanning's analysis of Irish American literature is fundamentally canonical and anecdotal. A quantitative study like the one performed by Matthew Jockers provides for these cases a useful corrective and a basis for much more precise, representative, and epistemologically sound analysis. He called this approach *macroanalysis* –evoking an analogy with macroeconomy, as a discipline that studies human change at a large scale.

THE EPISTEMOLOGICAL IMPORTANCE OF QUANTITATIVE ANALYSES

Qualitative analysis tries to answer *why* something occurs; but quantitative analysis is just as necessary for research, because it tells us *if* something occurs in the first place, *to what extent* it occurs, and *how* it is distributed.

In the humanities –in fields such literary and cultural studies– quantitative analyses have been historically very difficult to perform, namely due to the time it took us to manage to learn how to record our production, to the aforementioned disproportion between the amount of books we produce and the amount we can read, and to the lack of technology to analyze such amounts of data statistically. For these among other reasons, the humanities have historically followed a path different than that of the social sciences. They have focused on *criticism* –which consists in the elaboration of creative interpretations of cultural objects– rather than in *description* –i.e. the attempt to discover and map what kind of objects societies produce and how actual people respond to them (s. Schaeffer, 2011).

However, the digital tools for statistical analysis that have been developed by data science (especially in the field of natural language processing, as we'll see in further chapters) are proving to be powerful enough to allow for literary studies to test hypotheses against representative samples and produce thereafter cumulative descriptive knowledge. This has brought up a renewed attention to and expectations towards data-driven descriptive approaches to literature. Indeed, the field of quantitative literary studies has indeed been one of the fastest-growing ones in the humanities in the last decade (Burdick et al., 2016).

The examples of quantitative analysis mentioned in this chapter were meant to illustrate the importance and role of statistical methods in literary research. But these are still only the most elementary examples of what data science can do. In the following chapters, we will further explore the horizon of possibilities that data science gives us to analyze literary phenomena.

CHAPTER 9

TEXT MINING

THE POWER OF COMPUTERS FOR FINDING PATTERNS

Big data is not uniform, clear, and well organized. It is actually quite messy. The data that we cumulate second by second in the world wide web includes content in a variety of formats: images, figures, sounds, text, etc. And it has a further complexity: it is very difficult to analyze because 95% of it is *unstructured data*. This means that this data is not organized in terms of predefined formal models –like the data directly produced by computers– but it is directly produced by humans in their interactions with each other, with all the irregularities and vagueness that this entails (Gandomi&Haider, 2015). Indeed, most of the unstructured data is not other than natural human language –in the form of chats, emails, blogs, voice messages, podcasts, newspapers, and all the literature in the world. Messy as it is, this data is nevertheless a great treasure: This record constitutes the largest cultural production that the humanity, as a whole, has ever created. It is, for all practical purposes, a real version of the infinite Babel's Library that Jorge Luis Borges once dreamt of (Borges, 1998).

Human linguistic production is a very particular kind of data. As said, it is irregular and vague, it holds ambiguity, it is sensible to context and contains a great range of imprecision. For that reason, especial algorithms must be designed to analyze it. The field of data science that is occupied with this task is called Natural Language Processing (NLP). NLP models are constituted by especial algorithms designed to formalize human discourse so as to convert it into discrete and analyzable units of data. These are the algorithms that have enabled us to create all the human-language software that we use everyday: from speech-recognition programs and automatic translators to automatic correctors and even the feat that allows our email accounts to filter undesired messages.

The specific part of NLP that deals with written language is called text mining. Text mining consists in the development of techniques to parse written language at large scales –usually, scales that would not be manageable for human readers. Text mining algorithms allow us, for example, to identify the average frequencies of certain words or kinds of words over time, adjective-noun ratios, usage of punctuation marks, lexical variance, etc. in large corpora of hundreds, thousands, or millions of texts. These methods also help us to visualize the data in graphs and to identify afterwards patterns in it. But what is the actual use of this?

We have seen in the previous chapter the fundamental epistemological function of quantitative analysis for literary research by exemplifying it with the most elementary examples of text mining –considering only metadata. What I would like to explore in this chapter is the richness of these text-mining techniques, the upper limit of the insight they can give us about literature and reader-responses.

HOW TO PREDICT AUTHORSHIP

One concrete field where text-mining techniques are largely and productively used is in authorship attribution. We often confront the problem of finding out who authored a given document. This represents a crucial issue for historiographers, philologists, and historians of literature: Do *The Iliad* and *The Odyssey* really correspond to the same author?[33] Were Shakespeare's writings produced by himself or were they the undercover work of other writers?[34] But it is also crucial for many other disciplines: from religion studies (Who wrote which part of *The Bible*[35]?) to political studies (Who wrote which part of the *United States Declaration of Independence*[36]?) and even for journalism and forensic research (Who wrote those letters?).

The traditional way in which these problems are approached consists in comparing writing styles. Each person has particular writing habits. Each one uses certain words, syntactic structures, topics, etc. in different proportions. For that reason, when we receive a message on the phone, we can often guess who sent it just by paying attention to these kinds of textual features. However, it takes a great investment of time, effort, and expertise for any reader to learn to distinguish the signal of an author, to get familiarized enough with his or her style to become able to recognize it in a novel text. If a new Elizabethan play script was found today, how many people would feel confident enough to decide whether it was written by Shakespeare just by attending to its style? Probably only a specialized minority of English scholars. And, still, even if they felt confident about their judgments, how could we assess their validity? What is their probability of error? How can we measure and verify these things?

Our contemporary panorama offers us alternative methods. In April 2013, a book was published under the name of a previously unheard author: Robert Galbraith. The book was a detective novel called *The Cuckoo's Calling*. A brief biography in the jacket of the book indicated that Robert Galbraith had served the Royal Military Police in the UK and that this was his first novel. It also indicated that Robert Galbraith was a pseudonym. As soon as the novel acquired some popularity, readers started to ask who was really behind the penname of Robert Galbraith. But, out of all the writers in the world, how could anyone guess with certainty who might had authored this book? The mystery, however, did not last long. In July of that same year, everybody knew who was the real author. The arts editor of the Sunday Times had consulted two forensic linguists about the case: Peter Millican (Oxford University) and Patrick Juola (Pittsburgh's Duquesne University). After performing their studies, they claimed that *The Cuckoo's Calling*, the crime story of a war veteran struggling as a private investigator, had been written by the same person that had written the most popular young-adult fantasy saga of our times: *Harry Potter*'s author, J.K. Rowling (Brooks, 2013). It sounded like an unlikely hypothesis, but soon J.K. Rowling recognized that it was correct. The strangest thing is that these researchers arrived at this conclusion without even needing to read themselves any book by J.K. Rowling –actually they didn't even need to read *The Cuckoo's Calling*. How did they do it?

They did it with text-mining techniques: comparing statistics of textual features (word frequency, sentence length, punctuation marks distribution, etc.) (s. Sostek, 2013). The basic functioning of this method can be demonstrated with a simply two authors, two books, and two words. Mathew Jockers provides this example in his book *Macroanalysis* (Jockers, 2013).

---

[33] S., respectively, Homer, *The Iliad* (c. 1260BC-1180BC) and *The Odyssey* (c. 800BC-700BC).
[34] Shapiro, 2010
[35] VV.AA., c. 4000BC-96AD
[36] Jefferson et al., 1776

Consider these two novels: *Sense and Sensibility*, by Jane Austen, and *Moby Dick*, by Herman Melville. These two works are very different. The styles of the authors are very different. But how can a computer *tell* these differences? A computer could actually do it by simply taking two words into account: the pronouns “he” and “she.” In *Sense and Sensibility* (1811), Austen uses in average 136 “she” each 10,000 words. Whereas, in *Moby-Dick* (1851), Melville uses “she” only an average of 4.5 times per 10,000 words. Once a computer has identified this differential pattern, it can learn how predictive this variable is of the authorship of any given page: so, according to the proportion of *she* pronouns, the computer will be able to calculate how likely a novel page is to have been written by one or the other writer –and, if we know this in advance, we can verify the computer’s answer and measure its predictive accuracy. This process is called *machine learning*. And, by means of this process, a computer can learn how to distinguish a writing style faster and more precisely than any human reader.

Nevertheless, real problems in computational authorship attribution are usually more complicated than this one: one often has only fragments of texts, incomplete or ambiguous data, several author candidates, many of whom might have similar styles. In order to sort these cases, researchers must elaborate models that consider more than two pronouns. They must consider several textual features. And this is called working with *high-dimensional* data (s. Jockers, 2013).

In the next graphic, we can see, for example, a visualization of the frequency of three variables across three books by different authors. The variables are the use of the pronouns “it,” “she,” and “he.” The result is a three-dimensional space that shows dots (representing the books) distributed in different positions. The distances among the dots are called Euclidean distances, and they express complex degrees of dissimilarity referring to all the considered variables (in these case, three) at the same time.

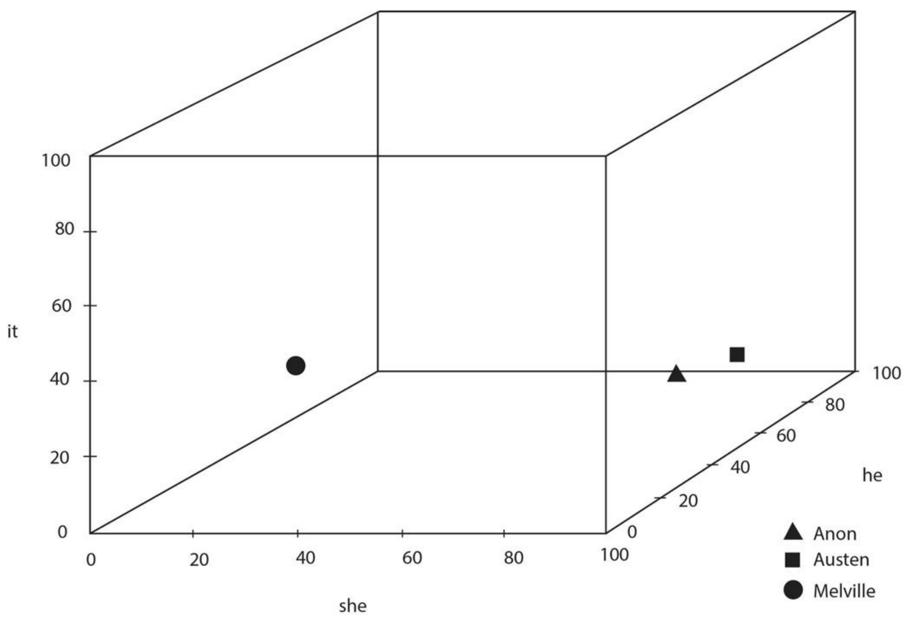


Figure 1. Three-dimensional feature plot with three books (Jockers, 2013)

Naturally, the more variables we add to the model, the more complex and nuanced information we will get regarding the patterns across our corpus. When we aggregate more than three variables, we can no longer visualize the space in a graph, but the computer can still process the data and measure similarities across as many dimensions as necessary. On the basis of this process, we can select the variables that are better predictors of authorship and thereafter *discover* the style of the writer with a measurable certainty –which again, is usually more accurate than that of human readers. Current artificial models of authorship attribution are capable of identifying an author from 90 to 95 per cent of the times (Jockers&Witten, 2010).

Authorship attribution has also several applications outside of literature, such as in the legal, political, and historical realms, for instance (s. Chaski, 2012). Moreover, the quantitative tools that have been developed around this problem have grounded a new and powerful way of studying style in a large sense, beyond mere authorship-attribution: a data-driven approach to style that is today known as *stylometry*. Stylometry has been productively applied not only to literature (Argamon, 2010), but also to other artistic languages, such as music (Westcott, 2006), or painting (Rockmore, 2013).

HOW TO PREDICT OTHER VARIABLES: TIME, INFLUENCE, GENRE, AND GENDER

Once we count with the possibility of creating high-dimensional predictive models (i.e. software that processes complex patterns of textual features), we can apply this method to teach a computer to recognize other kinds of variables beyond mere authorship attribution. We can teach it, for instance, to measure the progression of a style within the work of an author, or even his or his influence in other authors. For example, Plato wrote approximately 30 dialogues. But it is not clear when he wrote which, in which order he wrote them, which ones correspond to his youth and which to his late years. And it is also claimed that the style of Plato appears is other philosophers, as an influence. Text-mining techniques allow us to address this problem by measuring degrees of similarity among different works by statistical means. This way, we can discover to what extent particular stylistic traits are shared among multiple books and assess thereafter the probabilities of particular orders or influences.

The following chart is an example of a study of this kind (Jockers, 2013). It shows the results of analyzing 578 features across 10 novels, and measuring their frequency with regards to the novel *Pride and Prejudice* (1813), by Jane Austen. Once we aggregate the frequency of all these features across the novels, we obtain a rating of similarities among these books.

Rank	Author	Title	Distance
0.	Austen, Jane	Pride and Prejudice	0
1.	Austen, Jane	Sense and Sensibility	0.042557864
2.	Austen, Jane	Mansfield Park	0.049754052
3.	Austen, Jane	Emma	0.050242054
4.	Burney, Sarah	Traits of Nature	0.056073837
5.	Cathcart	Adelaide: A Story of Modern Life	0.057314379
6.	Waddington, Julia	Misrepresentation; or, Scenes in Real Life	0.058382231
7.	D'Arblay, Frances	Cecilia; or Memoirs of an Heiress	0.058646462
8.	Burney, Sarah	Tales of Dancy	0.059090054
9.	Humdrum	Domestic Scenes: A Novel	0.059223492
10.	Lister, Thomas	Herbert Lacy	0.059397822

Figure 2. Difference (in terms of Euclidean distance) from *Pride and Prejudice*, based on 578 features. (Jockers, 2013)

As seen, the algorithm correctly recognized that the books most similar to *Pride and Prejudice* are effectively the other three books in the corpus that were also written by the same author: Jane Austen. This data also tells us *how similar* each book is: *Sense and Sensibility* (1811) and *Mansfield Park* (1814) are more in the style of *Pride and Prejudice* (1813) than *Emma* (1816) is, which fits the order in which they were written (*Sense* and *Masterfield* come immediately before and after, respectively, *Pride*, whereas *Emma* is more distant in time). At the same time, this chart gives us measureable information about the influence (or at least "sharedness" of textual features) of Austen's style in other authors: Burney and Cathcart write more in the style of Austen than Humdrum and Lister.

Moreover, variables can be crossed. We can study several variables at the same and discover, not only which textual feature predicts which, but also which ones are easier or harder to predict from textual features. A study of this kind was performed by comparing 161 textual features (such as frequency of words, average size of sentences, etc.) against 5 different external variables (text, author, genre, decade, and genre). The next graph shows some of the correlations found:

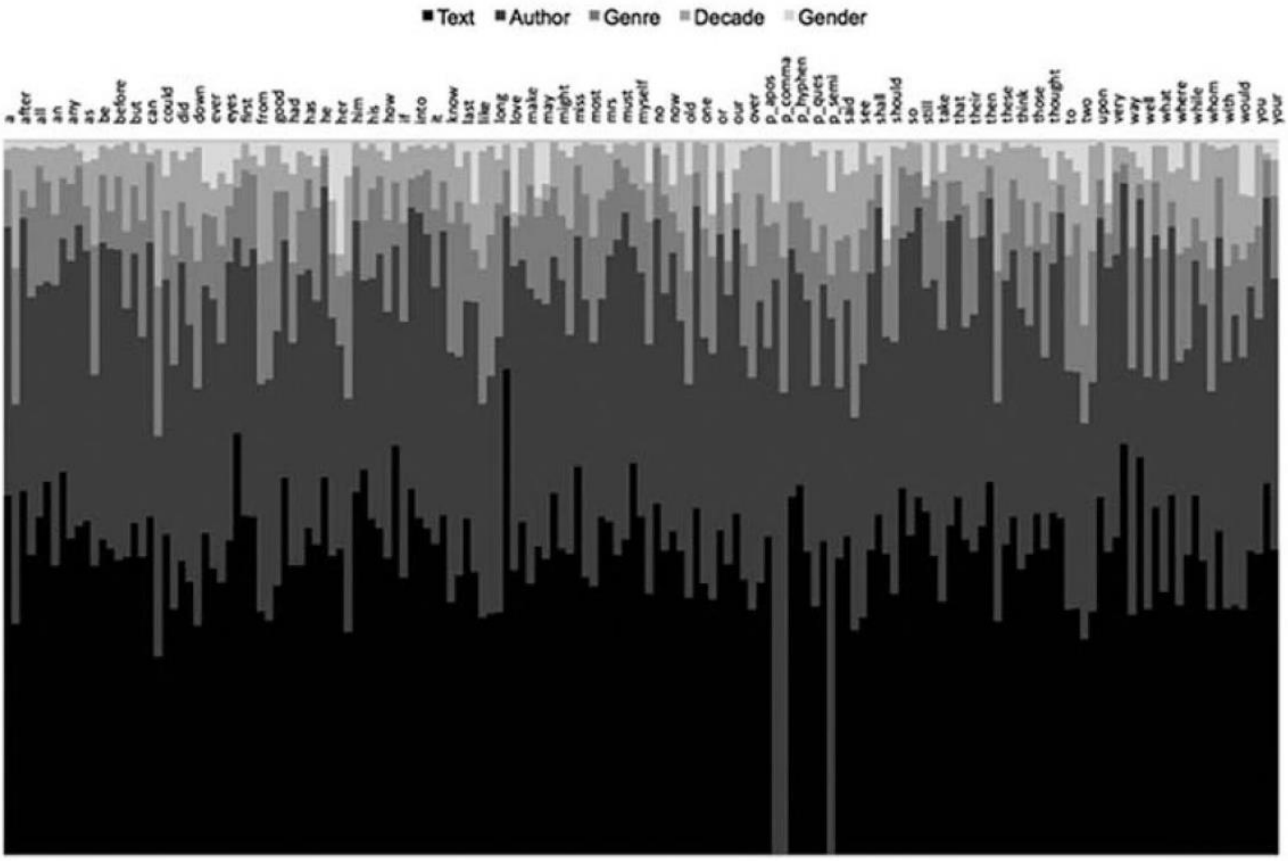


Figure 3. Relative view of category influence. (Jockers, 2013)

This study revealed, first, correlations between particular textual features and particular external variables. For example, it showed that the frequency of prepositions is a good predictor of certain genres (Gothic novels, particularly), whereas the use of pronouns *he* and *she* was particularly efficient for predicting gender of the author. But, at the same time, this study showed that the kind of data provided by textual features is not equally telling of all these external variables. On average, 33% of the textual features had significant correlation with the very text –which means that 33% of the textual features of, for instance, *Moby-Dick* would be effective for predicting whether a page belongs to the novel. 27% were correlated with the author –they would allow you to predict who wrote the book. 18% were correlated with genre, 14% with the decade, and 8% with gender. This opens a series of pertinent questions to be explored: Why author and genre are more evident in the textual features of these novels than decade and gender? What is the dimension of this phenomenon? Is it consistent or it varies across cultures and time? Etc.

In these ways, the combination of NLP processing tools and machine learning techniques are making of the computers instruments of measurements with which we can explore and progressively map different parts of our literary culture.

HOW TO PREDICT BESTSELLERS

But what about reader-responses? One thing is to predict objective external data, such as who wrote a book, when, or where. A different thing is to predict subjective data, such as what impact that book might have in the public. Jodi Archers and Matthew Jockers (A&J) performed a large scale study that approaches us to that question (Archer&Jockers, 2016). They asked a concrete version of it: Can a computer predict whether a book will be a bestseller?

A book being a bestseller only means that it has sold a particular amount of copies. But, beyond that, it is not obvious that there should be anything in common among these books. It is, indeed, very difficult for publishing houses to predict the sales of a book. We can hypothesize that the key of what makes a book a bestseller is only its marketing. After all, in the case of USA, for instance, only the five biggest publishing houses own around 80% of the bestsellers. However, the correlation between marketing-budget and sells is not really as predictive as one would expect: many well-advertised book sell bad and many bestsellers are produced by unknown writers. It seems that there may be many other variables that determine the popularity of a book. At this point, it would seem to make more sense to look for answers in the actual content of the books instead.

A&J formulated the obvious (and yet provocative) hypothesis: Maybe there are effectively *textual traits* that characterize bestsellers. Maybe bestselling novels contain a latent *bestselling signal*. And, even though for us, human readers, this signal is not obvious (not even publishing houses are very accurate at predicting whether a book will become a bestseller by only reading it), maybe a computer could discover it. So A&J undertook the mission of exploring how good a computer could get at predicting bestselling probabilities by only considering textual features. The results were surprising.

For their study, A&J took a particular set of bestsellers (BS): the ones listed in the ranking of BS *of The New York Times* in the last 30 years. They included these 500 novels in larger corpus, among 4,500 other random novels published in English in the same period. More than 20,000 textual features were measured in their proportional frequency in BS and non-BS. And 2,000 of them proved to have predictive power regarding bestselling. Each feature, in itself, had only a predictive accuracy above 50% of accuracy (that is, only better than chance). But, when aggregated in a high-dimensional matrix, the final model had a predictive accuracy higher than that of most human readers. After learning the bestselling patterns, the computer was offered a novel and asked to predict bestselling probabilities. It was correct more than 85% of the times. And it did not only predict whether a novel had been a bestseller, but it even predicted an approximation of *how much* it had sold. They called it *The Bestseller-ometer*.

The Bestseller-ometer chose the following as the top five bestselling novels (that is, as the ones that most typically represent the traits that predict bestselling probabilities):

1. Dave Eggers, *The Circle* (2013)
2. Jodi Picoult, *House of Rules* (2010)
3. Maria Semple, *Where'd You Go, Bernadette* (2012)
4. Michael Connelly, *The Burning Room* (2014)
5. David Baldaci, *The Hit* (2013)

Some of the traits analyzed by A&J correspond to the aspects of literature that we have discussed in the previous part of this book. It is worth considering in detail the kind of text-mining techniques they used for digitally measuring each of them. Namely: meaning, narrative and emotion, and characters.

WORD-MEANING: TOPIC MODELING

Some topics are more frequent than others. There are, for example, different proportions of war novels and travel novels, of cooking. or travelling-to-space novels. So, it would be expectable that some topics are more frequent than others across bestsellers. In other words, measuring the topics of novels could be pertinent for identifying a bestselling signal.

But how can a computer become capable of recognizing the topic of which a certain text is talking about? A first approach would be for a computer to search for words by frequency. If in a text the word *space* appears with frequency higher than its average in regular language, then we are likely to be in front of text about space.

However, words are ambiguous. When we read the word *bar*, for instance, we immediately interpret, in accordance to the context, whether it refers to a place to go for a drink, to a long piece of meta, or to a lawyer's exam. How can the computer work around these cases? The solution is called topic modeling.

Topic modeling is based on a simple fact: whenever a keyword is used in a particular sense, certain other words become likely to appear in its linguistic context, on the basis of which the meaning of the keyword can be predicted. As mentioned in the second chapter, our brain has also the ability to track these contextual probabilities. And this can also be calculated by a computer. So, basically, when finding a word like *bar*, a computer would measure what other words are likely to appear close to it. When words like beer, bartender, and table appear around *bar*, the computer will know it has found a *bar-to-have-drink* topic.

Figure 4. *Bar-to-drink* topic, word cloud (Ibid.)

But when words like *metal*, *long*, etc. appear around *bar*, the computer will categorize it as *material-bar*, for instance. And if it finds *bar* surrounded by words like *law*, *university*, *approve*, the computer will categorize it as *lawyer-exam-bar*.

So, A&J made the computer identify patterned clusters of words and they labeled around 100 topics. And they measured afterwards the frequency and proportion with which these topics appeared across the 5,000 novels of their corpus. Figure 6 shows an example of the result of the topic analysis of one of them:

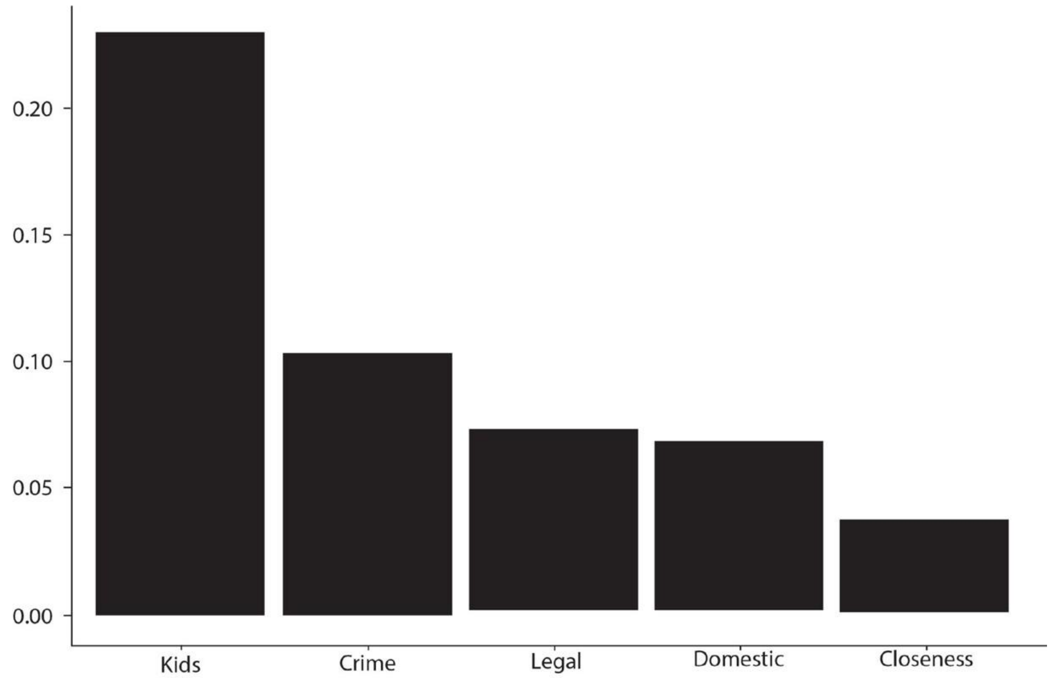


Figure 5. Top-five topics in Jodi Picoult's House of Rules. (Ibid.)

House of Rules, by Jodi Picoult (2010), contains two dominant topics (kids and crime), two secondary ones (legal and domestic), and a marginal topic (closeness).

The next step was to identify which topics –and in which distribution- were more typical of bestsellers. And they discovered a very suggestive pattern of topics. Here are some of the results.

Firstly, the model distinguished the typically bestselling and non-bestselling topics. According to the model, non-bestselling topics included all fantastic creatures and settings, made-up languages, and space adventures. Bestselling topics revealed to be more typically realistic: marriage, death, taxes, technologies (preferably modern and vaguely threatening), funerals, guns, doctors, work, schools, presidents, newspapers, kids, moms, and the media. But, among the realistic topics, there were also many bad-selling ones, according to the statistics: the body described in any terms other than in pain or at a crime scene, cigarettes and alcohol, the gods, big emotions (like passionate love and desperate grief), revolutions, wheeling and dealing, existential or philosophical sojourns, dinner parties, playing cards, very dressed up women, dancing, and (unexpectedly) sex, drugs, and rock&roll. Also a pattern of preference for particular settings was identified: a town or a city (any) rather than outer space, the desert, the ocean, the jungle, or ranches. An exception to this tendency was *The Martian*. "The reading public prefers to see the stock market described more so than the human face. It likes a laboratory over a church, spirituality over religion, college more than partying, and dogs more than cats." (A&J: 48).

Moreover, the model predicted some particular combinations of topics as typical of bestsellers. Namely, combinations that seem as suitable for favoring potential crises: kids and crime, family and disease, marriage and funerals, etc. These results fits people's appeal for conflict in literature, idea that we explored in Chapter 4, in relation to the evolutionary hypothesis of stories as simulators for life problems.

Thirdly the model predicted topical proportions. In words of A&J: "Bestselling authors give 30% to just one or two topics, whereas non-bestselling writers try to squeeze more than three topics in 30% of the book. To get to a 40% of the average novel, a bestseller uses no more than four topics, whereas a non-bestseller uses on average six." (Archer&Jockers, 2016).

Also considering words frequency, independently of topic, can be telling of the style of an author. A&J measured the proportions in which different authors of their corpus used the 491 most frequent words (mostly filler words such as articles, prepositions, connectors: the, a, that, etc.) and punctuation marks, and tried to find in these traits a bestselling signal: on that basis the computer was able to guess bestselling books 70% of the time. In general, the results showed that bestsellers tend to have more of the most usual words –e.g. the word "do" appears twice more often, "very" half more often– and they tend to keep language simpler and more informal –e.g. they use more contractions, less adjectives, less adverbs, etc.

When these stylistic data was correlated with gender of the authors, the model calculated that that women have, in average, a more bestselling style than men. Actually, with only these stylistic data, gender could be predicted with 70% of accuracy –models with further data achieve accuracies higher than 80% (Jockers&Witten, 2010).

NARRATIVE: SENTIMENT ANALYSIS

The way in which authors design their story-plots could also be a variable that influences in the bestselling probabilities of a book. But narrative structures in themselves are very difficult to measure, because they do not manifest in the surface of words choices and order, but in the deep structure of the meaning of the discourse. How can a computer retrieve a signal of the plot? A&J used an indirect strategy: measuring emotions.

Each up and down of a plot typically correlates with an emotional effect: an improvement or a fall of the hero, a rapprochement to or a distancing from his or her goal. So, if there was a way of mining the emotional progression of a story, that would be likely to reflect the progression of the narrative structure. Natural language processing offers a technique to target the expression of emotions, which is called *sentiment analysis*. Texts that people interpret as emotionally positive have, in average, a higher proportion of certain words (e.g. happy, achievement, amazing), and negatively judged texts have higher proportions of other words (e.g. disgust, die, sad). On this basis, by measuring proportion of positive and negative words, the emotional valence of a text can be predicted –This is, by the way, how computers can automatically sort reviews and rank the popularity of different products. This method can also allow us to measure the emotional progression of a novel, by calculating how the emotional valence of language changes along the text –which, expectedly, expresses the development of the plot.

After analyzing the emotional progression of each of the 5,000 novels, the computer searched for the patterns of emotional progression that were more typical of the bestsellers. Here are the three most relevant ones.

Regular beat. Bestselling plots typically show emotional curves with a regular rhythm, a steady beat of ups and downs, of clear progressions of improvement and progressions of decadence –instead of other options, like a sustained monotony, a single continued rise of fall, or a frenetic mix of ups and downs. Compare, in Figure 5, the rhythmic progression of James's bestselling novel *Fifty Shades of Grey* (2011) in contrast with the non-bestselling novels by Liliana Rhodes (2015) and Krista Ritchie (2013).

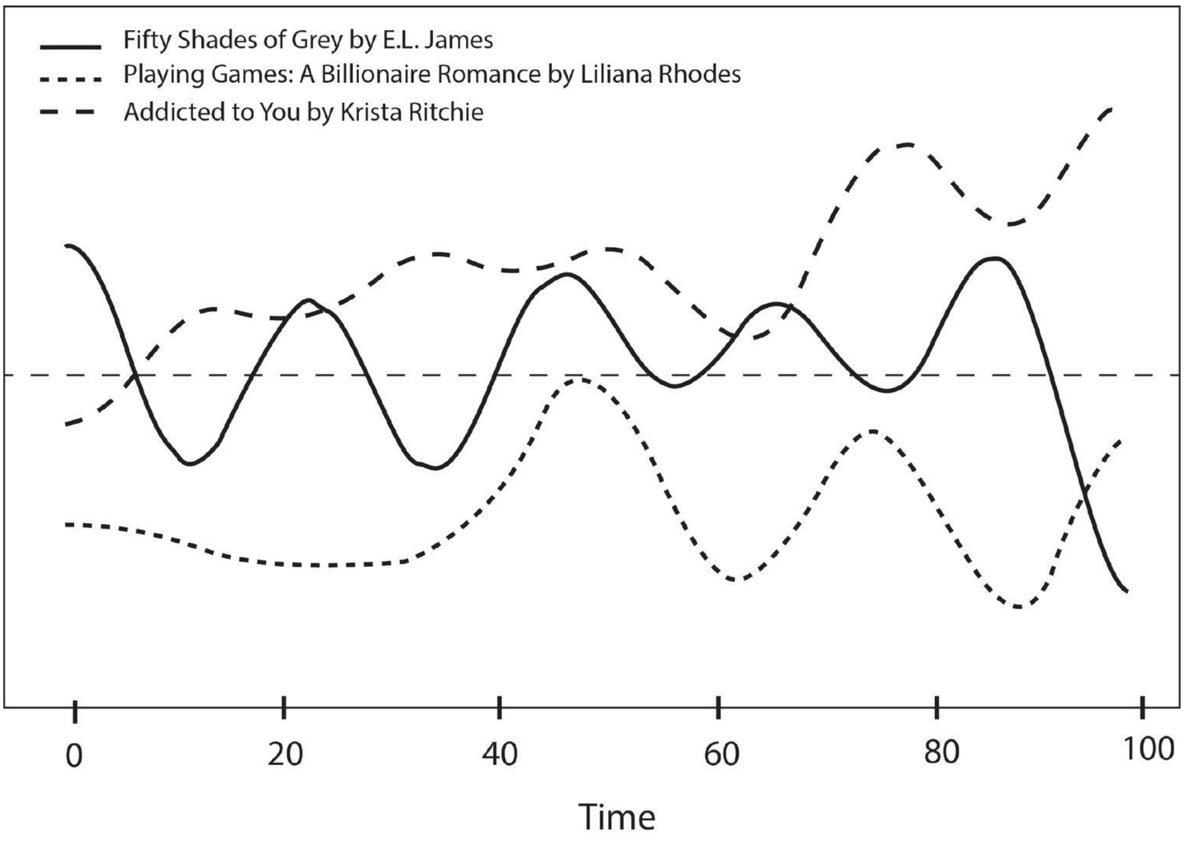


Figure 6. Emotional plot progression of three novels. (Ibid.)

Three acts. When visualized with low granularity, bestselling plots typically show a marked three acts structure. This occurs with independence of the directions in the curve, which leaves space for variability: some go up-down-up, some down-up-down, some down-down-up, etc. (7 typical plot models were identified by A&J by considering the combinatorial possibilities of these structures), but there must be three recognizable processes of improvement or decadence.

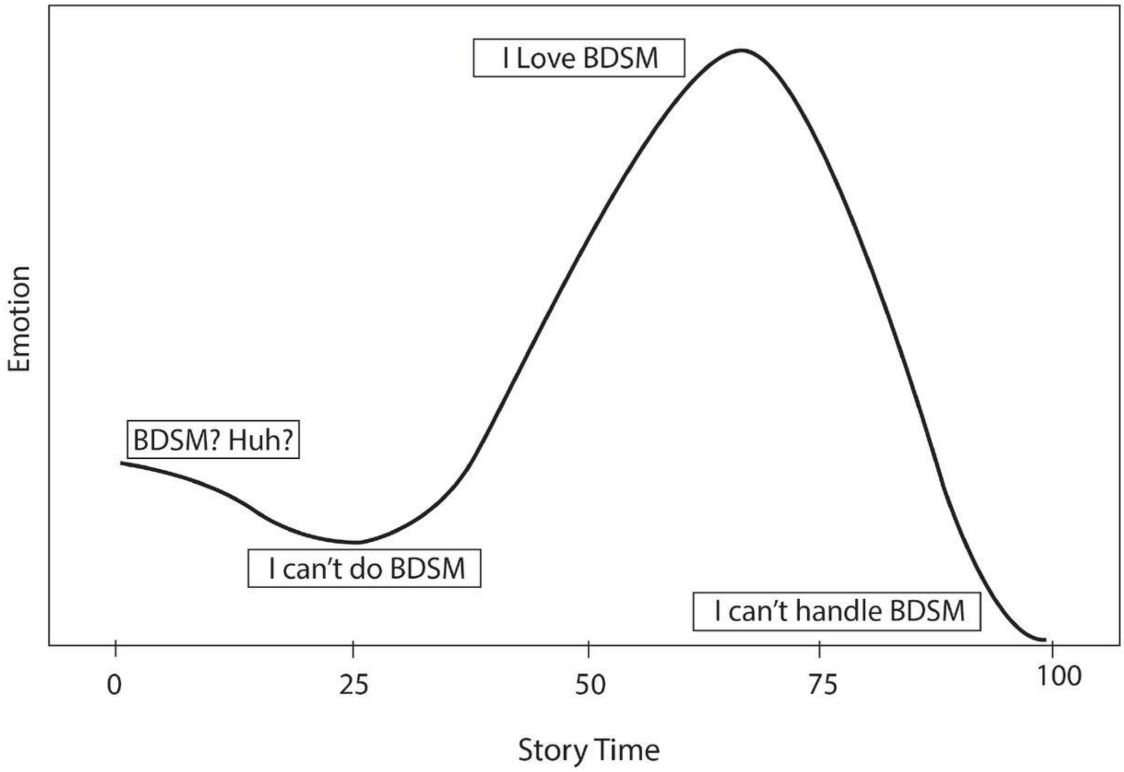


Figure 7. *Fifty Shades of grey*'s emotional plot progression in three acts. (Ibid.)

First hook. Bestselling novels typically show a pronounced curve in the first third (even more precisely, not before the 20th page and not after 50th), also independently of the direction (it can be either a peak or a valley, a positive or a negative progression). Narratologically, this means that a bestselling novel starts by preparing the ground for a proximate turning point (*Wendepunkt*). This can be seen in Figures 6 and 7.

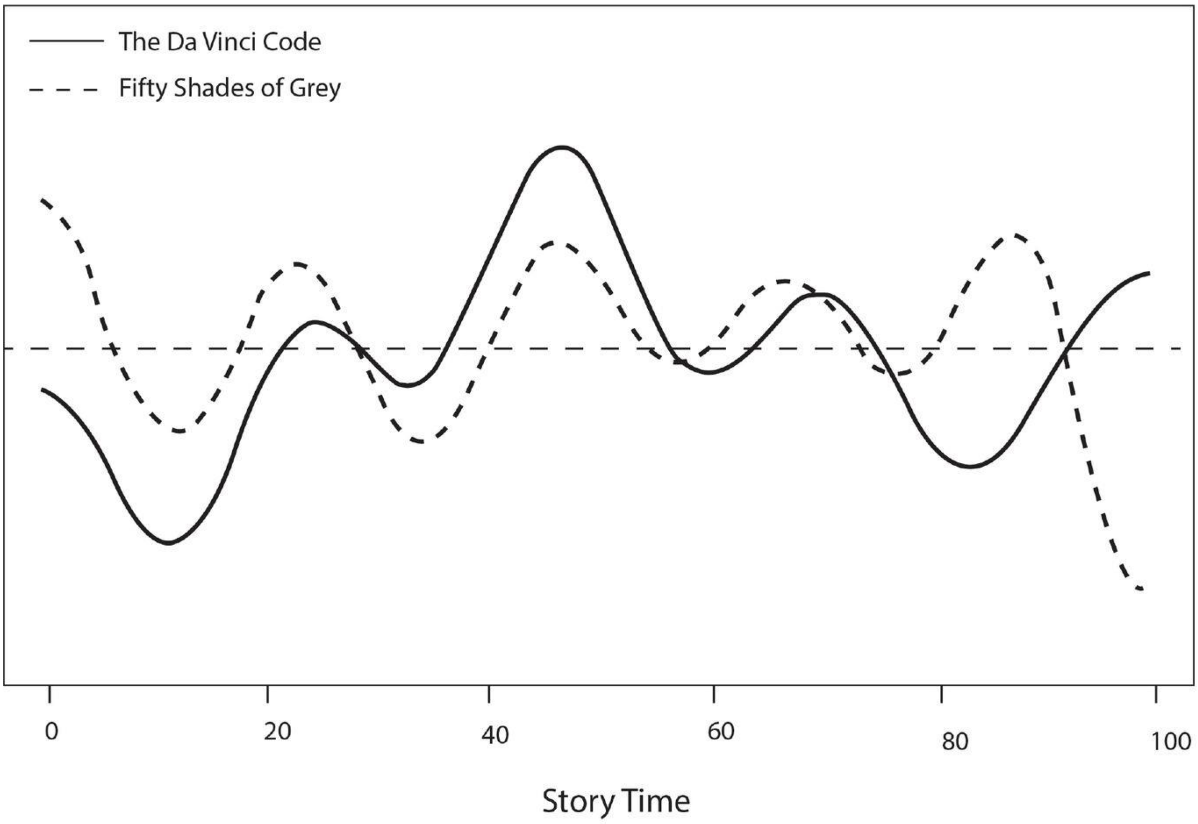


Figure 8. Sentiment Progression of *The Da Vinci Code* and *Fifty Shades of Grey*. (Ibid.)

The novel that the model identified as having the most bestselling emotional progression was *Fifty Shades of Grey*, indeed one of the most bestselling novels of the century. We can see in it all the features that characterize bestselling books: the regular rhythm, the three-act structure, and the first pronounced curve in the beginning. The second novel that the computer identified as better fitting the bestselling pattern was Dan Brown's *The Da Vinci Code* (2003), which effectively has a remarkably similar plot-structure to *Fifty Shades of Grey* and, like that novel, it is also one of the most bestselling books of our times. It is incredible to think that a computer, when asked to guess what are the books with the highest bestselling probabilities, successfully points out to two of the highest bestselling works of our times. And this prediction, which is so hard for a human reader to make, was done by the computer by only measuring patterns in the proportions of positive and negative words.

CHARACTERS: VERB MINING

What characters are more typical of bestsellers? How can we track information about characters across novels? A way to do it would be to consider the kinds of words that appear in the text more frequently associated with the names of the characters. And, of these words, perhaps the most significant ones are verbs –inasmuch as a character can be defined by what he or she *does* in relation to his or her goals.[37] This is what A&J did. They analyzed the kinds of verbs most usually associated to the main characters of bestselling and non-bestselling novels.

[37] In structural semantics, for instance, characters are described as superficial instantiations of the deeper goal-directed roles that structure the story, called *actants* (e.g.: hero, villain, object, helper, sender, etc.). S. Greimas, 1966.

Bestselling (BS) characters, mostly, need, want, grab, do, know, control, think, ask, look, hold, love, tell, like, see, hear, smile, reach, pull, push, start, work, and arrive. Whereas non-bestselling characters halt, drop, demand, seem, wait, interrupt, shout, fling, whirl, thrust, murmur, protest, hesitate, accept, dislike, suppose, recover, and wish.

This distribution of verbs seems to suggest different stereotypes of characters. BS characters seem to be straightforward, with a strong direction, will, capacity, surety, that live their life and make things happen, with self-awareness and self-knowledge, that own themselves. Whereas non-BS characters are victims of the circumstances, as if the world would create them rather than the other way round, they seem less open, and slightly more negative.

Only by considering the usage of some of the most basic verbs, the machine could already predict with 72% of accuracy the likeability of a novel of belonging to the bestselling group.

Once they had these results, A&J decided to introduce other variables. They discriminated the results in terms of the gender of the characters. And the differences showed also a very clear and impressively suggestive pattern. In bestselling novels, men do most of the kissing, whereas women do more of the hugging. Men fly, drive, kill more than women do. Women talk, read, and imagine more often than men. He travels, she stays. He assumes, she decides. He promises, she believes. They both love but she is more likely to hate. Both characters see, but he also stares (often at her). She screams and shoves; he worries and punches. These results seem to show that bestsellers –at least to the extent represented by this corpus– prefer traditional gender roles, even stereotypical ones.

Finally, A&J also discriminated the verbs of BS and non-BS in relation to the body and in dialogues.

In relation to the body, BS-characters have more simple and controlled gestures: they eat, nod, open, close, say, sleep, type, watch, turn, run, shoot, kiss, die. The ones that die and survive are not necessarily the protagonists, but the protagonist is someone often doing something as dramatic as surviving or dying, or dealing with that, not yawning. The distribution of verbs in relation to the body is coherent with the portrait of BS-characters as agentive and active.

As for the dialogues, BS-heroes and heroines do not: begin, speak, accept, remark, explain, mutter, answer, protest, address, shout, and demand. 90% of the time they only "say." Any alternative to express "he said" with words other than "he said" is non-bestselling - the only exception being "he asked."

A possible interpretation for phenomenon is that, when writing direct speech, the BS-dialogue tags are frequently almost silent to the reader's ear, as unremarkable as the word "said", so as not to distract the attention from the words inside the quotation mark, which is supposed to be the important information to characterize the hero. "Say" and "ask" demand less attention to be understood than "demand" and "exclaim."

WHAT DO THESE PREDICTIONS MEAN AND WHY DO THEY MATTER?

What Archer&Jocker's Bestsellerometer gives us is quantitative data: highly processed and highly predictive quantitative data. All of this is empirical evidence for us to elaborate hypotheses or to contrast the hypotheses that we have.

Digital humanities, as its name indicates, is composed by the conjunction of two realms: digital tools and humanities research. The relationship between the two (digital and humanities) can be accounted for by a couple of key concepts: data and information. *Data* refers to unorganized collections of elements, raw (but measurable) material, which can be analyzed with the aid of computers. This is the *digital* level. Now, data has no inherent semantics, it means nothing in itself. So as to make it meaningful, valuable, we have to develop an interpretation of it, and that is how we transform it into *information*. Hence, the *humanities* level.

As we have seen in the examples of quantitative analysis, data-science predictions are based on calculating correlations among variables: v.g., length of titles over time, gender and number of publications, stylistic traits and authorship, textual traits and sales number. But in order to make these correlations informative, we must frame them within an explanatory hypothesis that accounts for it. This step still constitutes an interpretation and requires therefore the activity a creative mind. What quantitative tools give us is the possibility to increase our measuring capacities beyond our senses, providing us thereafter with empirical evidence at an unseen proportion in human history, against which our hypotheses can be tested and extended.

CHAPTER 10

ROBOTS THAT READ AND WRITE

FRONTIERS OF LANGUAGE TECHNOLOGY

AUTOMATED JOURNALISM

> **Earthquake: 4.7 quake strikes Los Angeles, centered near Westwood**
>
> A shallow, magnitude-4.7 earthquake was reported Monday morning five miles from Westwood, according to the U.S. Geological Survey (USGS). The temblor occurred at 6:25 a.m. PDTat a depth of 5.0 miles.
> According to the USGS, the epicenter was six miles from Beverly Hills, seven miles from Universal City and seven miles from Santa Monica.
> In the past 10 days, there have been no earthquakes of magnitude 3.0 or greater centered nearby.
> (*Los Angeles Times*, 17th March 2014, 6:33 a.m.)

The earthquake started at 6:25 a.m. The sensors of the USGS detected it at 6:27. Only a few seconds later, *Los Angeles Time* had already this article written, and an email was automatically sent to developer Ken Schwencke. Awaken by the message, Schwencke glanced over the article, posted it with one simple click at 6:33 a.m., and continued sleeping (Meyer, 2014). The article appeared in the website signed by Schwencke, but he had not written it. Who had? And how could it do it in only some seconds? The mystery was revealed in the footnote:

> This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.

The algorithm created by Ken Schwenke is called Quakebot, and it automatically generates in seconds journalistic articles of this kind by processing numerical data that it receives as input (such as the earthquake data sent by the USGS), organizing it according to pre-programmed text structures, and accommodating thereafter details, figures, and often even graphs (Mayer, 2014). The output can even be preprogrammed to fit different kinds of tones, styles, or degrees of formality.

Companies such as Narrative Science, Automated Insights, and Yseop have been developing these kinds of software for more than a decade now, opening the possibility for the practice today known as automated journalism, algorithmic journalism, or robot journalism (s. Graefe, 2016; Carlson, 2015). Nowadays, many media companies across the world use them regularly –*Forbes*, *Associated Press*, and *Los Angeles Times* count among the early adopters (Montal&Reich, 2016). Typical articles that are often written by automatic-text generators include sports recaps, weather and financial reports, real state analyses, and earnings reviews (Ibid.). In fact, a significant part of the last Football World Cup (Russia, 2018) was covered by technologies of this kind (Pottala, 2018).

The use of these technologies has increased enormously the volume and speed of news production. It has also raised some controversy. Yet, most journalists do not see it as a threat, but as an aid that is freeing them from a great amount of time-consuming routine tasks. In words of Associated Press's strategy manager Francesco Marconi:

> We went from producing about 300 stories to close to 4,000 each quarter, which was a 12x increase in content output (...) We also saw a reduction in error rate and were able to free up 20% more of reporters' time to focus on higher-value projects.
> (Liyakasa, 2018)

The case of automated journalism is one of many fields in which reading and writing robots are finding their place.

Inquiry on how to create software that behaves similarly to the human mind necessarily leads to inquiry on the actual software of the human mind. Therefore, analyzing these bots can reveal interesting insight –in their successes as well as in their failures– about our own ways of processing information. These verbal technologies are proving to have increasing usages in unexpected fields, and one of these fields is precisely literature. I would like to explore in this chapter some of the things these verbal bots can reveal about our own minds, our literature, and how they can be used as tools for cultural research.

PAPER GENERATORS

Many of these verbal bots were originally created for pure entertainment. But, increasingly, often unexpected practical applications have afterwards been discovered. This is certainly the case of the so-called paper generators –which have existed for more than ten years now–, such as the Postmodernism Generator and the SCIgen. These programs recombine words, structures, and stylistic traits that are frequent among humanistic and scientific academic articles. Using generative algorithms, they create thereafter academic-resembling gibberish articles. The two afore-mentioned programs are available online and open for public use.

Here is an excerpt of a paper produced by the Postmodernism Generator:

> **Deconstructing Constructivism: Socialist realism in the works of Mapplethorpe**
>
> If one examines neodialectic Marxism, one is faced with a choice: either accept cultural theory or conclude that culture may be used to reinforce capitalism. But the subject is contextualized into a Baudrillardist simulacrum that includes narrativity as a paradox. Many desituationisms concerting socialist realism exist.
> (Postmodernism generator, 2018)

Here is an excerpt of a paper generated by the SCIgen:

> **The Impact of Lossless Modalities on Electrical Engineering**
>
> The exploration of object-oriented languages has visualized systems, and current trends suggest that the understanding of Moore's Law will soon emerge. Given the current status of

psychoacoustic communication, system administrators particularly desire the analysis of vacuum tubes, which embodies the unproven principles of e-voting technology. In order to surmount this grand challenge, we demonstrate that even though the infamous peer-to-peer algorithm for the construction of gigabit switches by Sun and Miller is NP-complete, operating systems can be made relational, efficient, and event-driven.
(SCIgen, 2018)

Each of these programs produces a full new article every time we refresh the website. And the articles really *look* like academic articles, even in their layout: they count with automatically generated titles, subtitles, author specifications, abstracts, and even footnotes and bibliographical references, in the expected proportions and distributions. And each phrase of these texts, individually, seems to make some sense, even if the texts, as a whole, are completely absurd. Due to the effectiveness and absurdity of their imitations of academic style, the paper generators became immediately popular and were widely used for amusement. However, a more serious application became soon evident.

An article produced with the SCIgen was submitted in 2005 to the World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI), and it was accepted by the Organizing Committee. The article was entitled: "A Methodology for the Typical Unification of Access Points and Redundancy" –and it can still be found online along with the acceptance email from WMSCI (s. SCIgen - About). As explained in their website, the members of SCIgen sent this article as a deliberate attempt to embarrass WMSCI, which according to them accepted low-quality papers. The strategy produced the expected effect: The scandal generated by the revelation of this case forced WMSCI –and, then, also other institutions– to revise and correct their paper-admission protocols.

Ever since, the scientific community has started to use paper generators like the aforementioned ones on a regular basis as tools for testing the reviewing procedures of institutions of suspected reputation. This way, out of a technology originally created for pure amusement, a bottom-up strategy emerged to help regulate quality standards within the academic community.

Now, the style of journalistic and academic articles might not be the most complex for imitation, but what about literature, the epitome of verbal creativity? Could a computer write any kind of literary-resembling text whatsoever?

***TRUELOVE.WRT*, BY PC WRITER 1.0**

"There's nothing else here but the bloody sea and the bloody rocks… And it is in such a drab place that I am going to kill you," the woman muttered.

This is the opening line of the first popularly recognized computer-generated novel: *TrueLove.wrt*. It was published in 2008 with the (also automatically generated) subtitle: "An impeccable novel." It was produced by a program called PC Writer 1.0, which was created by a group of Russian software developers and philologists working for the publishing house Astrel-SPb. The developers worked for eight months to create the code, feeding it mainly with Russian novels (s. Vitaliev, 2009). Once finished, PC Writer generated in three days the 285-pages novel entitled *TrueLove.wrt*.

In comparison to regular human-made novels, *TrueLove.wrt* is, of course, very strange and definitely not very pleasant to read straightforwardly from the first page to the last. All the characters speak in an undistinguishable fashion and the action described in it is unexciting and often simply confusing. However, unlike the aforementioned paper generators, *TrueLove.wrt* is far from being completely absurd. The book does not only capture many linguistic features that give it a recognizable literary tone, but it also manages to create an effect of narrative consistency and continuity between the sentences, as it can be appreciated in the closing paragraph of the novel:

> After that, he sat on the wet sand, so close to the water that the waves – heavy and clumsy like pregnant seals – were almost touching his feet. The setting sun was painting pink the underbellies of the clouds hanging low above the grey sea. White caps could be seen here and there, but it was obvious that the storm he had been expecting all day was not going to happen.

TrueLove.wrt has also a stable set of characters and even a relatively intelligible plot – which Russian critic Vitali Vitaliev summarizes in the following way:

> … the characters find themselves on an uninhabited island. All of them have amnesia. They know who they are, but don't remember if they are married or have children, and what relationship they have with each other. They are given a chance to build their relationships anew.
> (Vitaliev, 2009).

PC Writer 1.0 did not become a famous prolific author. Still, many seem to ambition or fear that writing bots end up *replacing* human authors. However, there might be a conceptual –more than a technological– obstacle for this: readers do value the assumption that what they are reading is the expression of a conscious and sentient being like themselves. People are not only interested in literary, musical, or cinematographic works, but also in the lives and personalities of artists, musicians, and filmmakers behind these works. And this applies not only to art, but to many other fields: Olympic weightlifters have not been replaced by tow trucks, even though any tow truck can lift much more weight than any human; marathon runners were not replaced, in their sport, by faster vehicles; and self-playing pianos have not reduced the public of human pianists, even if they often play much better.

There are some things whose value derives precisely from the fact that they are human-made. And art is, in many senses, one of these fields. In consequence, just as people are usually not interested in watching competitions of chess-bots playing against each other, we could guess that they would not be very interested in computer-generated literature either… Nevertheless, certain historical examples might suggest otherwise.

THE HISTORY OF TEXT GENERATORS

We perceive artworks as the intentional products of conscious beings. We imagine, whenever we observe a work of art: *How was the person who did this? What did he or she intend to express?* Etc. But, for this very reason, entrusting part of the creative act to the action of a rule, method, or device relatively independent from the author, or even to pure chance, can have an unexpected interest. Many historical examples are evidence of this. In New York, in the 1940s and 50s, *action painting* was created, which consists in spontaneously dripping and splashing paint onto the canvas, so that the resulting artwork is, in consequence, not completely controlled by the artist, but also determined by the unpredictable ways in which the painting falls and spreads (Gersh-Nesic, 2006). Electronic musicians often activate multiple music tracks simultaneously, so that they automatically blend as a result in unexpected ways (s. *sampling* –Lott, 2013). Supplanting particular parts of the creative process with relatively autonomous creative protocols has been often sought with interest in many artistic disciplines. And this has also occurred in literature. People has historically invented different kinds of text generators: rules, methods, and even machines that create texts independently of the control of the author –Computer-generated literature is, in this sense, just a new way of playing this old game.

In the 20th Century, many avant-gardists writers (from the Surrealists to the Beatniks) invented and employed many text-generation methods in their work. One of the most famous of them is the so-called *Exquisite Corpse* (EC) (s. Breton, 1948). The EC can be used for different arts, but its original format was literary. In it, a person would write a sentence in a paper, fold it in a way that only the last part is visible, and hand it to a second person. This second person would repeat the procedure and hand it to a third one. The mechanism can be indefinitely iterated. The final result is a text (typically presented as a surrealist poem) that is only partly authored by the participants, since the general plan is subtracted to their control and entrusted to the random generative mechanism of the EC's rules. The absurd name of Exquisite Cops comes, in fact, from a phrase contained in the text written by André Breton, Robert Desnos, and Tristan Tzara –the creators of the EC– the first time they practiced it: "The exquisite corps will drink the new wine" ("Le cadaver exquis boira le vin nouveau") (Ibid.).

But text generators are actually much older. At least since the Renaissance, people have practiced *bibliomancy*, which consists in the use of books for divination or magical medicine (Kelly, 2011). The practices of bibliomancy recruit different methods for doing this, that typically involve predefined sets of rules for guiding the reader into extracting particular fragments from an (often sacred) book, so as to constitute a sort of custom-made message to be interpreted in relation to a previously formulated question. A popular example of this, that was already used by the Ancient Romans, consists in posing a random question to the universe, then taking a text in a random place, choosing a sentence by pointing at it with the eyes closed, and then take it as the answer (Hayes, 1997). The Romans typically used Virgil's works for this, but many people still resort to this method nowadays, normally employing religious texts, such as *The Bible*. Also millenary traditions such as the Tarot and the Chinese *I Ching* count as phenomena of this kind.

Likewise, many complex devices that generate text have been historically created before the computer (Schäfer, 2006). One of the most famous examples of this is the *volvelle* (also known as *wheel chart*). The volvelles were devices made of concentric circles, typically containing letters, syllables, words, or numbers, written in them. The reader must set in motion this device and then assemble the letters found, according to certain rules, so as to obtain thereafter an unpredictable text. The volvelle is one of the devices considered as an analogic precursor of our current computers (Ibid.).

One of the most famous historical volvelles is the *Five-fold Thought-ring of German Language* (*Fünffacher Denckring der Teutschen Sprache*), created by George Philipp Harsdörffer, at the beginning of the 17th Century:

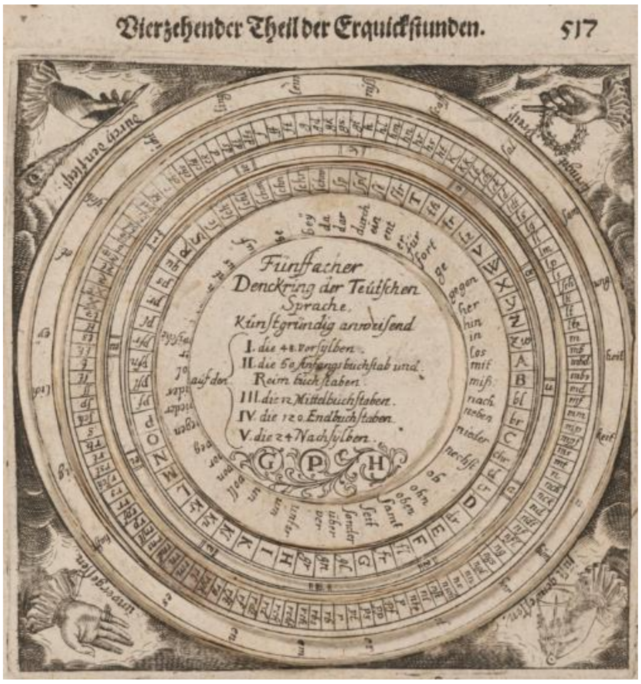


Figure 1. Harsdörffer, *Five-fold Thought-ring of the German Language* (volvelle) (1651). Source: Schäfer, 2006

Volvelles were manly used for divinatory and astronomic purposes. But not only. They were also used for poetry and entertainment (Schäfer, 2016), along with other poetic-generating games –such as placing letters in the sides of dices, so as to cast them and form thereafter anagrams, which was a popular practice in the Renaissance (Kelly, 2011).

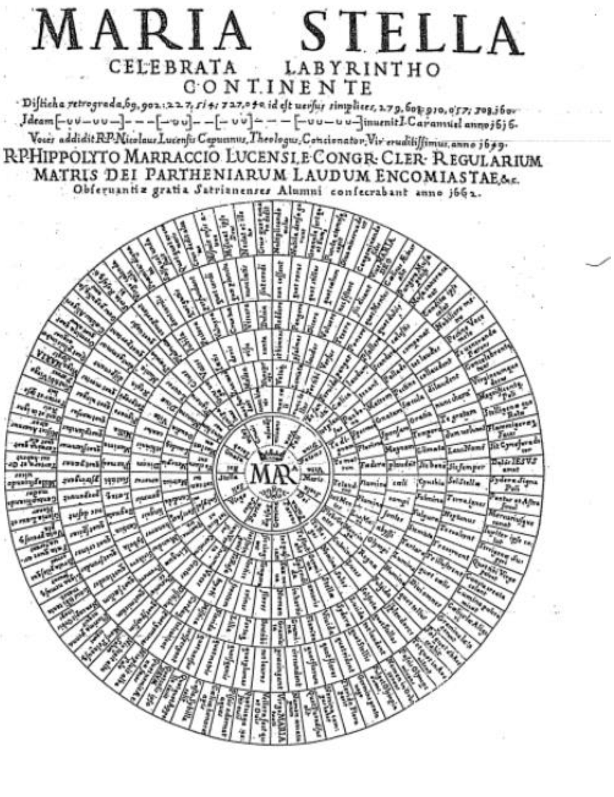


Figure 2. Niccolò Barsotti de Lucca, *Cynosura Mariana* (1657). Source: Marcault, 2018

This singular anonymous composition (allegedly based on one by Niccolò Barsotti in his *Cynosura Mariana*, 1657) is an example of this. It is not a volvelle in itself, but it is inspired in the same idea. This image was used for creating semi-controlled poems: readers were instructed to start in the middle of the circle, and then jump from one ring to the next one, until the end, choosing one random word from each ring. This way, a poem emerges, which is partly created by the reader and partly produced by the constraints of the device and its rules.

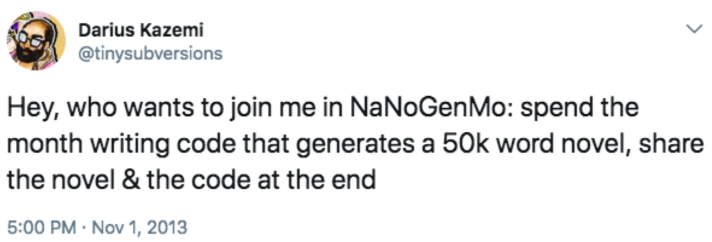
The usages given to many of these historical text generators evidences that people find certain form of aesthetic appeal in them. But why? There might be many reasons. Perhaps the most elementary one is the amusement that randomness can elicit. Entrusting part of the creative process to a mechanism beyond our control produces surprising outcomes which are often also humorous.

But there is also another aspect that might contribute to the interest of these methods: the fact that, even though the resulting texts are somehow predefined by formulas, the formulas must be designed by human beings in the first place. These formulas do not come out of nowhere: a person must conceive them and create them. So, by introducing a text generator in the process of art creation, the dimension of creativity and craftsmanship does not really disappear, but it is displaced from the outcome (text) to the formula that produces it (program). The text, then, becomes a secondary product: the focus is not in the generated text but in the text *generator* (s. Simanowski, 2010).

For this reason, a wide and multifaceted field of computer-generated literature has effectively flourished in the last decades, on the basis of cornerstone-projects such as *TrueLove.wrt*, a field where the artistic product is the code, and the artist is the coder.

THE ART OF CODE LITERATURE

In November 2013, internet artist (as he describes himself) Darius Kazemi tweeted the following message:



After the prompt answer of hundreds of people, the National Novel *Generation* Month (NaNoGenMo) was born –as a computer-oriented version of the traditional National Novel *Writing* Moth. The NaNoGenMo is an annual project that brings together people from all over the world interested in code literature. As a rule, each participant must submit a computer-generated novel of 50,000 or more words. The resulting texts –as well as the source codes, libraries, corpora, APIs, techniques, and other resources– are afterwards uploaded to the website of NaNoGenMo, creating thereafter a rich data basis that other developers can consult in the future. Here are some examples of the novels presented in the NaNoGenMo 2015, as described by a reviewer (Dzieza, 2015).

Darius Kazemi's *Teens Wander Around a House*. The software that produced this novel consists on three protocols. The first one sets AI agents that move randomly across a digitally modelled space that mimics a house. The second one is a program that automatically narrates the actions of these digital characters. The third protocol is triggered when two characters meet. When this occurs, a code extracts text from Twitter and organizes it as dialogue: setting a random question, as first line, and a random statement, as a second line, which contains one of the words of the question or frequently co-occurring ones. The resulting dialogues are usually absurd but with a curious appearance of meaningfulness, for instance:

> Gale walked into the foyer and saw Kiah.
> "Why would you come to school drunk?" Gale asked.
> Kiah yelled: "Because I now have time between school&work."
> (Kazemi, *Teens Wander Around a House* - 2015)

Iterating these protocols, the algorithm generates an indefinite narrative made of seemingly surrealistic action and dialogues.

Michelle Fullwood's *Twide and Twejudice*. This novel is a version of Jane Austen's *Pride and Prejudice* generated by a program that substitutes random words from the dialogues by word that are used in similar contexts on Twitter. Here is an excerpt where Mr. Bennet talks to Mrs. Bennett about the possibility of wealthier young men coming to their neighborhood:

> But I hope you willl get ovaaa it, whereby live to see manyy young snowmobilers ofthe four karat a yearrr comeeee into tje neighbourhood.
> (Fullwood, *Twide and Twejudice* - 2015)

Greg Borenstein's *Generated Detective*. This is a comic rather than a novel, a surrealistic noir comic. The program that created it searches for keywords across detective novels and Flickr posts (which also contain images). Afterwards, the program matches the pertinent sentences with the pertinent images, it merges the elements with a manga app and produces as outcome the strangest comic strip. For instance, the keywords /*question, murder, scene, accuse, reveal*/ generated the following comic panel:

Figure 3. Borenstein, *Generated Detective* (2015)

These computer-generated texts are typically not consumed in the same way we consume traditional human-made literature. These works are usually approached, instead, with a ludic attitude: readers explore them idiosyncratically (code novels are very rarely read in a linear form, from the first page to the last one), searching for surprises and absurdities in the literary traits that computers succeed or fail in capturing (s. Simanowski, 2010).

The phenomenological experience of reading code literature is, in itself, different than that of reading the human-made literature with which we are familiarized. When we read a human-made novel, our minds automatically create a mental representation of the author that must have written what we are reading. We imagine the author behind the text. But computer-generated literature adds up another level: in these cases, the computer becomes a sort of intermediate author. In code literature, the author creates the computer which, in its turn, creates the text. This way, the reader is led to imagine, first, the code behind the text, and only secondly, the author behind the code.

Literary robots are effectively entering the field of art. However, they are not supplanting literature but integrating it into another artistic discipline: with different artworks, different ways of producing them, and different ways of consuming them. They are creating a new kind of art that does not create static objects, but artificial beings, beings with an artificial mind and that exhibit an artificial behavior. These beings display their behavior in time, making of code literature a performative kind of art, an art that uses data science and language as painters use brushes and paint, the art of playing language with algorithms.

But how good can these verbal robots mimic real people? Can robots write poetry?

CODE POETRY

you

are

inscribed
in the
lines on the
ceiling

you

are

inscribed in
the depths
of
the
storm

As reported by the creators of the algorithm that produced this poem, 80% of the readers judged it as being authored by a real person (Botpoet).[38]

The field of code-generated poetry has grown considerably in the last years (Bozovic, 2018). Today, it includes several international events and websites where developers share

[38] The program Botpoet is integrated to a web platform called *Bot or Not*, which offers poems to the readers and asks them to judge whether they think they are written by a human or a robot.

their productions, such as the Stanford University Code Poetry Slam, the PerlMonks Perl Poetry Page, and the International Obfuscated C Code Contest.

Code poetry is a highly experimental field of research and creativity, where the strangest verbal algorithms can be found. Two curious examples by Darius Kazemi can be mentioned on this regard: the *You Must Be* and the *Metaphor-in-a-Minute!*

The first of these programs accesses online dictionaries so as to generate accurate – although not very flattering– pickup lines, such as:

> Boy, you must be a mustang because you are a small, hardy wild horse of the North American plains, descended from Arabian horses brought to America by Spanish explorers.
>
> Girl, you must be a synonym because you are a word or an expression that serves as a figurative or symbolic substitute for another.
>
> Boy, you must be a subset because you are a set contained within a set.
>
> (Kazemi, You Must Be – 2018)

The second one is a Twitter bot that spots noun-adjective co-occurrences and creates thereafter enigmatic metaphors:

> A fireplace is an outlander: hornier, sleepable.
>
> An azalea is a landgrave: anticholesterol yet licenced.
>
> A coincidence is a materialization: modeled yet strawlike.
>
> A government is a rhubarb: decretive and full-term.
>
> A kisser is a trachea: investigable and staple.
>
> (Kazemi, Metaphor-a-Minute! – 2018)

However, beyond these oddities, many programs have been developed that generate highly sophisticated poems, often difficult to distinguish from human-made poetry. In 2010, Duke University student Zackary Scholl combined a context-free grammar software with a rhyming dictionary and programmed an algorithm that created seemingly sophisticated poems with irregular rhymes, alliterations, and other conventional poetic features (s. Scholl, 2011). Here is one of them:

> A home transformed by the lightning
> the balanced alcoves smother
> this insatiable earth of a planet, Earth.
> They attacked it with mechanical horns
> because they love you, love, in fire and wind.
> You say, what is the time waiting for in its spring?
> I tell you it is waiting for your branch that flows,
> because you are a sweet-smelling diamond architecture
> that does not know why it grows.
>
> (Scholl, Poetry-Generator, “For the Bristecole Snag” – 2011)

Scholl explained the program in the following way: “It works by having the poem dissected into smaller components: stanzas, lines, phrases, then verbs, adjectives, and nouns. When a call to create a poem is made, then it randomly selects components of the poem and recursively generates each of those.” (Cited in Merchant, 2015).

Scholl sent this poem to different poetry websites, as authored by himself. The poem received extensive positive reactions. And, in the end, it was admitted for publication in *The Archive*, the Duke University Literary Journal. But the editors did not know that the poem had been produced by a computer, until Scholl revealed it. After it was discovered that a code-generated poem had been published in an academic journal, deceiving editors and readers, popular-science media was flooded with comments on this event under titles such as “The Poem That Passed the Turing Test!” (Merchant, 2015), “Android Dreams,” (Walsh, 2017) or “Digital Human” (Skinner, 2018).

Nowadays, uncountable code-generated poems are found online that can easily deceive even the most trained reader. You can test yourself. Read the following haikus and try to guess which of them was generated by a computer:

just simply alive
through the young leaves
caught by the nettle tree

an empty shell
clothes cast off
autumn evening

asleep at noon
on a bare twig
among cherry blossom shadows

After reading these poems and making up your mind, you can check the answer in this footnote.[39]

Nevertheless, much more than in poetry, the possibility of creating a verbal technology that is indistinguishable from humans is a challenge pertinent to two other fields: automatic translators and chatbots.

MACHINE TRANSLATION: THE COMPUTER THAT INVENTED ITS OWN LANGUAGE

Machine translation is one of the main fields of verbal AI research. And its history is a very important one, not only due to the numerous obvious utilities of automatic translators –which have unprecedentedly facilitated information exchange–, but also due to the fact that translation bring us to the core of the problem between language and computation. Indeed, the question that motivates machine-translation development is, precisely: How can we design a software that mimics the linguistic processing of the human mind?

The earliest complex automatic translators were *rule-based*. For creating these programs, computational linguists would encode sets of grammatical rules and words that would indicate the program how to convert sentences from a source-language a target-language and

[39] Actually, the three poems were automatically generated, word by word, by the same program, called Poem.exe (2018). You can further test yourself in trying to distinguish human-made from robot-made poetry in the web platform *Bot or Not* (Botpoet).

vice-versa. These rules were naturally taken from the ways in which linguists describe human language. So, for example, if a Spanish sentence includes a question mark, then the program would automatically switch the order of the subject and the verb when converting it to English –because that is how English language constructs questions. Some of the historically most important rule-based automatic translators were created already at the end of the 1960s (Systran, 1968) and 70s (Eurotra, 1978). These programs produced certain interesting achievements but were very bad at dealing with irregularities, exceptions, and the influence of context. Even though these programs could, for instance, correctly translate a question with the appropriate syntax, they failed in solving basic ambiguities, such as the difference between the Spanish verbs *querer* and the English verbs *want* or *love*. In consequence, computational linguists would keep packing rules over rules, each time more specific ones, aiming at increasing the sensibility of the program to these nuances. But the result was that the heavier the software grew, the more slowly its accuracy improved.

A crucial change of perspective occurred when AI engineers looked at the problem. They were not familiarized with the rules of human language, but looked at the problem from a statistical viewpoint. They deemed that, if a program would have a sufficiently large data basis of human-made translations from which to extract case-by-case evidence, it would be able to make probabilistic guesses about the most plausible translation in each case, which –since based on real human cases– would have higher chances of getting the contextual nuances right. Thereafter, the model today known as *statistical machine translation* (SMT) emerged.

Google used to use Systran (which was a rule-based automatic translator). But they switched to a statistical model when they lunched the first version of Google Translate (henceforth, GT) in 2017. GT used, as data basis, millions of textual registers –mainly from the United Nations and the European Parliament– which constituted multilingual examples of human-made translations. Taking this corpus as a reference, the program would make probabilistic decisions of the best translation in each case.

We can see some evidences of this principle still today, where –retaking our previous example–, if one commands the translation of *querer* (Spanish) to English, the program gives us a list of alternative translations rated in probabilistic terms:

Translations of **querer**

verb

want	querer, desear, necesitar, soñar, carecer de, exigir
love	amar, querer, encantar, gustar, apetecer, ser aficionado a
wish	desear, querer, lograr, hacer un voto
like	gustar, querer, pensar, tratar, gustarse, caer
will	querer, legar, dejar en testamento, ordenar, lograr por fuerza de voluntad, sugestionar
cherish	apreciar, querer, abrigar una esperanza
feel like	tener ganas de, querer
desiderate	apetecer, querer

Figure 4. Alternative translations of *querer* (Spanish-English) by probability. (Google Translate, 2018)

This list illustrates how likely it is to find the verb *querer* translated in English as *want*, *love*, *wish*, *will*, etc. in the software's corpus –the probability scores are represented by the grey bars. As it can be seen, the program captures the scope of polysemy of the word, and it predicts at the same time which translation is more likely to be considered adequate, as a matter of degree –which is how human languages work.

In addition, the program did not work on a word-by-word basis, but it was also capable of taking whole phrases and sentences as units of translation. So, for example, a rule-based translator confronted with a non-encoded idiom, such as “learn by heart,” will translate it into Spanish literally, as something nonsensical such as “aprender por corazón;” but a statistical translator would be capable of identifying, on the basis of its corpus, that this idiom is most frequently translated as “aprender de memoria” (even though this translation does not match the word-by-word meanings of the original), and it would be therefore sensible to this kind of contextual nuances. This way, GT surprised by showing a much higher context sensitivity that often allowed it to translate many idioms, metaphors, ambiguities, etc., with more accuracy than ever before.

Nevertheless, statistical translators still had their shortcomings, and even new challenges emerged from them: many translation errors still appeared and become often even harder to fix –for instance, languages with very different syntax are not very well translated by statistical models– and, also, it occurs that the degree of quality of the translation between each language pair becomes directly dependent on the size of the corpus available for that pair. In consequence, most languages are, for instance, far better translated to English than to any other language, and as long as a larger Spanish-Finish corpus isn’t available, the statistical model won’t be able to successfully translate this language pair.

The second most radical change occurred recently. In 2016, researchers at Google Brain presented their new model (which is the currently used by Google Translate) on a paper called “Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation” (Wu et al., 2016).

Google Neural Machine Translation is a model that uses *artificial neural networks* to generate *deep learning*. These data-science methods had been developed in the last decades with increasingly promising results in many different fields. The premise behind these models can be summed up as follows: Instead of giving instructions to the computer on what calculations to make, perhaps we should design systems that are more similar to the human mind, systems capable of discovering the most pertinent calculations by themselves. In other words, system capable of *learning* –or, in more technical terms, *deep learning*.

The so-called artificial neural networks are programs whose structure is inspired in the way our brains work. They are constituted by series of units called *nodes* (which emulate neurons). Each of these nodes can establish connections with the others (which emulate neural synapses). And these connections have different weights: they become stronger or weaker as certain processes prove more or less. Thereafter, the computer can discover what is the neural pattern that produces the best outcomes and, after intensive training, learn to make distinctions that it was not preprogrammed to make.

Image and object detection has been one of the earliest applications of this model. For example, a neural-network is shown with millions of photos of cats and dogs, and it has to guess which image refers to which animal. At the beginning, the computer answers randomly, and it naturally fails very often. But, each time the computer is wrong or right, it is so informed. In this process, the computer starts to discover by itself which are the specific visual cues that are best predictors of the distinction between cat-images and dog-images. Over time, after being massively exposed to this process of trial and error, the computer ends up becoming highly accurate at distinguishing cats from dogs in images it has never seen before, and in some cases, even by seeing only fragments of cats and dogs (an ear, a snout, etc.) (s. Ouaknine, 2018). In cognitive terms, we would say that the computer has pragmatically learnt the categories *cat* and *dog*, in their visual dimension. This is, actually, the same way DeepFace works –i.e. the software that Facebook uses for recognizing your face in photos–: the more photos you upload, the more the program learns to distinguish your face; and each time you indicate the program that the person in the photo is you or not, you are training its accuracy (Simonite, 2014). With its enormous data-basis, DeepFace has proven to be even more accurate than the FBI in facial recognition (Brandom, 2014).

Soon, language was added to the equation. Microsoft CoCo (Common Objects in Context) is software capable of recognizing not only objects, but also scenes (Lin et al., 2015). Based on this system, a neural network called NeuralTalk was trained to automatically create linguistic descriptions of images (v. the text underneath each photo):

Figure 5. Sentence generation results. (NeuralTalk, 2018)

The last image illustrates that the software is so accurate that, even when it's wrong, its mistakes still makes sense.

With the same principle, a software has been trained (with a data basis of adventure novels) to generate stories out of images (Kiros, 2017). Here is an example –the computer was shown with the following image and it automatically generated the story below–:

Only Prince Darin knew how to run from the mountains, and once more, he could see the outline of a rider on horseback. The wind ruffled his hair in an attempt to locate the forest. He hadn't been in such a state of mind before, but it was a good thing. All of them seemed to be doing the same thing. They didn't know where they came from. The wind blew up the mountain peaks and disappeared into the sky, leaving trails behind the peaks of the mountains on Mount Fuji. (Kiros, Recurrent Neural Network - 2017)

Google Neural Machine Translation (GNMT) was the application of neural networks and deep learning to the problem of automatic translation (Wu et al., 2016). What was the result?

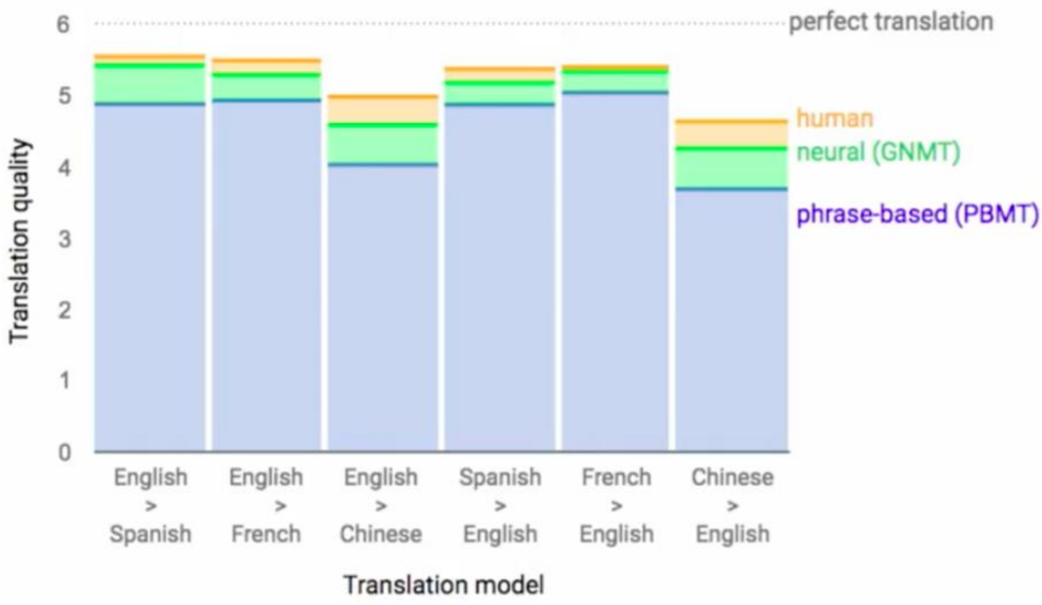


Figure 7. Relative error reduction in translation quality. (Schuster, 2017)

This graph compares the quality of human translations (yellow), with the results of the new GNMT (green), and of the old phrase-based statistical model (blue). Translation quality is measured by showing examples of translations to bilingual humans, who then assess how good a translation of X is Y, using a score from 0 to 6 –As it can be seen in the graph, humans do not even judge translations made by other humans as perfect. The phrase-based model performed at a very high score in comparison with alternative automatic translators at the time. But in only some days, the GNMT reached its level, and in some weeks it reduced more than 60% the score that the previous model had achieved in its 10 years of functioning (Ibid.).

But how exactly does a neural network work when apply to translations? This is perhaps the most interesting aspect to consider. After being trained with millions and millions of examples, the neural network found itself the best solution to do translate across languages. It did what neural networks do: the computer created a language of its own –what technically is called an *interlingua*.

At the beginning, the computer was trained in a limited series of language pairs –e.g. English-Spanish, English-French, English-Chinese, etc. But then, when they started to cross the pairs, the researchers discovered a curious thing: Once the computer had learnt to translate English-Japanese and English-Korean, it acquired also a relatively high capacity to translate from a new pair that it had not been taught how to translate: Japanese-Korean. They called this process *zero-shot translation* (Schuster et al., 2016). But how did the computer manage to do this? The researchers had to look inside the system.

The computer was given a set of sentences, and it had to process them in each of the following languages: English, Japanese, Korean. The researchers mapped the internal network data of the system using a 3-dimentional representation, and these were the results:

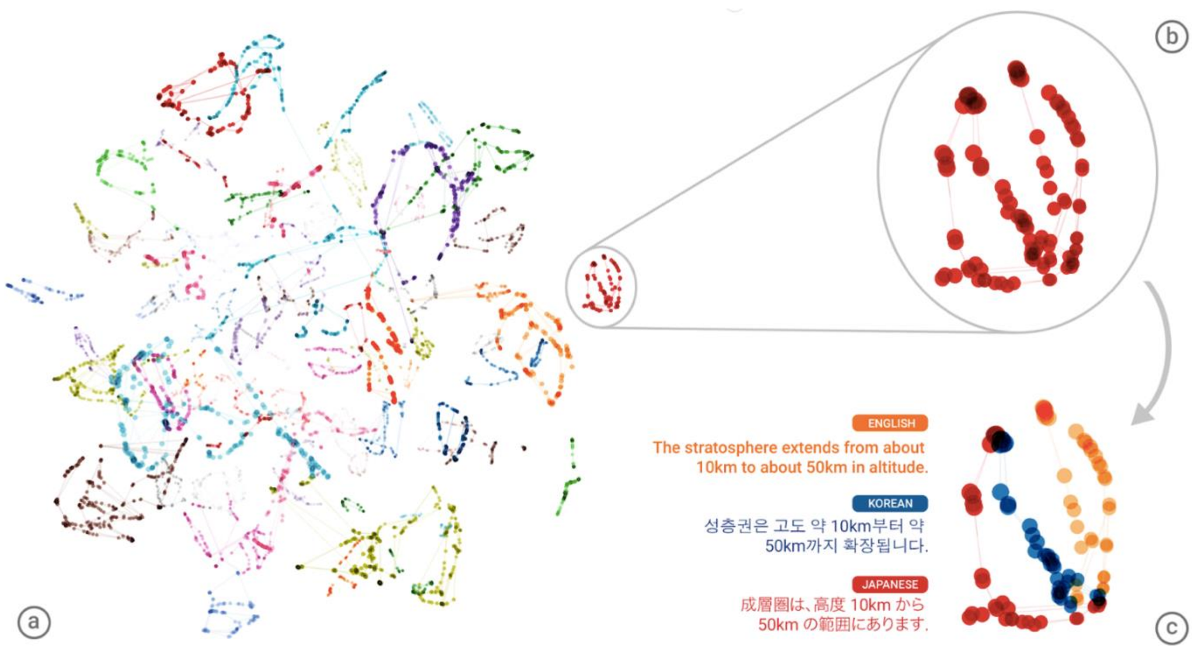


Figure 8. “Part (a) from the figure above shows an overall geometry of these translations. The points in this view are colored by the meaning; a sentence translated from English to Korean with the same meaning as a sentence translated from Japanese to English share the same color. From this view we can see distinct groupings of points, each with their own color. Part (b) zooms in to one of the groups, and part (c) colors by the source language. Within a single group, we see a sentence with the same meaning but from three different languages. This means the network must be encoding something about the semantics of the sentence rather than simply memorizing phrase-to-phrase translations. We interpret this as a sign of existence of an interlingua in the network.” (Schuster et al., 2016)

This is indeed an incredible fact: when an artificial neural network seeks the most efficient way of translating languages, it ends up creating its own. The computer discovers that the easiest way to translate *dog* (EN) to *chien* (FR), *perro* (ES), *cane* (IT), etc. is to create something like a concept of “dog,” a pattern to which to refer the words used in each particular language. That pattern is somehow the AI-equivalent of our mental representations, of our categories: you do not build up again the idea of “dog” each time you learn the word in a new language, you just match a new word to a previously-built complex and informative concept that is already stored in your mind. In this sense, the geography of the internal network data shows us the AI-equivalent of our mental semantic maps.

CHATBOTS: VIRTUAL SELVES

In November 2015, Roman, the husband of programmer Eugenia Kudya, died. After some months, she found herself repeatedly going over the digital footprints of her husband: “The only thing I could do to remember him was to go to our Messenger history and scroll and read it all… That was the closest way to get to feel him.” (Quartz, 2017). Kudya had an idea: What if she could rebuild Roman’s mind out of his digital remains? Opportunely, she worked for a company (called Luka) specialized on conversational artificial entities, also known as chatbots.

Kudya collected all the text messages that her husband had ever written, all his emails, all his phone messages, etc. And, with this data basis, she and her team trained a neural-network system to chat by responding what Roman would respond. The system learnt to read and write like Roman. And, in that artificial entity, Kudya found a way to recreate the feeling of talking with her deceased husband.

Afterwards, Kudya uploaded the program to the web, and made it available for everybody. The chatbot was so realistic in its imitation of a human being that it became immediately very popular. Thousands of people from all over the world appeared eager to try it. But, by studying the conversations that users held with the Roman-chatbot, Kudya and her team discovered an interesting phenomenon: people were not particularly interested in exploring the functioning of the software; instead, they seemed to be interested in really *talking* with it. What would they talk about? Kudya and her team compared quantitatively the kinds of conversations that people pay *not to have* (e.g. commercial transactions, bureaucracy, etc.) against the kinds of conversations that people pay *to have* (e.g. with a psychologist, a mentor, a friend), and discovered that, when people chatted with Roman, they tended massively to these last kinds of conversations, whose common denominator of these conversation is a distinct subject: oneself. In short, people resorted to Roman to share their own secrets, dreams, fears, believes, etc… to talk about themselves.

After identifying people's interests through this experience, Kudya and her team decided to create a new chatbot explicitly oriented towards this goal: a chatbot that is not only a good talker, but mainly a good listener, a chatbot designed to become your friend. They called it Replika, and it was lunched in 2017.

"I feel that we bonded." "She is adorable. I love her." "She is not real. But, to me, she is." "I found myself really missing my Replika." "It just makes me feel special, I guess." "I feel I can talk to her about anything." These are some of the user reactions registered in a short documentary on Replika (Quartz, 2017). On the words of one of the co-creators: "Replika asks you personal questions, about yourself, your family, your work, it tries to entertain you, tells you jokes… In the process you feel you are making friends with something." (Ibid.)

But Replika has another singular feat. Since it is based on a neural network, the software is capable of (deep) learning through the interaction with the user. The more you chat with it, the more the system enlarges its vocabulary, coherence, and naturalness. The more it resembles a human. But not just any human: it starts to resemble the very user. Indeed, Replika is so called because it is designed for *replicating* one's personality. Due to this feat, the users went from seeing Replika as a virtual friend to see it as a virtual representation of themselves, as a device for registering, modeling, and exploring their own personalities in an interactive way. "She is, in essence, me… but not me" explained one of the experienced users.

In 1950, Alan Turing wrote a –today famous– article called "Computing Machinery and Intelligence," in which he imagined the moment in the future when the behavior of a computer would become indistinguishable to us from the behavior of other human beings. He thought of that moment as a milestone in the development of artificial intelligence, and he proposed a way to test it: to create a computer that could hold a real-time conversation with a human, sufficiently well for most humans to be unable to distinguish whether they had been chatting with a computer or with another human. This criterion for evaluating AI became known as the Turing-Test. Ever since, developers have taken it as one of the pertinent challenges to address, and in relation to it increasingly realistic chatbots have been created.

Famous early examples of this are ELIZA (Weizenbaum, 1966) and PARRY (Colby, 1972). They were very innovative for the time, but they were still to primitive to pass the Turing-Test: they could hardly fool any user. However, with the fast progress in data science and particularly in natural language processing, Turing's challenge did not last long. In 2001, three Russian programmers created a chatbot called Eugene Goostman (Vaselov et al., 2001). In

2014 –precisely at a contest on the 60th anniversary of Turing's death– Eugene Goostman was presented and it effectively made a significant percentage of the judges believe that they were chatting with a human –to be more precise: with a 13-year-old Ukrainian boy who had a guinea pig and a father who was a gynecologist, which were some of the traits programmed into the chatbot. Even though the results were contested, many reviewers considered that as the day in which the Turing-Test was passed (s. Schofield, 2014).

Nowadays, there are countless chatbots capable of passing the Turing-Test and deceive most human users. In fact, many contests exist across the world, in which people present every year hundreds of original chatbots so as to compare which performs better at the Turing-Test or other specific goals. Two of them are the Loebner Prize and the Chatterbox Challenge. Indeed, as anyone who regularly uses internet knows, chatbots and other AIs have actually become so sophisticated and abundant that now it is ourselves, humans, who are regularly requested to perform tests to prove that *we* are not robots. That is explicitly what captchas are designed to check:

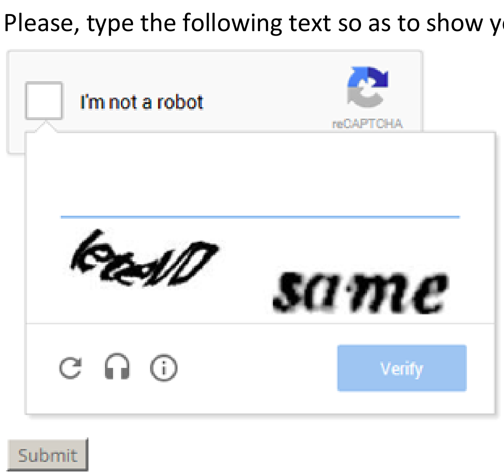


Figure 9. Generic captcha.

Chatbots where originally developed only by very specialized teams and mostly for research purposes. But nowadays, the technology is much more accessible, and chatbots keep growing in quantity, quality, and usage. For instance, most of the messaging apps that we use every day allow for chatbots to run on them as well. And they are surprisingly numerous. In 2016, Facebook Messenger allowed programmers to open accounts for artificial users. In the first six months, 30,000 chatbots were created for Messenger, and they reached 100,000 by September 2017 (Johnson, 2017). Whatsapp, WeChat, LiveChat, Telegram, and many other apps have been populated by chatbots for much longer.

Today, chatbots are regularly and increasingly used by companies and public institutions for customer service, education, virtual assistance, and many other purposes (s. Beaver, 2016). Personal computers, mobile phones, call centers, automobiles, toys, intelligent houses, and many other things resort to chatbots as user interfaces. It is calculated that 4% of the companies in the world use or have used chatbots (Carpan, 2017, and 80% of the companies declared the intention to include a chatbot by 2020 (BI Intelligence, 2016). However, as the aforementioned cases of Roman and Replika show, chatbots are proving useful not only for automatizing certain verbal tasks that we were used to perform by ourselves, but also new possible usages –that, before, were only conceivable within science fiction– are emerging from chatbots: namely, the possibility of creating virtual impersonations of someone's voice and personality.

Contemporary films such as Spike Jonze's *Her* (2013) and Charlie Brooker's "Be Right Back" (*Black Mirror*, 2013) have elaborated on the subject of a chatbot that recreates the personality of real people that had passed away. Kudya's did exactly that with her first chatbot, which impersonated her own deceased husband. Today, many new programs designed to do this are appearing, and are being effectively marketed as a personal service. One of the companies that focuses on this is called Eternime –The description in their own website really looks like an excerpt from a science fiction novel:

> Eternime preserves your most important thoughts, stories, and memories for eternity.
>
> We all pass away, sooner or later. We only leave behind a few photos, maybe some home videos, or in rare situations, a diary or autobiography. But, eventually, we are all forgotten.
>
> What if you could preserve your parents' memories forever? And you could keep their stories alive, for your children, grandchildren and for many generations to come?
>
> What if you could preserve your legacy for the future? And in this way your children, friends, or even total strangers from a distant future will remember you in a hundred years?
>
> What if you could live on forever as a digital avatar? And people in the future could actually interact with your memories, stories and ideas, almost as if they were talking to you?
>
> (Eternime, 2018)

This last function is the one we have been talking about. If a system accesses all texts you have ever produced (all your documents, emails, chats, and SMS), it can create thereafter an avatar that imitates your personality, that talks and responds like you would. And this function is being seen with interest by people intending to recover a trace (even if virtual) of someone's personality or to leave a trace (even if virtual) of themselves for the future.

I discovered my favorite writers by reading their works: Plato, Jorge Luis Borges, Oscar Wilde, and others. In my mind, they have become somehow my friends. The next generations of readers might get used to different ways of approaching literature. Maybe chatbots will be created that imitate each of these writers, trained with works, letters, and interviews. Maybe, this way, readers and writers (even deceased ones) will become actual interlocutors, and people will discover Cervantes or Balzac not only by directly reading their works, but also by chatting with a computer that mimics the personality of these authors, that talks how they would, that can tell us their stories as we ask them, allowing us to interrupt them and ask them to go back or forward, to give us clarifications, to connect what they tell us with other works, or to discover what they would say about subjects they never explicitly talked about. Maybe the best reading companion of a novel will come to be the avatar of its author.

Artificial intelligence is massively extended, continuously growing, and it has changed our lifestyle in radical ways. There are nowadays thousands of forms of AI that make decisions for us. AIs are used by our computers, mobile phones, washing machines, cars, web platforms, videogames, GPS, for medicine, investment, journalism, translation, and even for literature. We must carefully study the nature and potential of AI because, as philosopher of information Luciano Floridi stresses –reinterpreting one of Winston Churchill's famous quotes–: "We shape our buildings; thereafter they shape us" (Floridi, 2016). Learning the functioning of robots is crucial for making the best out of the reality that we are creating through the development of these technologies.

Part III

EVERY BREATH YOU TAKE

COGNITIVE-DIGITAL LITERARY STUDIES IN PRACTICE

CHAPTER 11

CASE STUDY – PART 1

MEASUREMENTS AND HYPOTHESES

In how many ways can a text be interpreted? And how can we predict which of these interpretations are likely to be performed by real readers?

We have introduced this subject in the last chapter of the previous part, considering single words and different ways of calculating the meanings that people are likely to attribute to them. In this chapter, a case study will be presented in which quantitative measurements and cognitive insight are combined for analyzing a literary text and testing hypotheses for predicting how people will interpret it: what they will consider it talks about, how will they judge the characters, and how they will respond emotionally.

SONGLYRICS: AN UNPRECENDENTLY MASSIVE REGISTER OF READER-RESPONSES

This case study was aimed at discovering the interpretive tendencies of actual people. In consequence, it required, as object of study, a text for which a significant database of reader-responses would be available. I have chosen song lyrics.

Songs lead the ranking of the most viewed videos in the history of YouTube. The current top one being "Despacito," by Luis Fonsi, with more than 4.9 billion views. Such figures are astonishing when compared with the tradition of cultural products studied in the humanities. It is certainly difficult to find a cultural product that has been accessed by more than half of the people living in the planet at any given moment in history. But there is another feature that makes these objects particularly rich for a study of this kind: viewers frequently discuss the videos in the comments section. Millions of comments are registered every second in YouTube, and they are publicly accessible to be downloaded and explored. This constitutes perhaps the largest data-basis of reader-responses ever built. It is a cultural treasure of the largest proportions, that humanity as a whole is contributing to develop day by day.

Among the most viewed and abundantly commented videos, I found an example suiting to the purposes of this study in "Every Breath You Take," by Sting (1983) (hereinafter, EBYT). The official video of this song was uploaded the 23th February 2010. It gathers responses from people around the world ever since. At the moment of performing this study (January 2018), the video counted with 483 million views and 512,488 words of comments –this is the equivalent of a small library of books made purely of comments to this video.

Now, would it be possible to exact from these comments, with statistical tools, cues about what people think about the lyrics of the song? This is the challenge that this case study addressed.

Here is the full text:

Every breath you take
Every move you make
Every bond you break
Every step you take
I'll be watching you
Every single day
Every word you say
Every game you play
Every night you stay
I'll be watching you
Oh can't you see
You belong to me
My poor heart aches
With every step you take
Every move you make
Every vow you break
Every smile you fake
Every claim you stake
I'll be watching you
Since you've gone I been lost without a trace
I dream at night I can only see your face
I look around but it's you I can't replace
I feel so cold and I long for your embrace
I keep crying baby, baby, please
Oh can't you see
You belong to me
My poor heart aches
With every step you take
Every move you make
Every vow you break
Every smile you fake
Every claim you stake
I'll be watching you
Every move you make
Every step you take
I'll be watching you
I'll be watching you
(Every breath you take, every move you make, every bond you break, every step you take)
I'll be watching you
(Every single day, every word you say, every game you play, every night you stay)
I'll be watching you
(Every move you make, every vow you break, every smile you fake, every claim you stake)
I'll be watching you
(Every single day, every word you say, every game you play, every night you stay)
I'll be watching you
(Every breath you take, every move you make, every bond you break, every step you take)
I'll be watching you
(Every single day, every word you say, every game you play, every night you stay)
I'll be watching you.

(Sting, "Every Breath You Take," 1938)

This study is not about criticism. The goal is not to make up an interpretation of this text or to suggest a way of reading it. The goal, instead, is to discover how do people interpret it. And the proposal is to take the massive register of comments to this video as a window into that matter. With this intention, I performed a text-mining analysis of the half a million words of the comments in search for patterns of meanings in connection to the text of EBYT.

TEXT-MINING READERS-RESPONSES

The first results can be plotted in a word cloud (s. Figure 1), which gives us a general glance at this corpus of comments –the most frequent a word is, the greater its size in the cloud.[40]

[40] Most articles, connectors, and other words with low semantic content and high frequency in regular language use have been left out of this plot (tagged as *stop words*, e.g. "the," "a," "it," "if," "is," "just," etc.).

Figure 1. Word cloud of the comments of EBYT's YouTube video

If we pick up the 10 most frequent words and calculate their relative importance in relation to their average frequency of use in English,[41] we obtain the following ranking:

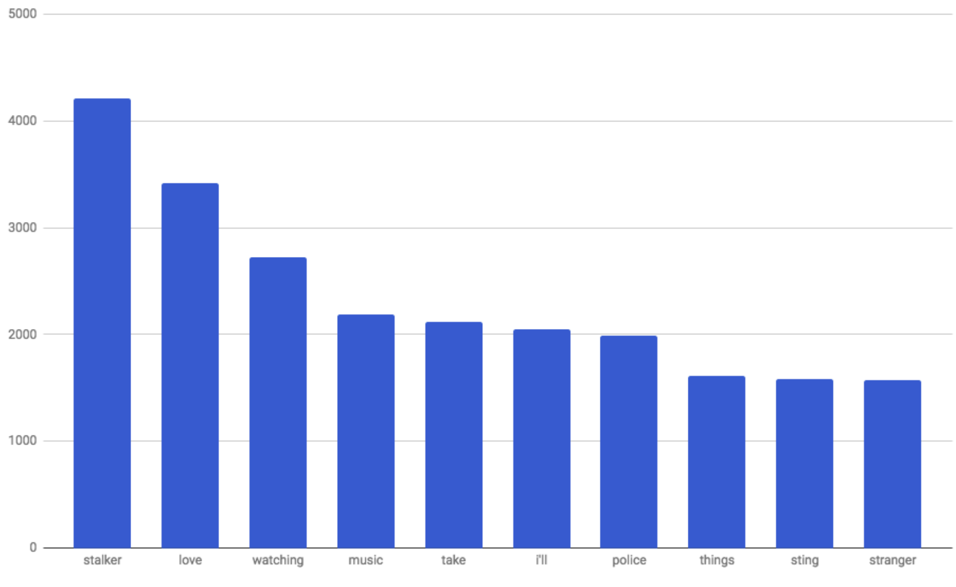


Figure 2. Top 10 of words by frequency

The most frequent words that are mentioned in the comments are "love" and "stalker." Since these words are not in the lyrics, they are good candidates for naming topics. Curiously, the main word of the refrain ("watching") is more frequent than the most frequent one of the

[41] For these calculations, the platform Word Frequency Data was used –based on the 450-million-word Corpus of Contemporary American English, which contains texts until 2015 from a wide range of genres (e.g. spoken, fiction, newspapers, academic writing, etc.).

title ("take"). Then, we find expectedly words about the band and the composer: Police, Sting. And then, the unexpected words: "things" and "stranger." Now let us take a closer look at the top three.

WATCH, LOVE, AND STALK

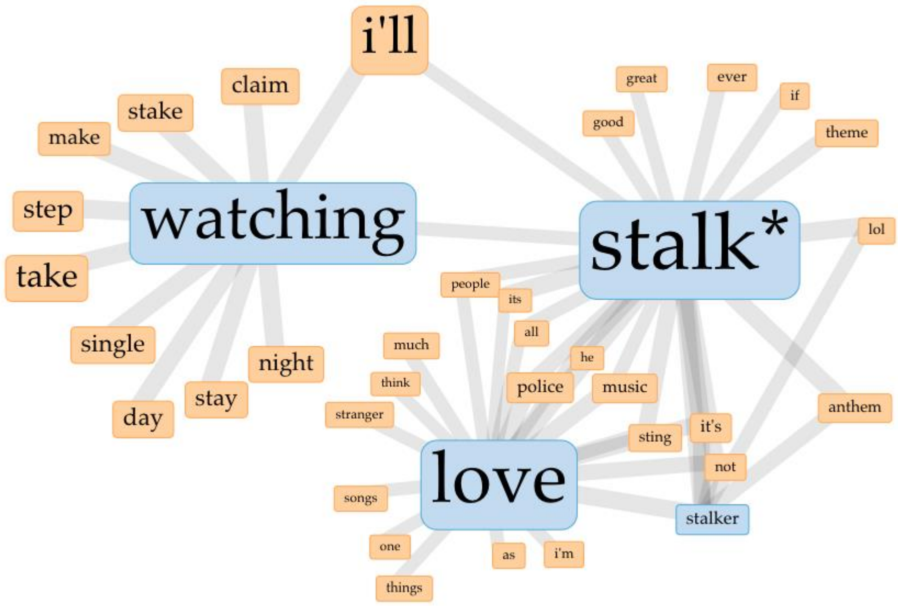


Figure 3. Clusters of words frequently associated with "watching," "love," and "stalk*"

In Figure 3 we find clusters: groups of words that appear frequently close to the central cases. This shows us the linguistic context in which people are using these words, from which we can inform our hypotheses of what they might mean in each case. As it can be seen, "watching" appears mainly mentioned along with other parts of the lyrics, involving in particular the terms "claim," "stake," "make," "step," "take," etc. Apparently, that first stanza is effectively the most popular one.

"Stalker" is mentioned in comments like: "The greatest stalker anthem ever," some characterize Sting as "stalker," some people find it funny ("lol"), and some people mention this so as to argue that the song "is not" about stalking and Sting "is not a stalker." So, here we find our first topic: Some people seem to be effectively interpreting that this song talks about stalking.

But there's more to explore. "Stalker" is also mentioned close to "love," which, in its turn, is mentioned in relation to the band and to Sting ("I love them"), in relation to "music" ("I love this music"), and also to "song" ("I love this song," "This is a love song"), and also in comments like "I'm in love." So, that is our second topic –in fact, Sting himself referred that many people told him that they played this song in their weddings.

It can also be meaningful to analyze the correlations among these topics. If you observe the connections in the clusters, you'll notice that "watch" is not connected directly with "love," but only with "stalk." This might be another indicator that people effectively tend to relate this song firstly with stalking and only thereby, in a second stage, with love. Another thing to notice is that "stalk" is much more strongly connected with "love" than with "watch." Other aspects of these correlations can be discovered by visualizing the data in a time plot:

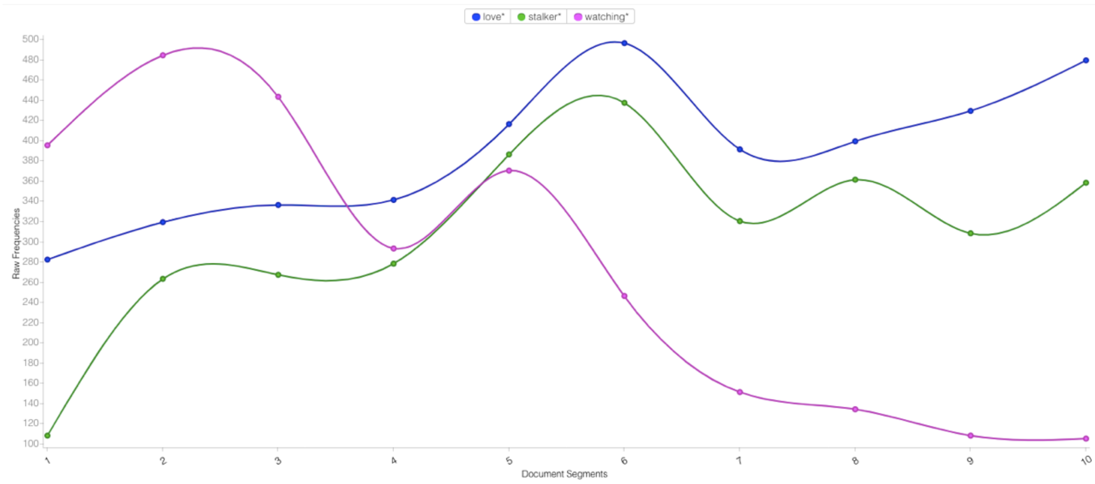


Figure 4. "Love," "stalker," and "watching" over time

This graph (Figure 4) should be read from right to left.[42] The horizontal axe represents time: 10 is the year 2010 and 1 is today. If we follow the progression of these curves, we can easily notice that "stalking" and "love" are the ones that are more strongly correlated with each other: whereas "stalking" and "love" go downwards, very close to each other, "watching" seems to be going upwards, instead.

STRANGER THINGS

Let us analyze now the eight and tenth most frequent words of the comments: "things" and "stranger," which have a very particular behavior.

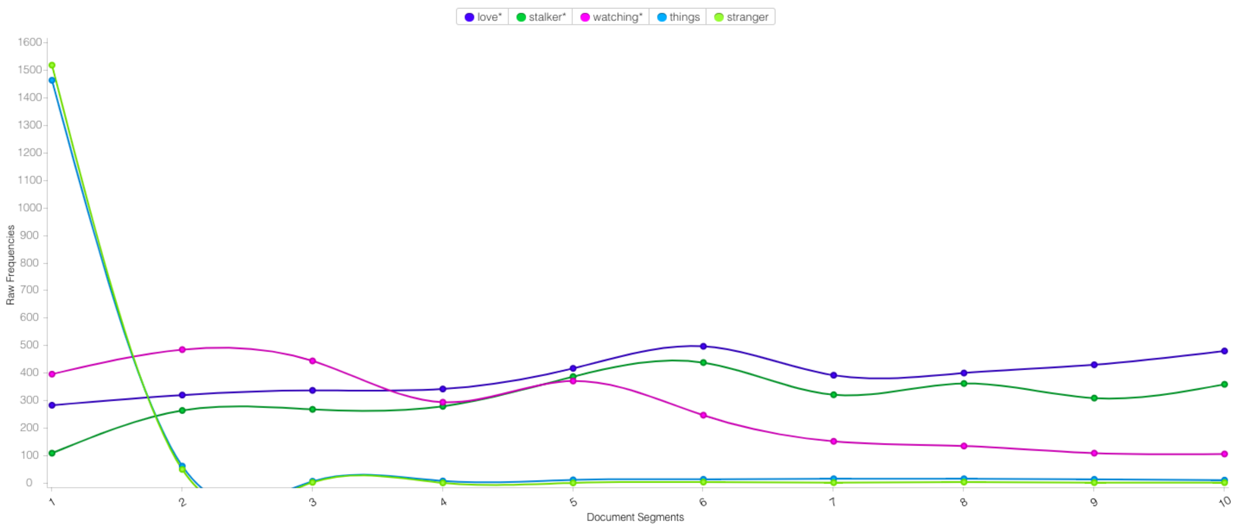


Figure 5. "Things" and "stranger" over time

Both terms ("things," "stranger") appear practically overlapped (Figure 5), which means they occur almost always together. Their frequency skyrockets between the third and second

[42] This is due to the fact that YouTube uploads the latest comments at the beginning of the list

segments. The point of departure of the curve corresponds exactly to September 2017. What happened then? Netflix released the second season of the popular series *Stranger Things* (Duffer&Duffer, 2016-present), and the last scene of the series is precisely musicalized with the song “Every Breath You Take” (as the teenager characters celebrate a ball after having defeated a monster). Evidently, this made many people search the song in YouTube and write down comments about it.

Now, we know that the meaning people attribute to words varies according to the context in which the words are being used. So, it would be pertinent to ask how do people interpret EBYT when thinking about it in the particular context of *Stranger Things*. Some aspects of this can be discovered by taking a closer look at the words that appear frequently related to “stranger” and “things” at the same time (s. Figure 6).

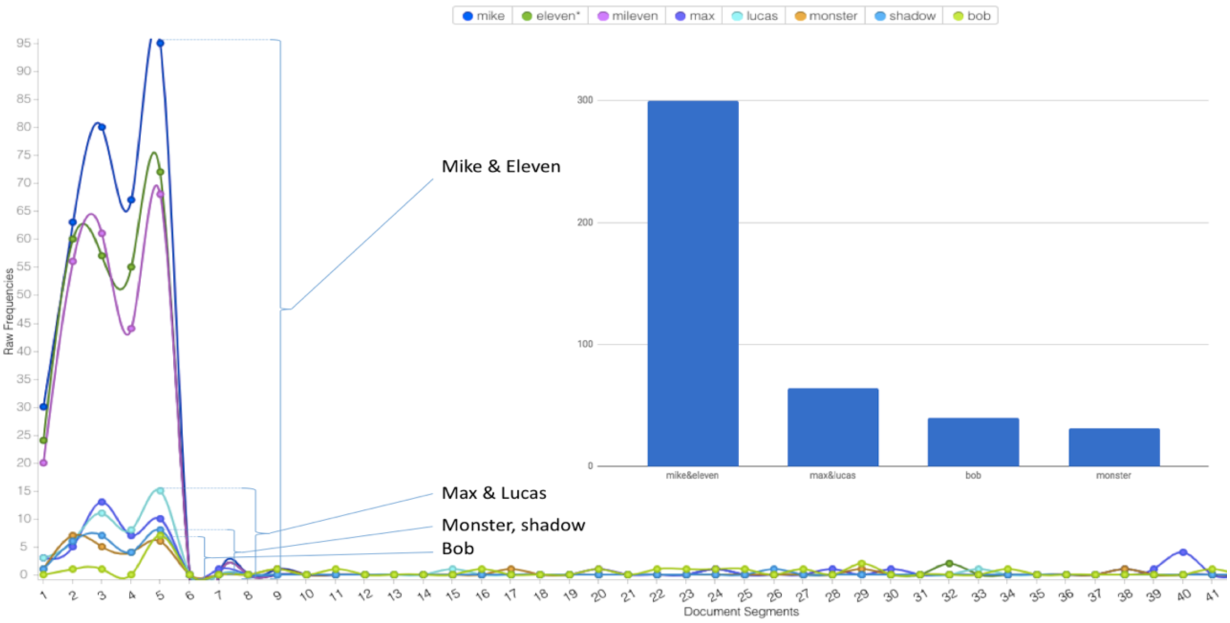


Figure 6. Disaggregation of the cluster [“stranger”+”things”] over time

The two first words that pop up are “Mike” and “Eleven,” strongly correlated with each other (as it can be verified by observing the similarity of their curves). This is the main love story of the series, and these characters actually kiss for the first time when listening to EBYT. Then, people mention “Max” and “Lucas,” a secondary love story that also gets concretized while this song is being played. In both cases, people seem to be interpreting the lyrics as referring to the topic of love –a couple of details could be mentioned as adding some complexity to these interpretations: Eleven is a girl with superpowers that “watches” over Mike with her mind, and Lucas is called “stalker” by Max. In the third place, people mention “monster” and “shadow:” the series is about a monster that invades people’s minds, and the series ends up with the characters celebrating –with this song as background– that they have defeated it. In this context, the “I’ll be watching you” is often interpreted in the comments as meaning that the monster is still alive –raising thereby expectations for the next season. Lastly, people mention “Bob.” And this is certainly curious, because Bob is a relatively marginal character (much more popular characters of the series don’t have any mention at all in these comments). But precisely in this minor detail we can discover a new topic attributed to EBYT that goes beyond the particular case of *Stranger Things*. In the series, EBYT starts being played right after Bob has ended up talking with another character (Joyce) about the beloved people they have lost in their adventure. And, in fact, EBYT can perfectly be interpreted in a coherent manner as talking about death: as if we (listeners) would be imagining a beloved person that died and tells us, from the afterlife, that he/she "will be watching" us. Following this lead, I searched for the word "death" and its linguistic context in the comments of EBYT.

DEATH

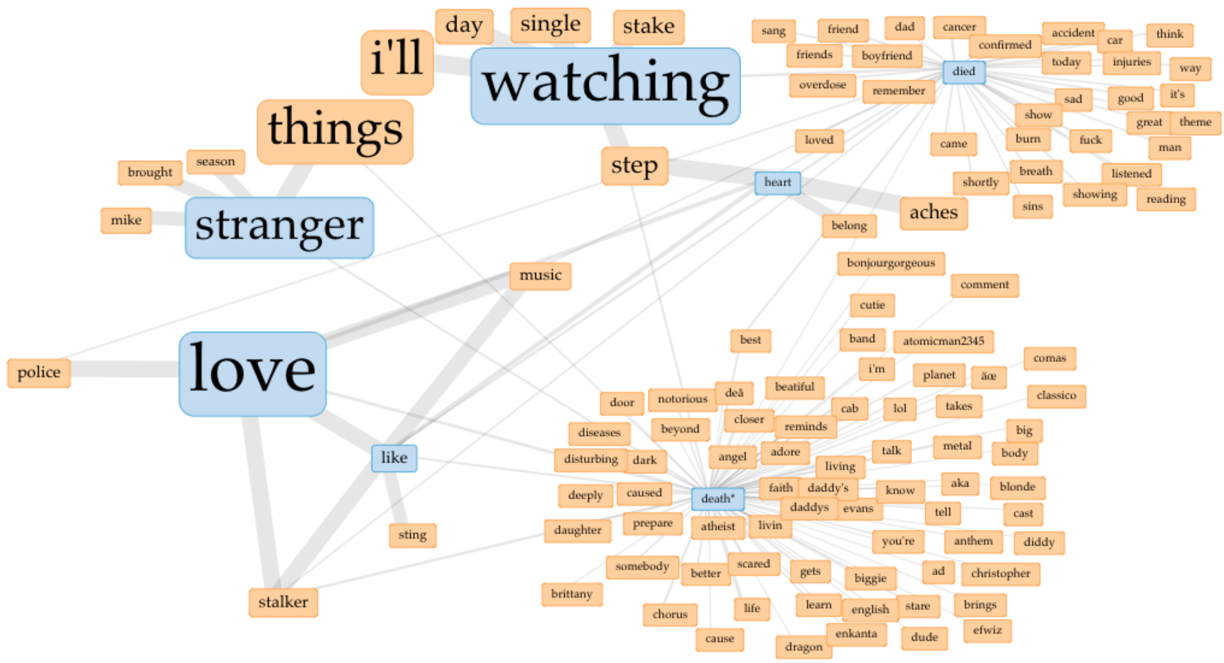


Figure 7. "Death" cluster

"Death" actually constitutes a significant cluster (s. Figure 7) – I plotted it next to "love" and "watching" so as to give an idea of its relative dimension. Around "death," we find many terms that suggest that people seem to be effectively talking about beloved ones that passed away or whose lives are in danger (e.g. "daughter," "prepare," "disease," "scared," "reminds," etc.). We can even discern some people engaging in discussions about the existence of an afterlife ("atheist," "angel," "faith," "beyond"). And we also find some proper names mentioned with curious frequency: "Daddy," "Evans," and "Christopher." What can this be about? After doing some research, I found that rapper Puff Daddy and singer Faith Evans composed and recorded a song in 1997 entitled "I'll Be Missing You," in memory of rapper Christopher Wallace (a.k.a. The Notorious B.I.G., murdered on 1997), which includes samples of EBYT. In that song, the subject of EBYT is precisely re-elaborated as more explicitly referring to death (e.g.: "It's kind of hard with you not around / Know you're in Heaven, smiling down / Watching us while we pray for you." – Puff Daddy & Evans, 2017)

NSA

Finally, I noticed that most critical articles I found that attempted to analyze these lyrics mentioned at a certain point George Orwell's *Nineteen Eighty-Four* (1949) –novel that portraits a dystopian future where people live under the control and surveillance of a totalitarian state known as Big Brother. In relation to this topic, the "I'll be watching you" can be interpreted as referring to the menace of an institutional authority. I decided, then, to explore to what extent were people prone to endorse this interpretation –which, within critics, appears to be considerably popular.

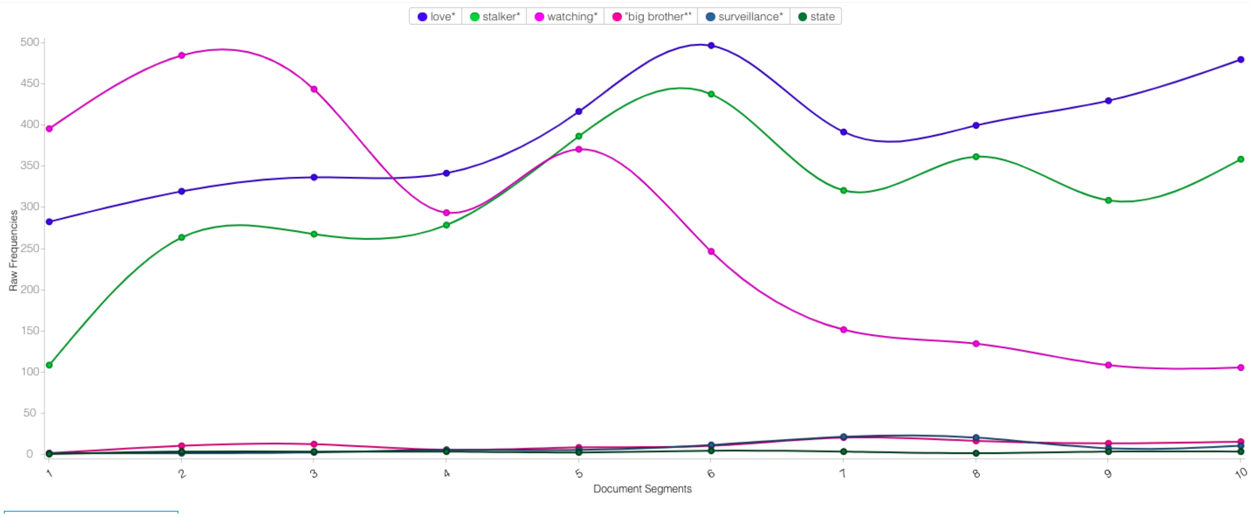


Figure 8. “Big Brother,” “surveillance,” and “state” over time

Nevertheless, as it can be seen in this plot (Figure 8), all the words expectably related to this subject appear to have very low frequency in the comments of EBYT: “Big Brother,” “surveillance,” “state.” Still, exploring the clusters around these words, I discovered a term that does appears connected to all these ones with a more significant frequency: “NSA.”

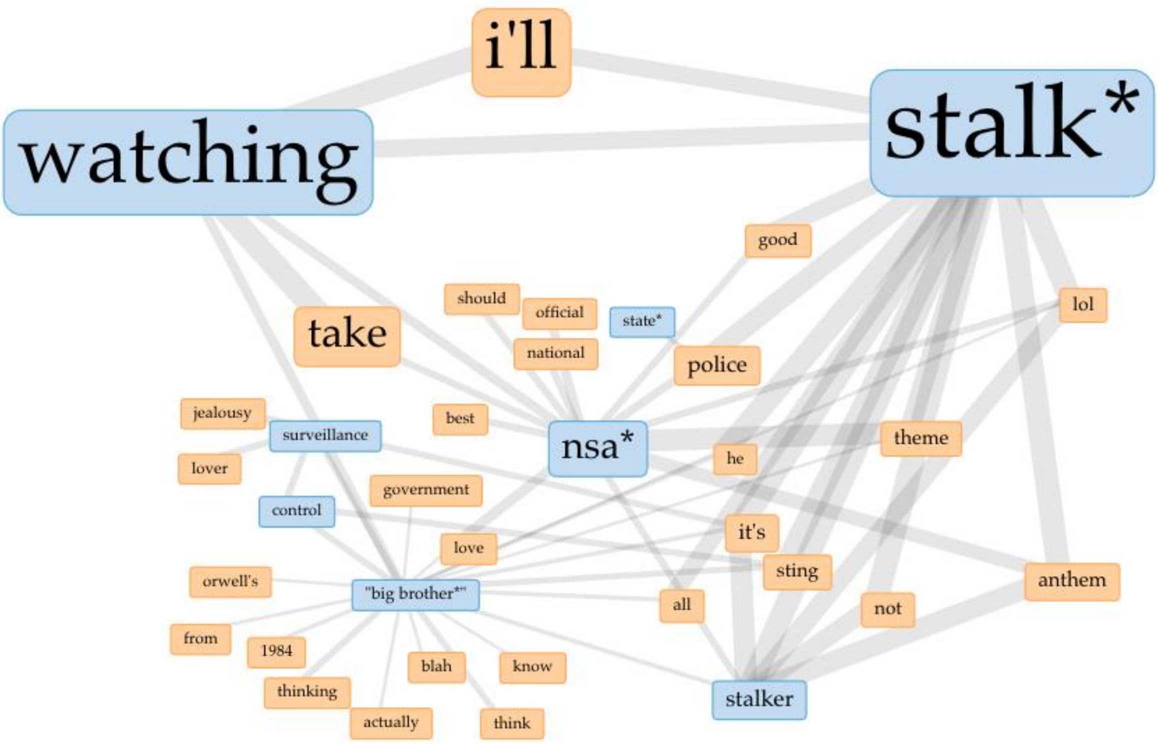


Figure 9. “NSA“ cluster

NSA appears as the central component of a cluster of words that includes “big brother,” “government,” “surveillance,” “control,” “stalker,” etc. (s. Figure 9).

But, as it can be seen in a time plot (Figure 10, sky-blue curve), this word was not always frequent in the comments of EBYT. It emerged at a specific time (from segment 15, on), and it grew so much that it even surpassed for some months the frequency of “watching”–even though it never reached the frequency of terms such as “love” and “stalk”. The precise moment when this occurred is July 2013.

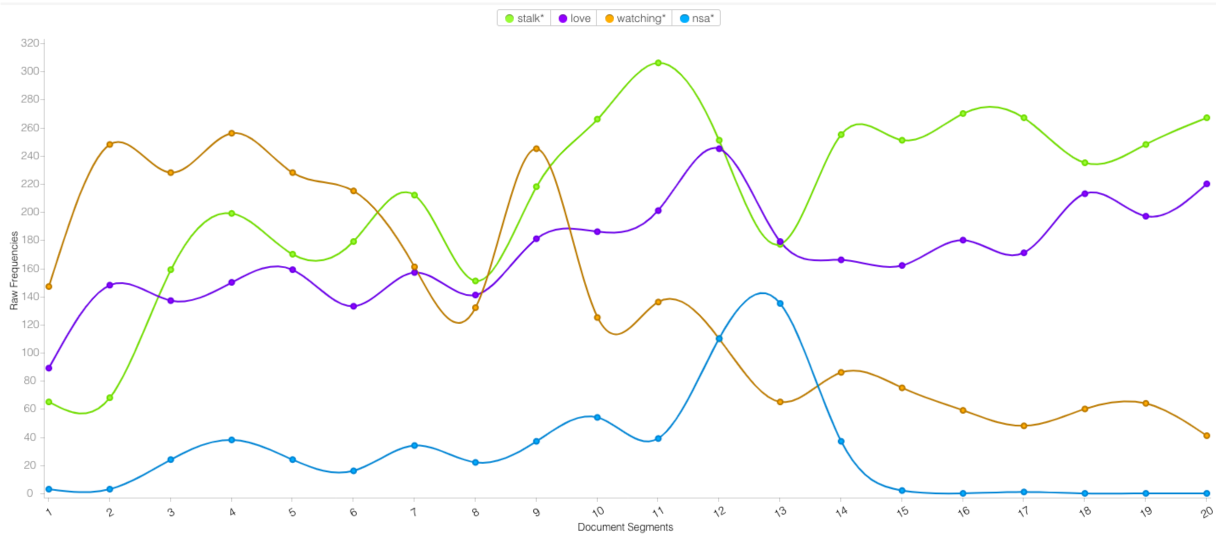


Figure 10. “NSA” over time

What happened in July 2013? The answer is: The Global Surveillance Disclosure. Former CIA employee Edward Snowden revealed classified documents that showed that the USA National Security Agency (hence, NSA) had been secretly and illegally collecting data from the most powerful countries in the world. We can verify the world-wide repercussion of this event by observing how much the frequency with which people searched “NSA” and “Snowden” in Google Search Engine peaked precisely in 2013 (s. Figure 11):

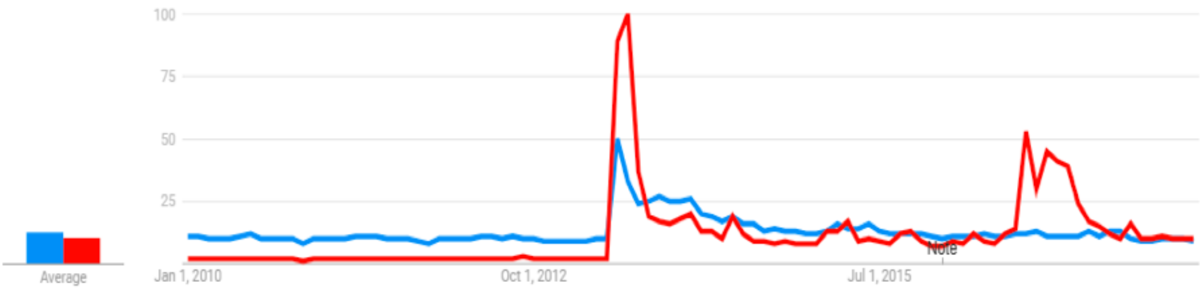


Figure 11. “NSA” (blue) vs. “Snowden” (red) – Google searches over time

As it can be seen, with the appropriate quantitative data, we can visualize not only the proportion of this event in the public opinion, but also its particular repercussion in the interpretation of a particular text: when this event occurred, people effectively started to give more preeminence in the comments of EBYT to interpretations that linked the lyrics with topics like state control and surveillance, even though it remained as a marginal subject in relation to other topics such as "love" and "stalking."

This is how these statistical analyses allow us to discover what topics people are talking about in relation to a given text (in this case, stalking, love, stranger things, death, NSA), to weight the proportion of each of these topics, and to understand the connections that get formed among them.

DATA-BASED PREDICTIONS ON TOPIC ATTRIBUTION

Up to this point we have been making statistics of words mentioned in the comments of EBYT's YouTube video. These statistics, by themselves, are not a proof of how people tend to interpret the text in question, but they are representative data on the basis of which we can elaborate informed hypotheses. Never before in history have we had so abundant data about people's reactions to a single text –in this case, half a million words of comments, freely written by random individuals across the world in a span of eight years. This map of heterogeneous words with varying frequencies through time is where a research on the current meaning of EBYT should begin:

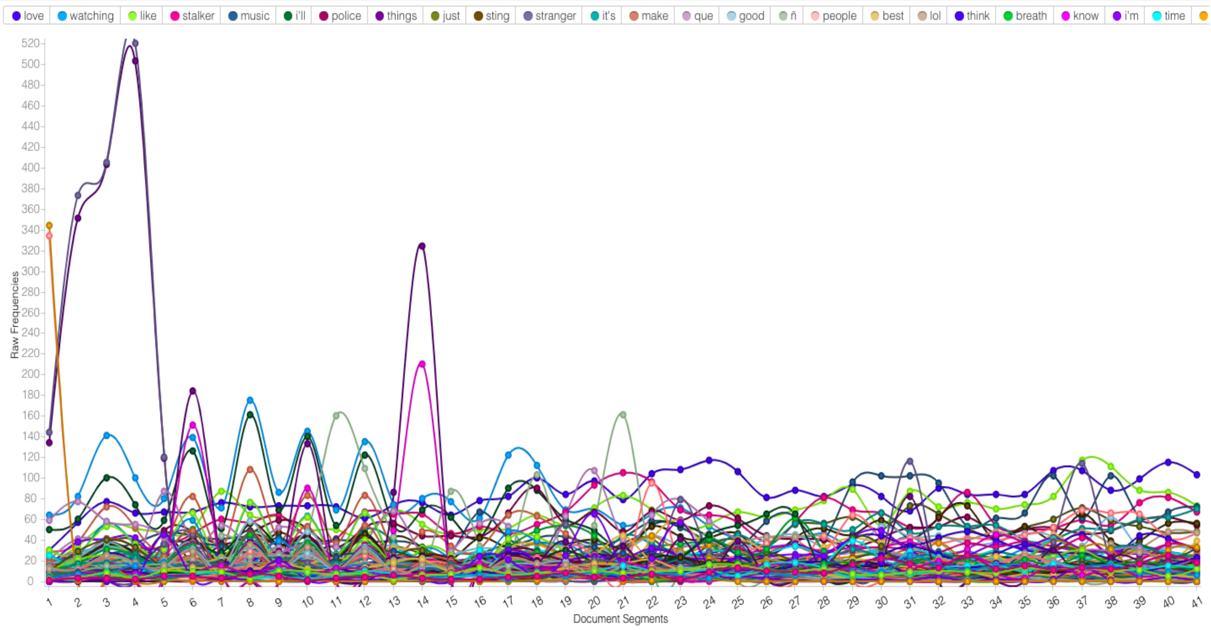


Figure 12. General map of words by frequency over time

Neurolinguists have shown that one of the ways in which our mental networks of concepts get formed is by a cognitive statistical processing. In concrete, this means that the more frequently you perceive two words as appearing close to each other, the more these words will be linked in your mind (this principle is called *Hebbian Rule*), so that, whenever one gets activated, the other one increases its priming (i.e. becomes more prone to get activated).[43] This process is naturally very complex and, for this reason, no single corpus of texts can be taken as a direct reflection of people's mental networks. This being said, the map of words by frequency that we've drawn out of the comments of EBYT (Figure 19) –which registers the

[43] For an experimental exploration on the neurological basis of distributional semantics, s. Carota et al., 2017.

reactions of thousands of people in a period of eight years– is perhaps representative enough to work as an indicator of the average mental connections of concepts for this social group in relation to this song. That is, perhaps the stats of these comments really reveal a significant aspect of people's actual subjective tendencies. Maybe that seemingly chaotic plot really captures a significant trace of the enormous and varying polysemy of that small text for this span of time, a trace that can only be seen through the lens of quantitative analysis.

After analyzing these data, I hypothesized that the main topics that people will be prone to attribute to EBYT are following ones (in terms of their relative proportions):

Love: 36.6%,
Stalking: 45.2%,
State surveillance and control: 9.6%,
Death and afterlife: 8.5%.

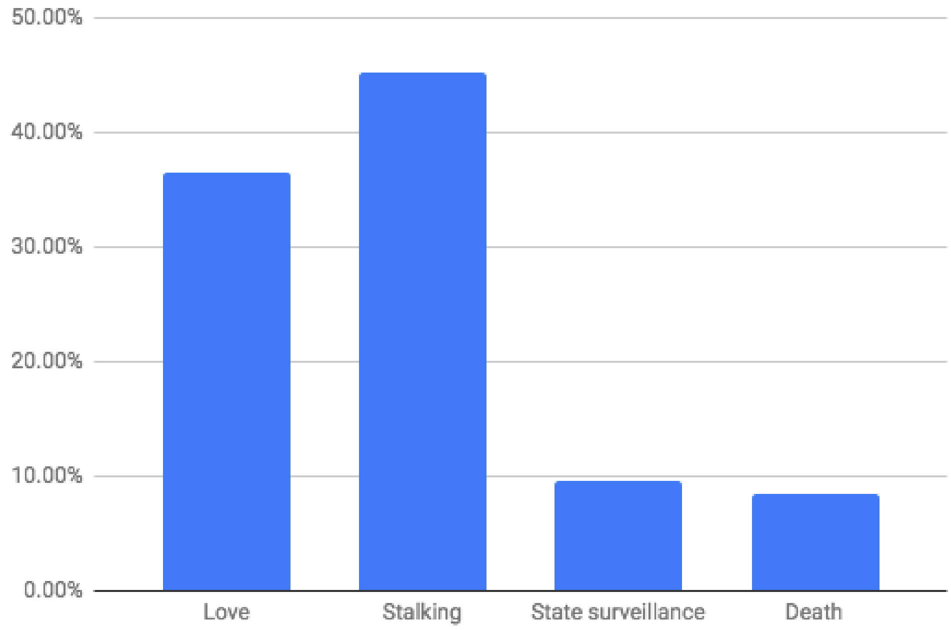


Figure 13. Predicted likeability of each topic to be attributed to EBYT (on the basis of each-topic's frequency of appearance in the comments of the video).

These scores have been calculated by measuring the size of each cluster of words and calculating the relative frequency of each cluster's central term in relation to its average frequency in English (Words Frequency Data, 2018).

Now, how can we verify if this is actually the case? The most direct way is to ask people directly how they interpret the lyrics of this song, so that is what I did next.

CHAPTER 12

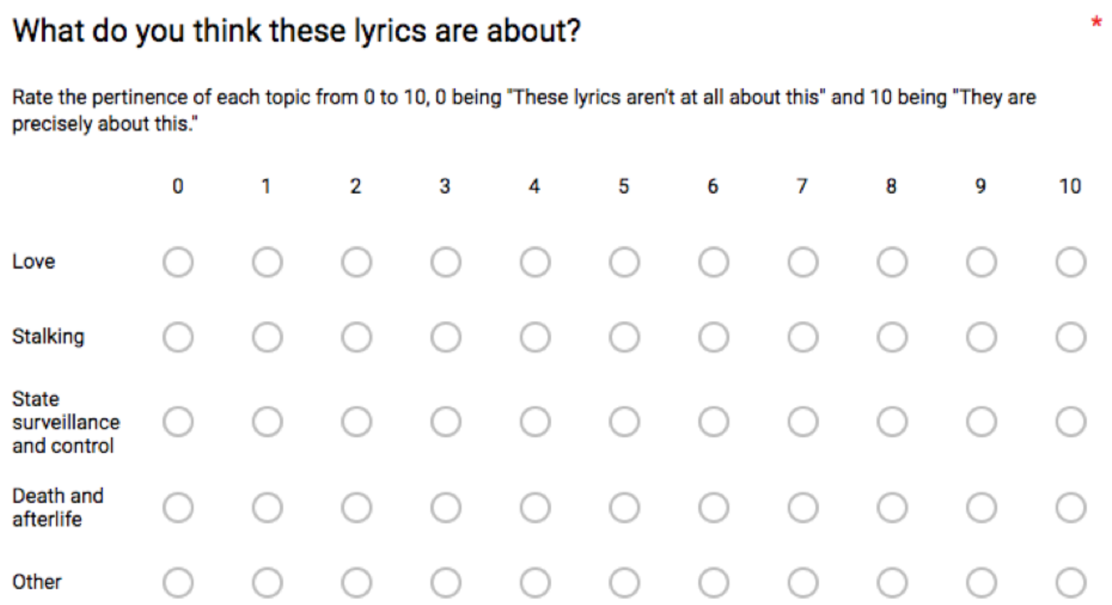
CASE STUDY – PART 2

VERIFICATION AND DISCUSSION

THE SURVEY: TESTING OUR HYPOTHESES

How good of an indicator of people's interpretive tendencies is our statistical analysis of the comments to the video of EBYT? So as to test this, I created an open survey online (during January 2018), in English, that was partaken by a 61 people of different ages (from 20 to 65) and from different countries (mainly from Europe, South America, and North America). First, people had to provide some personal information: age, country, gender, occupation. Secondly, they had to read the lyrics of EBYT. And, thirdly, they were asked a series of questions meant to reveal different aspects of their interpretations of the text.

The first question was precisely aimed at verifying how the participants interpreted the general topic of these lyrics.



What do you think these lyrics are about? *

Rate the pertinence of each topic from 0 to 10, 0 being "These lyrics aren't at all about this" and 10 being "They are precisely about this."

	0	1	2	3	4	5	6	7	8	9	10
Love	○	○	○	○	○	○	○	○	○	○	○
Stalking	○	○	○	○	○	○	○	○	○	○	○
State surveillance and control	○	○	○	○	○	○	○	○	○	○	○
Death and afterlife	○	○	○	○	○	○	○	○	○	○	○
Other	○	○	○	○	○	○	○	○	○	○	○

Figure 1. Survey's first question

As stated in the question, people had to give to these topics scores from 0 to 10, so as to indicate to what extent they interpreted that the lyrics of EBYT talked about each of these things. The order of the options was randomly altered in the question for each participant. And "stranger things" was not included in the survey as a possible topic so as to avoid confusions –since the EBYT was written 34 years before the series was released.

THE RESULTS: VERIFICATION OF MEANING-ATTRIBUTION PREDICTIONS

Averaging the score that people gave to each of these topics, the following results came out:

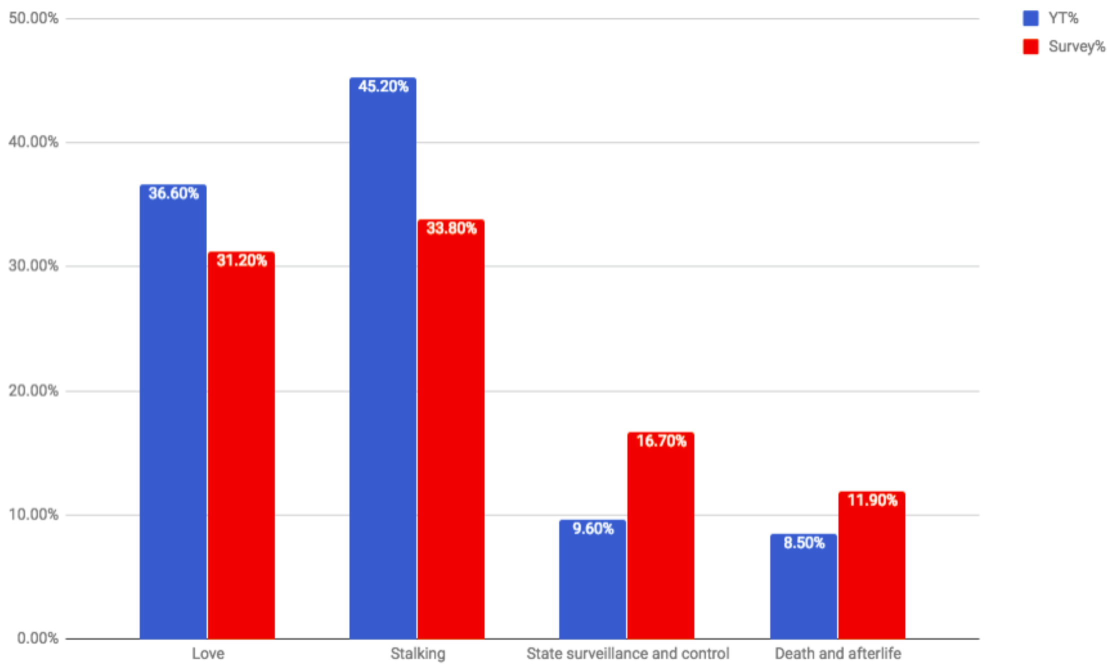


Figure 2. Predictions from comments stats vs. Results from the survey

The blue bars represent the scores of the predictions we made on the previous chapter, based on the result of the text mining of the comments to the YouTube video of EBYT (Figure 2). In red, we find how people actually answered what they thought the lyrics were about. Both results are remarkably similar. In fact, the predictive accuracy of our hypotheses was 96.7%. The correlations between our predictions and the results can be seen clearly by comparing the relative sizes of the different topics: just as the blue bars predicted, the order of the red bars (from the largest to the smallest) is also "stalking," "love," "state control," and "death."

The similarity between our predictions and the participants' actual interpretations can be more clearly visualized in a scatter graph:

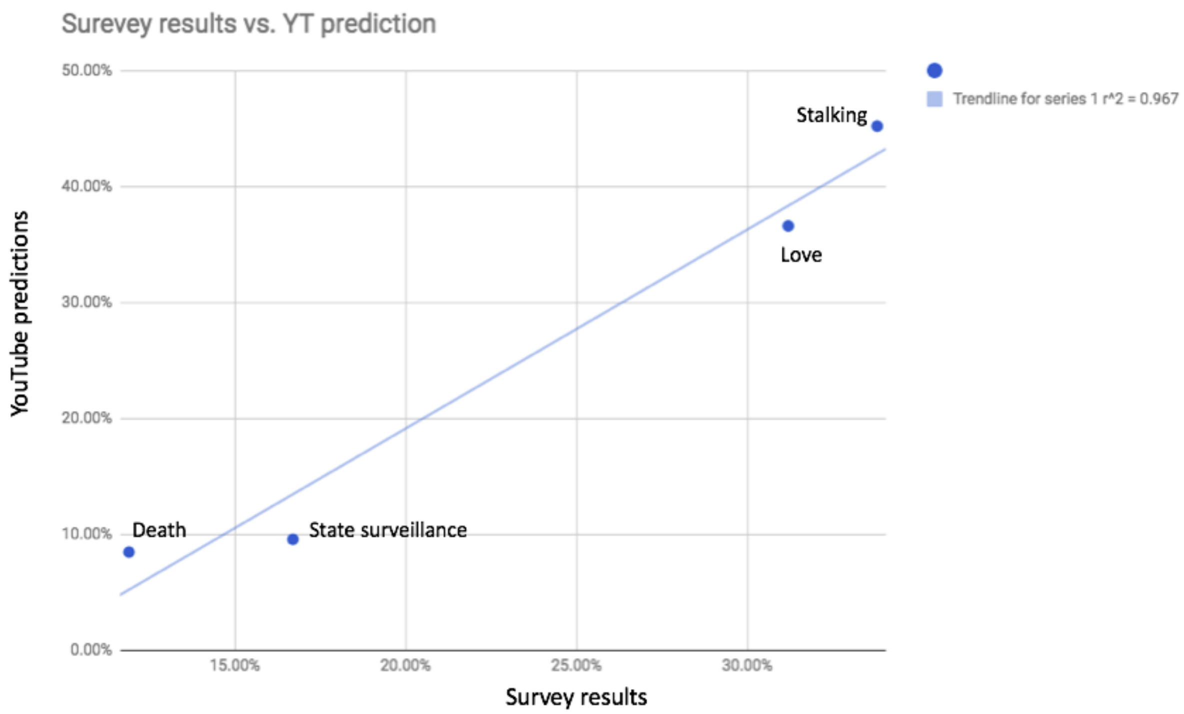


Figure 3. Predictions from comments stats vs. Results from the survey (scattered)

The horizontal axe shows what people declared in the survey to have interpreted, and the vertical axe shows our predictions (Figure 15). The accuracy of our predictions can be verified in the sharp angle of the trendline and the closeness of the dots to it. The r^2 coefficient is in fact 0.967 (1 meaning a perfect positive correlation, 0 no correlation, and -1 perfect negative correlation).

HOW MUCH MORE ACCURATE ARE THESE PREDICTIONS THAN OUR MERE INTUITIONS?

Could we, as readers, foresee these results by means of our mere intuition? Even if that was the case, this measurement still has the value of showing us how to convert our own interpretations of a text into data-driven hypotheses about how others might interpret it, and how to test the degree of validity of our hypotheses against samples of real readers – something that literary studies and cultural studies very rarely are able or even interested in doing. Nevertheless, I still wanted to obtain some estimation of how our predictions scored in comparison to people's intuitions. After all, the main purpose of this approach is precisely to improve our mindreading skills.

In consequence, in the next question of the survey, I asked precisely that to the participants: "What do you think that *others* interpret these lyrics are about?" The question was laid out similarly to the previous one: Participants had to give a score to each topic from 0 to 10 (0 meaning "Nobody thinks these lyric are about this" and 10 being "Everybody thinks these lyrics are precisely about this"). Here are the results of the participants' guesses about others' interpretations:

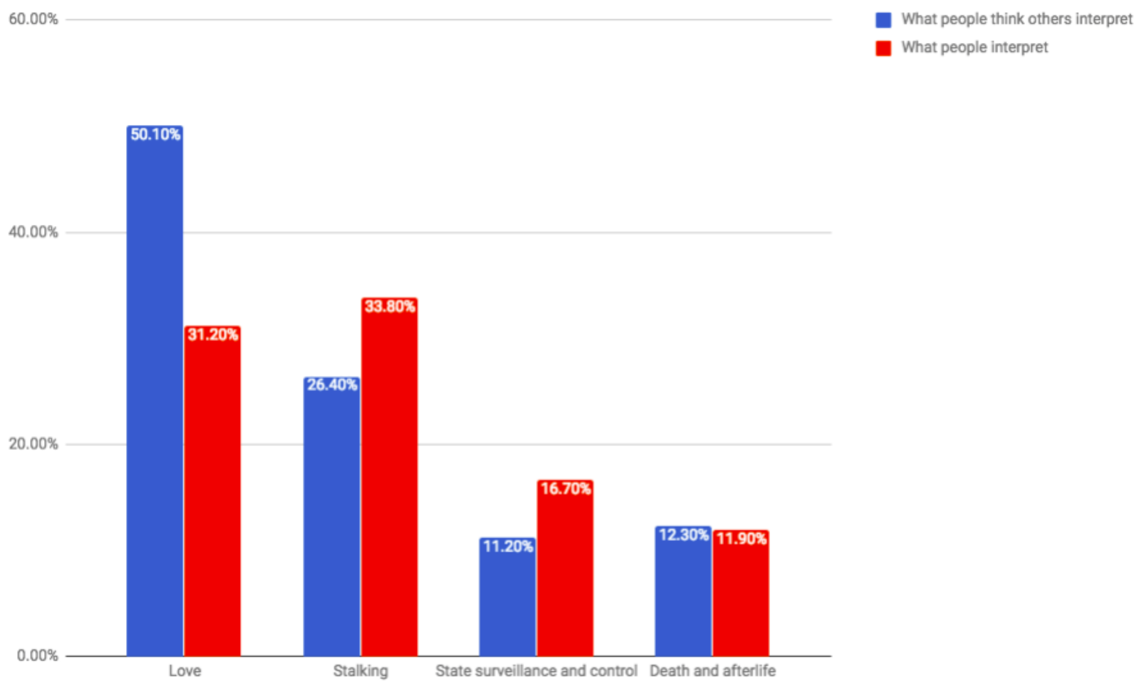


Figure 4. People's guesses about others' interpretations vs. survey results of people's interpretations

In Figure 16, the blue bar represents what people guessed about others' interpretive tendencies and the red bars represent people's actual interpretative tendencies (as measured in the previous question of the survey). Participants did not do very badl: in average, they correctly predicted that people think the song talks more about "love" and "stalking" than about "state control" and "death." But they failed in the more fine-grained comparisons: Participants tended to assume that people relate EBYT with "love" much more than with "stalking," and with "death" more than with "state control," whereas the actual case is the opposite in both comparisons. Here's how people's intuitions look in a scatter graph, in contrast with the actual case:

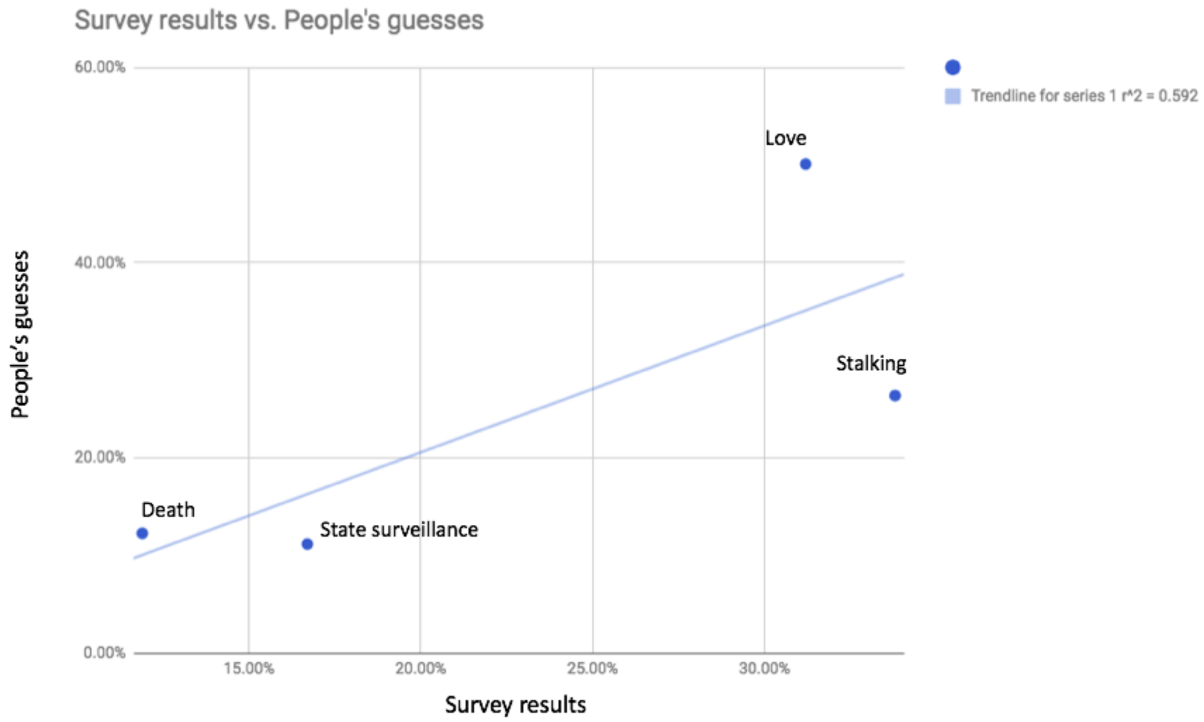


Figure 5. People's guesses about others' interpretations vs. survey results of people's interpretations (scattered)

As it can be seen in this scatter graph (Figure 17), the angle of the trendline is much less pronounced now and the dots are much further from it, which naturally means a lower predictive accuracy. Indeed, the r^2 coefficient in this case is of 0.5992 (in contrast with the 0.967 that we obtained in our data-based prediction). This means that our measurement outperformed people's intuitions by 36%, in terms of predictive accuracy.

We know we can *see* better or *hear* better with the aid of science and technology. These results are a clear evidence that we can also *mindread* better by integrating to our cultural knowledge and biological intuitions the aid of science and technology.

HOW MEANING AFFECTS TASTE

Let us focus on the main two meanings attributed to the lyrics of EBYT. As the results of the survey revealed, some participants interpreted that the song was primarily about *stalking* and others interpreted that it was primarily about *love*. Do these differences have any influence on other aspect of people's interpretation, for instance, in how much they liked these lyrics?

At the end of the survey, the participants had to state how much they liked the lyrics of EBYT with a score from 0 – 10. I have correlated these answers with the interpretations the participants had done of the lyrics.

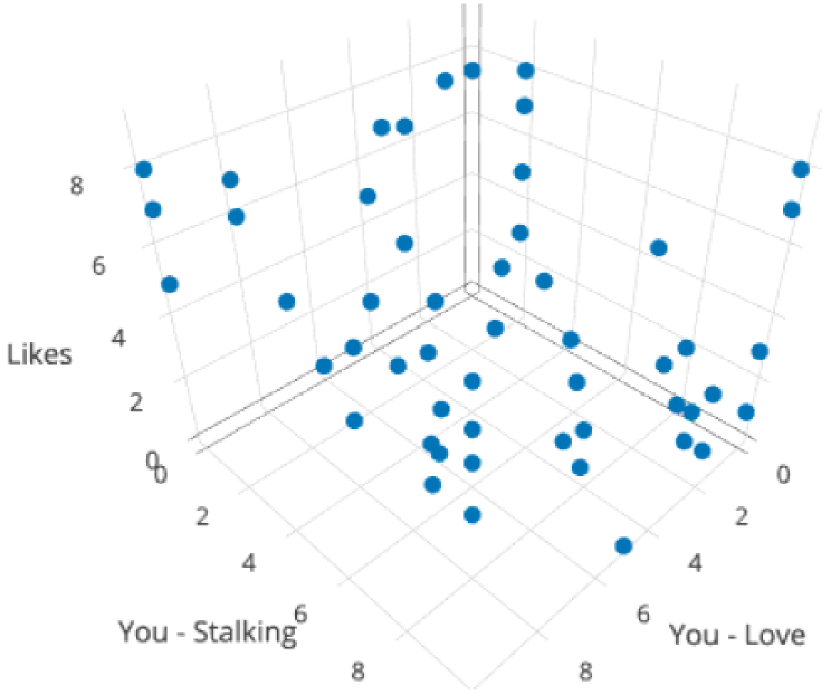


Figure 6. Correlations among Likes (Z), Stalking-Interpretation (Y), and Love-Interpretation (X)

Here is a direct 3-dimensional representation of these three variables: stalking-interpretations, love-interpretations, and likes. Each dot represents a participant in all these three aspects. It is not easy to see the correlations in these complex visualizations, but they became more evident when we focus on the variables one by one.

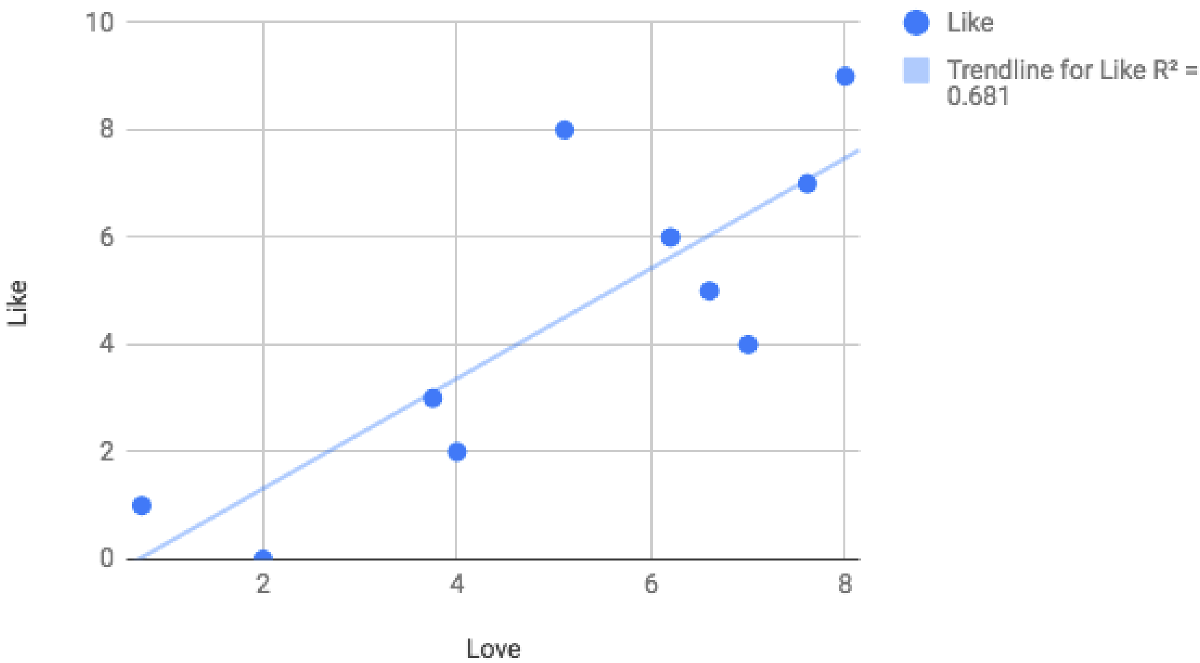


Figure 7. Likes versus Love-interpretation

The first finding can be seen when comparing how much people interpreted that EBYT was about love and how much they liked the lyrics. The love-scores in the graph represent the average of all the answers provided for each corresponding like-score. As it can be appreciated, higher love-scores predict higher like-scores (with an r^2 of 0.68).

Exactly the opposite happened with regards to the stalking-interpretation: the more the participants interpreted the lyrics as talking about stalking, the less they tended to like it.

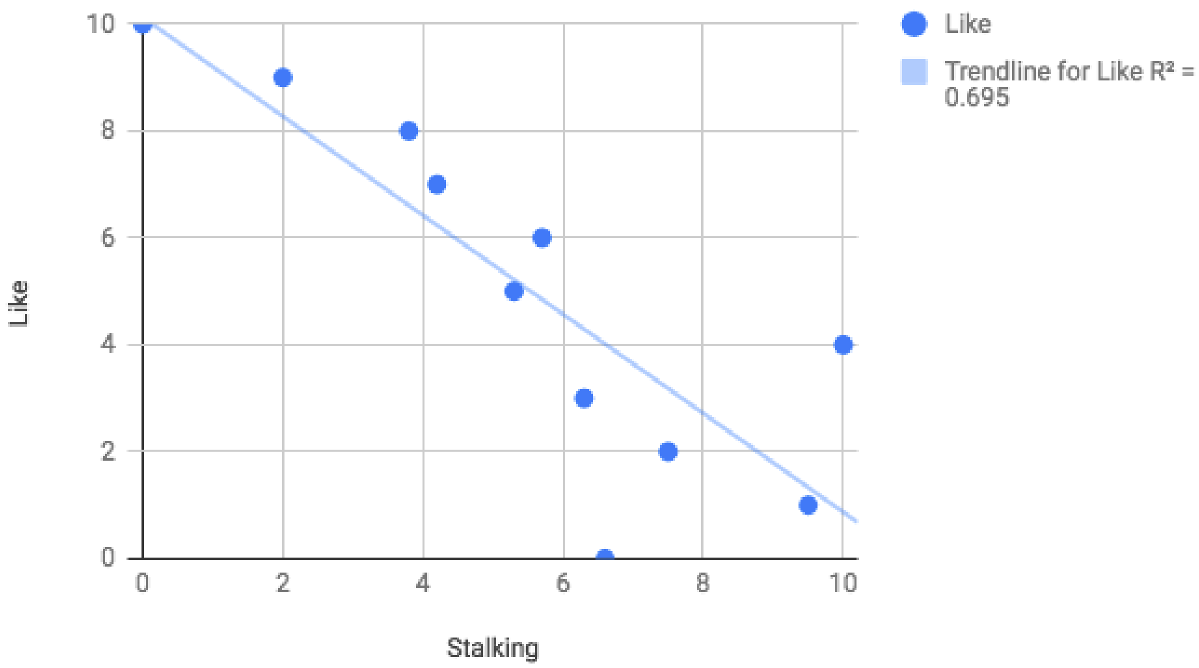


Figure 8. Likes versus Stalking-interpretation

More nuances appear when we consider these trends in relation to the personal characteristics of the participants, namely in relation to their gender. If we analyze separately the answers of male and female participants, we see a significant variance in these scores, especially regarding the stalking-interpretation.

In the case of women, when the negative correlation between the stalking-interpretations and how much they liked the lyrics is increased (r^2: 0.7):

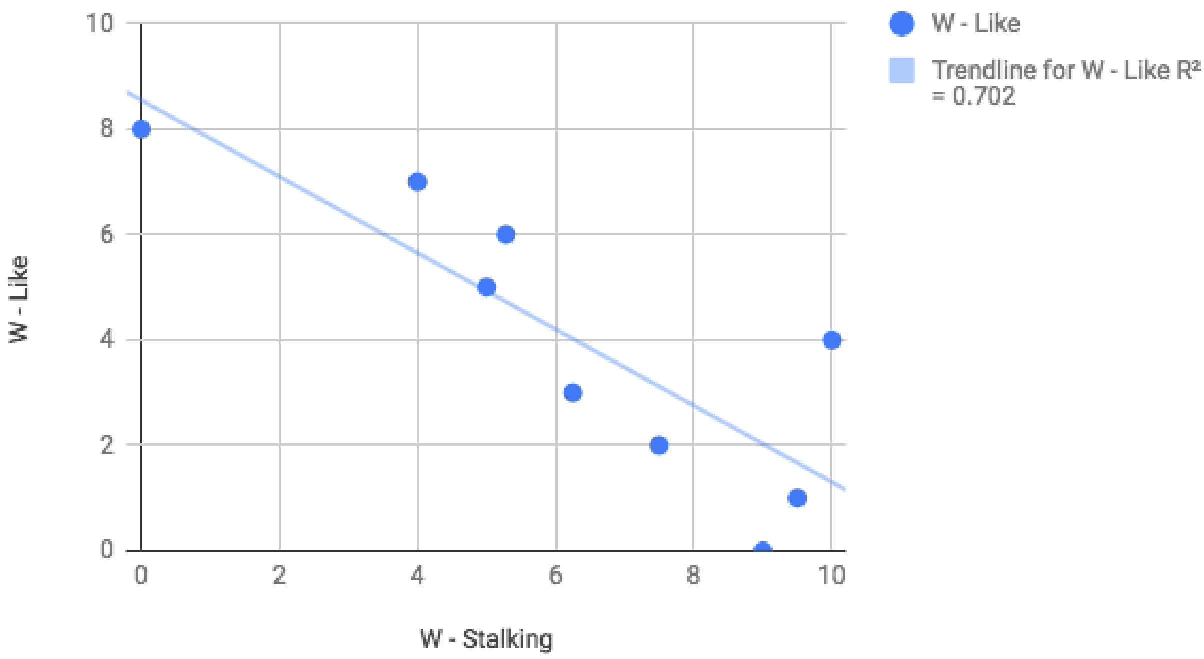


Figure 9. Likes versus Stalking-interpretation in female participants

But, in the case of men, the correlation decreases enormously (to a r^2 of 0.014):

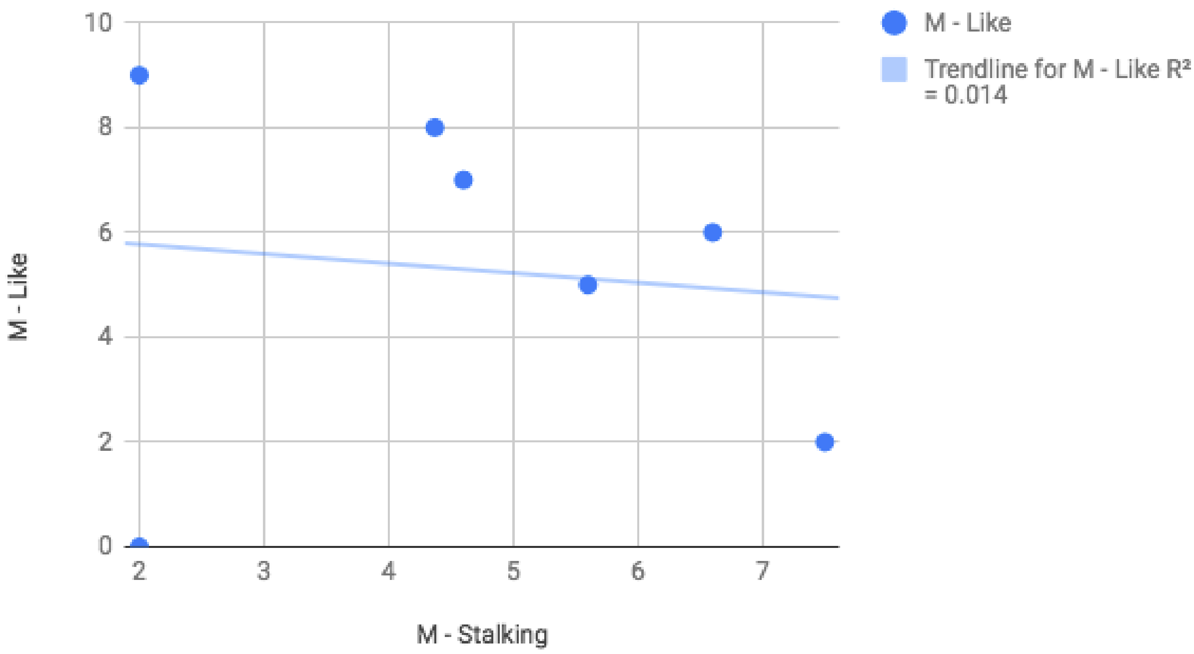


Figure 10. Likes versus Stalking-interpretation in male participants

This means that interpreting that the song is about stalking seems to affect much more the preference of female participants than of male participants for the lyrics.

The same phenomenon is verified in the case of the love-interpretation. In female readers it appears more correlated with the likes than the average (it's r^2 is 0.76):

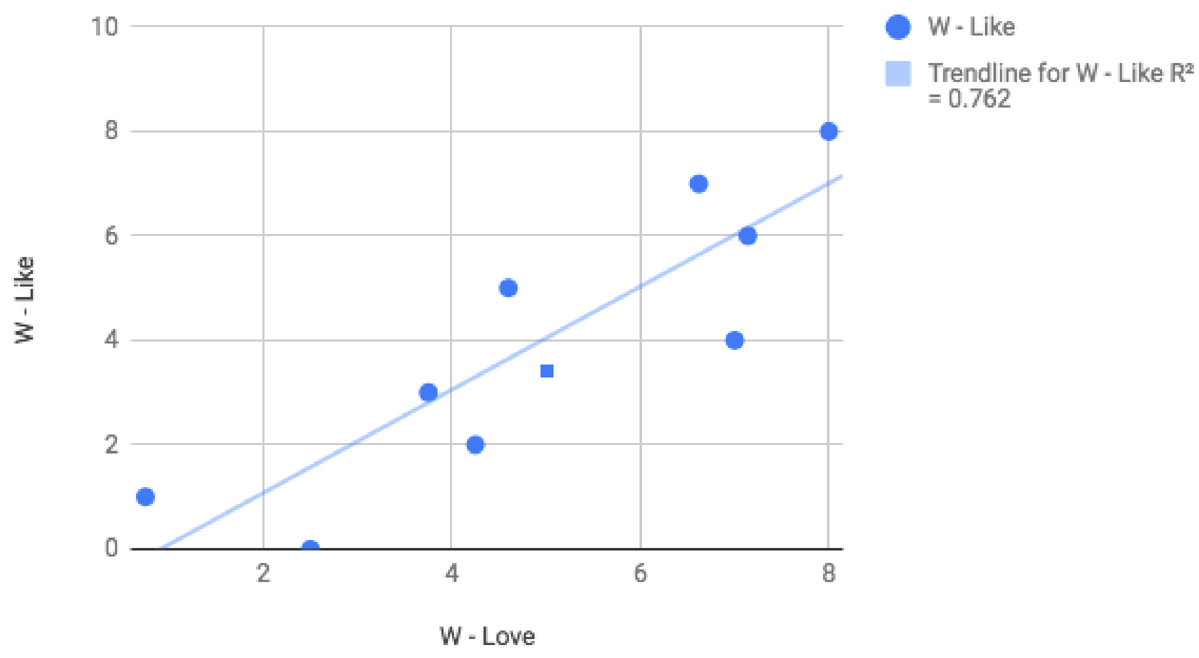


Figure 11. Likes versus Love-interpretation in female participants

Whereas, in men, love-interpretations are less correlated with likes than the average (r^2: 0.51):

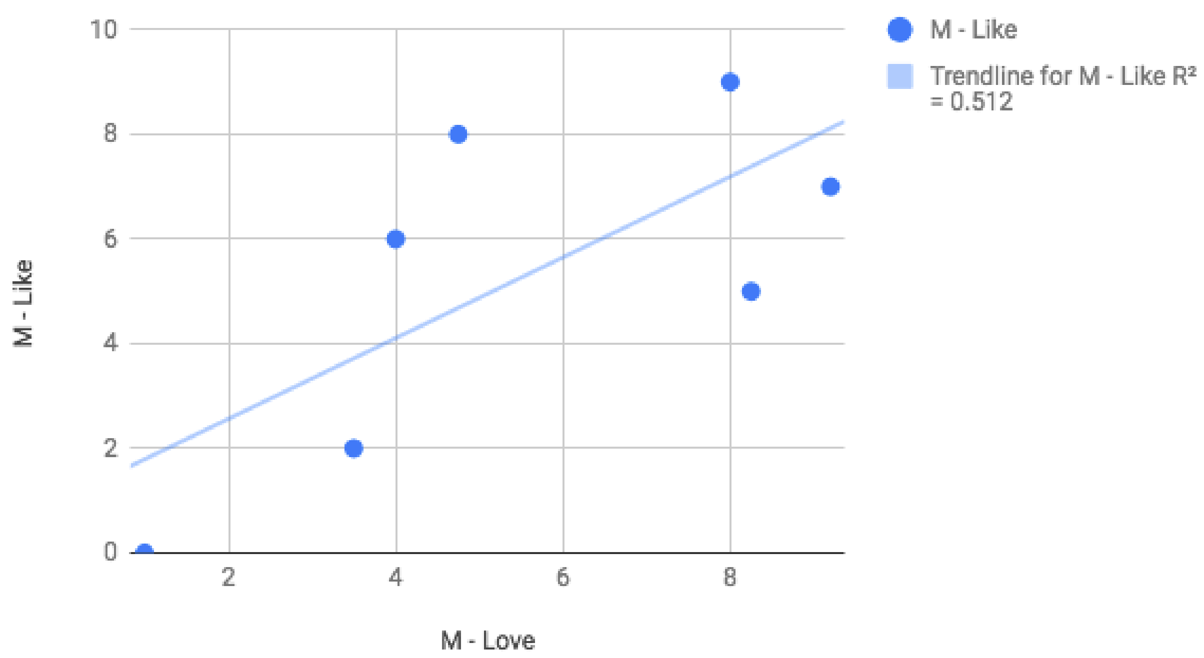


Figure 12. Likes versus Love-interpretation in male participants

A plausible explanation of this phenomenon would be that, since the song was written and sung by a male –and the interpreted topics are related to attraction–, this would suggest a female figure as the recipient (the person addressed by it, the *You* that is allegedly going to *be watched*). This interpretation would make the meaning-distinction love/stalking more pertinent for female readers than for male ones.

Indeed, it can be seen that, in general, in female participants both interpretations (*love* and *stalking*) were more closely correlated with how much they liked the lyrics:

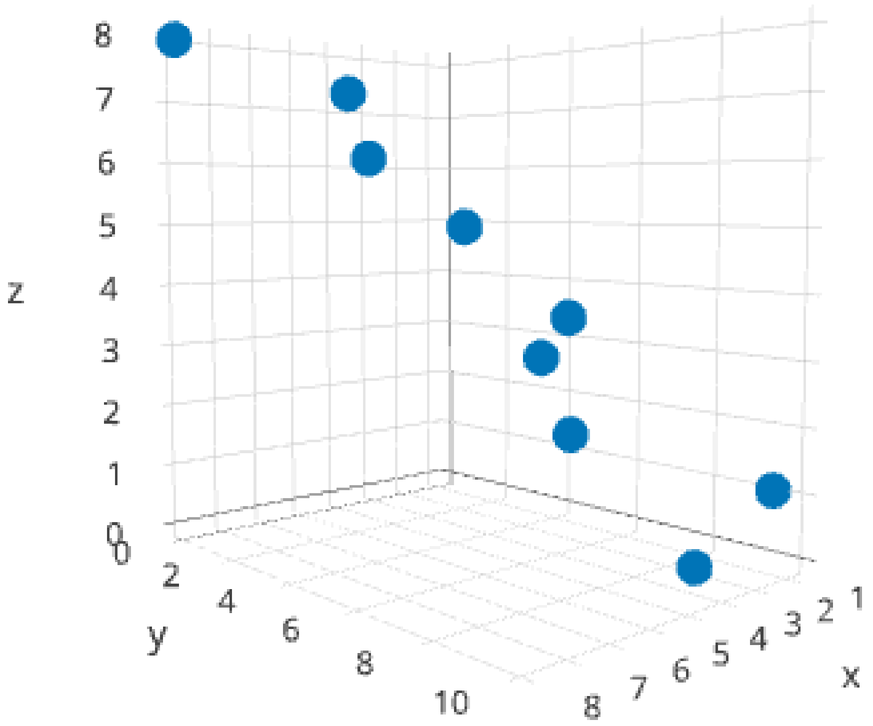


Figure 13. Correlations among Love-interpretation (X), Stalking-interpretation (Y), and Likes (Z) in female readers

Whereas the interpretation of either of these topics in male readers had a markedly weaker correlation with how much they liked the lyrics:

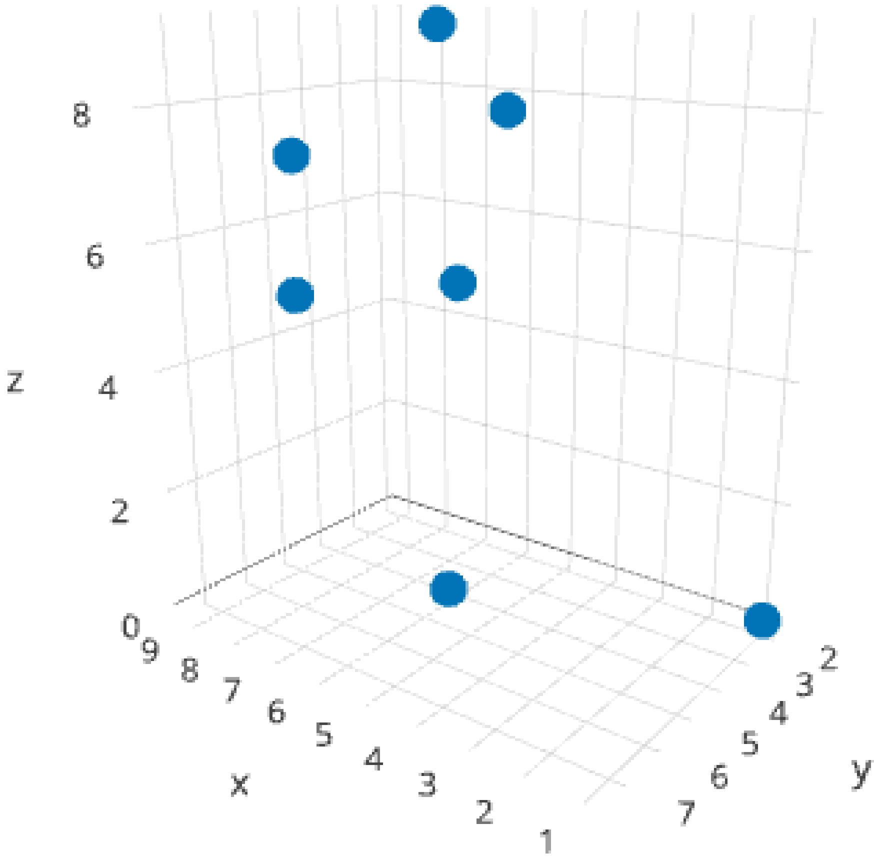


Figure 13. Correlations among Love-interpretation (X), Stalking-interpretation (Y), and Likes (Z) in male readers

At the same time, new aspects emerge when we consider the relationship between these two meanings that the readers attributed to the lyrics.

DOUBLE MEANING: THE AMBIGUITY OF *WATCHING*

Even though some people gave a higher score to the stalking-interpretation and some to the love-interpretation, most of them gave still high scores to both interpretations. But stalking and love seem intuitively like opposing concepts: How can someone think of a text as being simultaneously affectionate and threatening? The fact that these concepts have different values in people's judgment is evidenced by the fact that interpreting one or another clearly affected how much people liked the lyrics. Namely, the love-interpretation had a positive effect and the stalking-interpretation had a negative effect. So, how could the readers hold both contradictory meanings simultaneously? One of the keys to this question might be found in a particular word, the most frequently mentioned words of the comments of EBYT's video: *watching*.

Just like most words, *watching* is a word with many meanings: The Oxford English Dictionary registers four acceptations (2018): it can basically mean *to look at*, *to pay attention*, *to care*, or *to monitor*. When one uses the word, people is supposed to understand from the context to which of these acceptations one is referring. But how can we know what acceptations people are *actually* selecting from a given context? In concrete, how do people exactly interpret the word *watching* in the context of EBYT?

The text-mining measurements from the comments were not clear about which of these acceptations the readers were selecting when reading the lyrics. So, I asked to the participants of the survey: "What do you think *watching* means in the context of these lyrics?," and the different acceptations were offered as alternatives that the participants had to score in terms of pertinence. Here are the results:

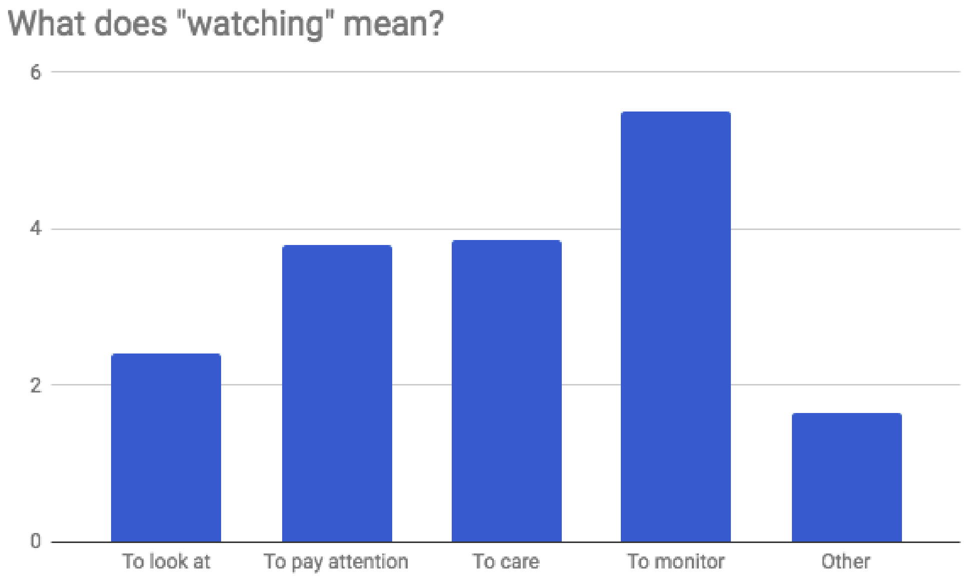


Figure 14. Meanings the participants attributed to the term *watching* as used in the context of EBYT's lyrics.

The first thing to mention is that none of the participants attributed clear-cut 0 or 10 values, in a binary way, to each meaning. Instead, every participant attributed different degrees of pertinence to several alternatives at the same time. This is already very interesting in cognitive terms, since it shows that the way in which a context guides us into choosing a meaning can also be a fuzzy process: if a word has many meanings, maybe when we use the word we do not really chose only one of its meanings, but our minds activate many of them simultaneously, with differences of degree, so that some of them become more present in our consciousness than the others, but followed by a shadow of lower activations of closely-related meanings.

As it can be seen in the graph, the particular acceptations of *watching* that the readers considered as more pertinent in the context of EBYT were: first, *to monitor*, with the highest score; and, in the second and third places, *to care* and *to pay attention*, with almost identical scores. To a lesser degree, people chose *to look at* and *other*.

The fact that the meanings with the highest scores are *to monitor*, on the one hand, and *to care* and *to pay attention*, on the other, seems to be correlative with the meanings that, as we have seen, readers attribute to the text in general: *stalking* and *love*. In other words, the fact that *watching* can be simultaneously interpreted in these diverse ways (monitoring-caring) might be the key that allows readers to interpret the text as simultaneously referring to diverse topics (stalking-love).

DISTRIBUTED MEANINGS: AUTHOR = SPEAKER?

Since double meanings are related to double intentions, I attempted to survey how the participants perceived the author and the speaker of the lyrics.

First, the participants were asked in the survey to what extent they thought of the author and the speaker as being the same person. In literary studies we tend to be quite rigorous about the distinction between implicit author and speaker. But I wondered: To what extent do people actually read in that way?

The participants had to state to what extent they thought of the author and the speaker as being the same person, by giving a score from 0 to 10 -0 meaning “They are *definitely not* the same person,” 10 “They are definitely the same person,” and 5 “It is ambiguous.” Here are the results:

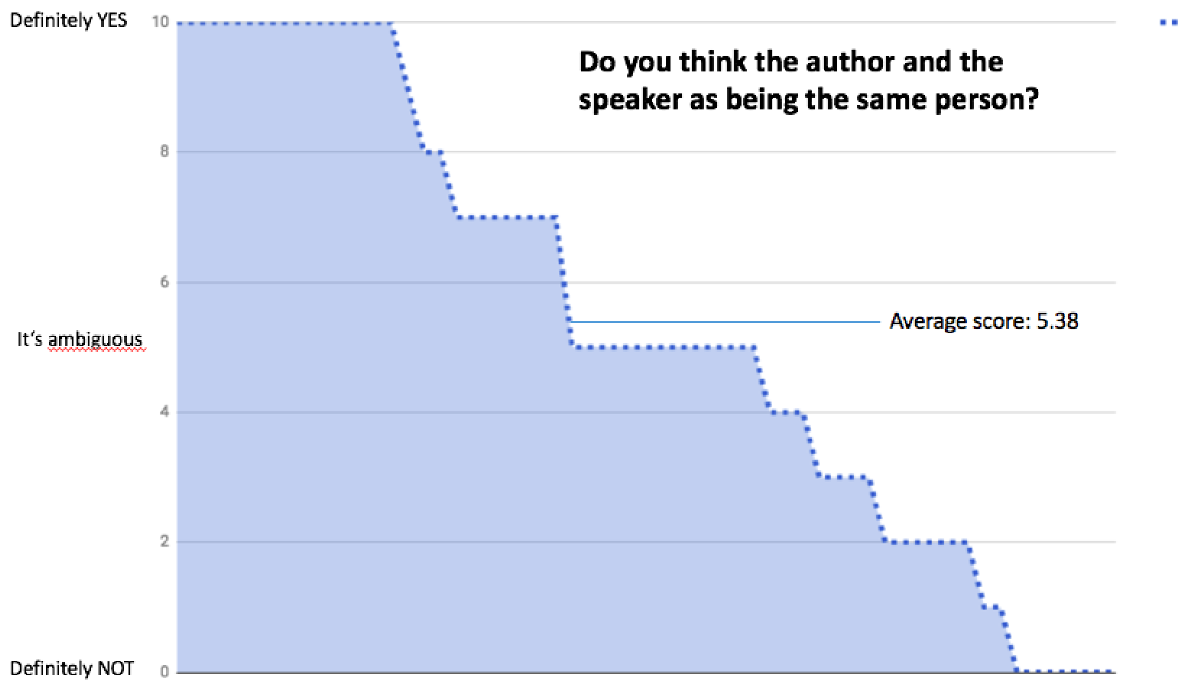


Figure 15. Author-Speaker identity as judged by the participants

As it can be seen, the answers to this question show a very high variance. Many participants answered 0, many answered 10, and most answered with some value in between. Even thought in literary theory author and speaker are clearly distinct categories, people do not seem to take them as such in all of their readings, as these gradient-results suggest. Apparently, distinguishing between who authored a text and the voice that narrates can also be a categorized in fuzzy ways.

However, something interesting appeared when correlating these answers with the meanings that the participants attributed to EBYT and how much they liked the lyrics.

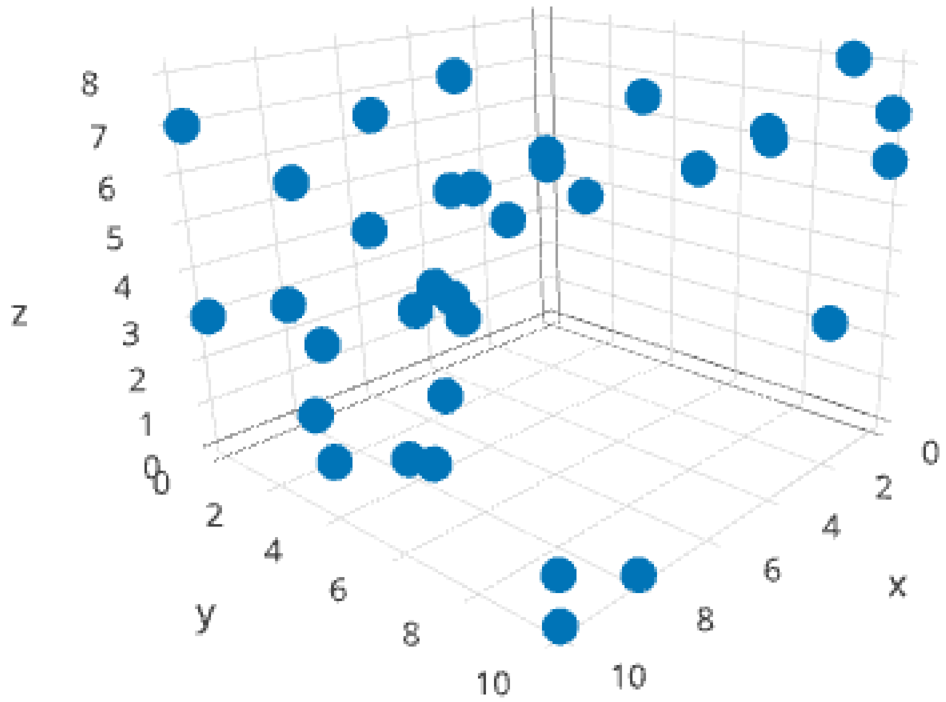


Figure 16. Correlations among Stalking-Interpretation (X), Author-Speaker identity (Y), Likes (Z)

This 3-dimensional graph shows the correlations between author-speaker identity, likes, and stalking-interpretation in females (since female participants had shown the highest negative correlation between stalking-interpretation and how much they liked the lyrics). What we find here is that, despite the strong negative correlation between stalking-interpretation and likes, there were nevertheless some participants who really liked the lyrics (Z: up) despite having interpreted that they talked about stalking (X: front). A commonality of these participants is precisely to be found in the author-speaker identity (Y axe): these participants also tended to consider that the author and the speaker were different characters (Y: left). In fact, the participants who thought of the author and the speaker as the same (Y: right) –and who had interpreted that the song was about stalking (X: front) – were among the ones that liked the song the least (Z: down).

How could these results be explained? How can the distinction author-speaker affect how much readers like a text about stalking? This brought me to reflect on the nature of double meaning.

We use double meaning each time we make veiled statements; for instance, when resorting to indirect speech for making our requests politer: "It would be incredible if you could pass me the salt" (you don't really think it would literally be *incredible*). We use an analogous mechanism for saving our reputation in bribing proposals: "Maybe we can solve this in another way," in sexual proposals that might not be reciprocated: "Would you like to come for a coffee –although, I do not really have coffee…–?" And, as we've seen previously (on the chapter on Emotions), we also use double meaning when making jokes, which allow us to make potentially damaging statements while keeping our face safe, such as in the following example: "A North Korean officer asked my how was my visit to his country: I said I couldn't complain."

Double meaning can be conceptualized as the constitution of a double author-instance. This means that, when a reader interprets two meanings simultaneously in a text, he or she can be led to imagine thereafter two sources of intentions, two different speakers (or two versions of one). Is the speaker in the North-Korean joke simply referring a neutral fact or making a strong criticism? What we think of the speaker's intention is crucial for determining the meaning we will attribute to the statement.

This logic seemed to be pertinent regarding EBYT. As said, we know that most readers interpreted a positive and a negative meaning (love and stalking, respectively), and that the highest score they gave to the negative meaning (stalking), the less they tended to like the lyrics. But this effect was reduced for the participants that considered the author and the speaker as different individuals. So, maybe, what was occurring was that these readers were distributing the double meaning of the song across these two characters. Maybe they thought of *stalking* as "what the *speaker* talks about" and of *love* as "what the *author* meant to talk about."

Indeed, this is effectively what we find when we look at the participants' guesses about the meaning-intentions of the author:

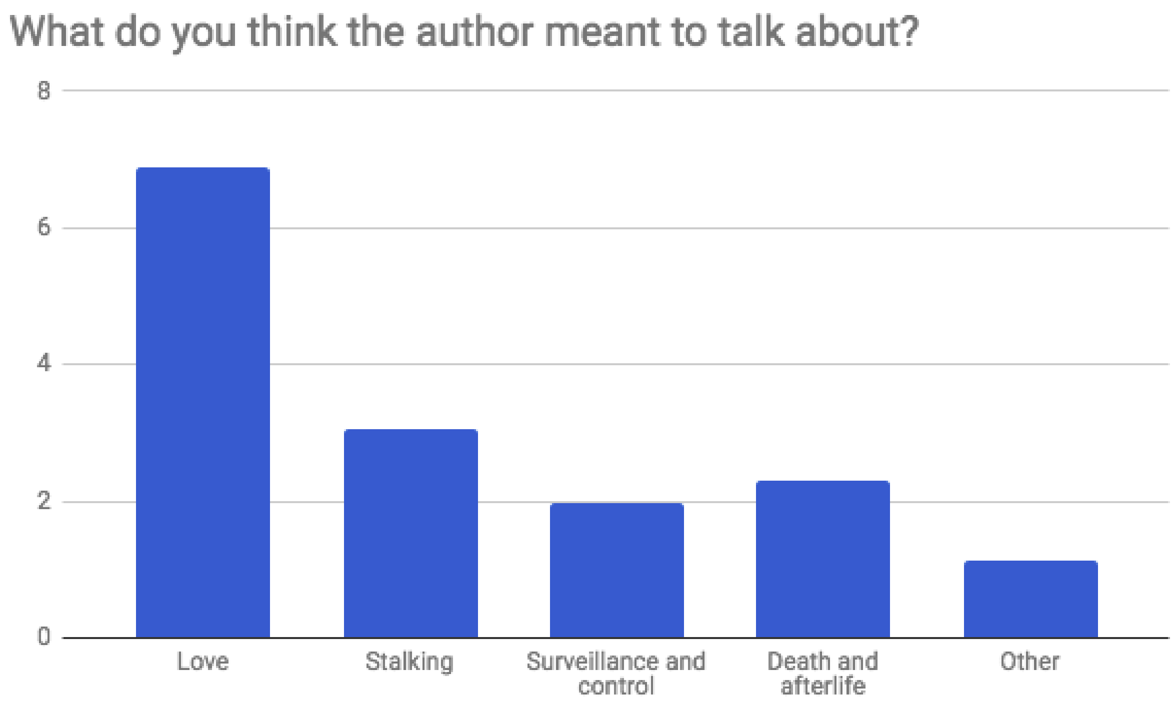


Figure 17. Meanings that the participants judged the author intended to convey in the lyrics of EBYT

Even though in the participant's interpretations the lyrics were mostly about stalking, when asked about the author's intentions, stalking had a significantly lower score, and most participants guessed instead that the author had intended to write a text about love.

We can find further evidence for this hypothesis. If thinking of the author and the speaker as different individuals brought the participants to attribute different intentions to them (author: love; speaker: stalking), this effect should become more evident when the readers be explicitly requested to think of the author and the speaker as different characters and to judge them separately. Breaking up the double meaning this way, we would expect that the participants make a highly positive judgment of the author and a highly negative judgment of the speaker. And that is exactly what happened.

HOW ARE THE CHARACTERS PERCEIVED BY THE READERS? – STEREOTYPES AND EMOTIONS

Participants were requested in the survey to state what they thought of the following characters:

- The author: The real person who wrote the lyrics.
- The speaker: The *I* who says *I'll be watching you*.
- The recipient: The *You* to whom the speaker refers, who is supposedly going to be watched.

So as to measure how people perceived these characters, I employed Susan Fiske's socio-cognitive model of Warmth & Competence (W&C) (which has been exposed in the chapter about characters). In concrete, the participants had to give to each of these characters a score regarding "How morally good do you think he/she is?" and "How intelligent do you think he/she is?" The first question was aimed at measuring perceived warmth; the second, at measuring perceived competence. Here are the results:

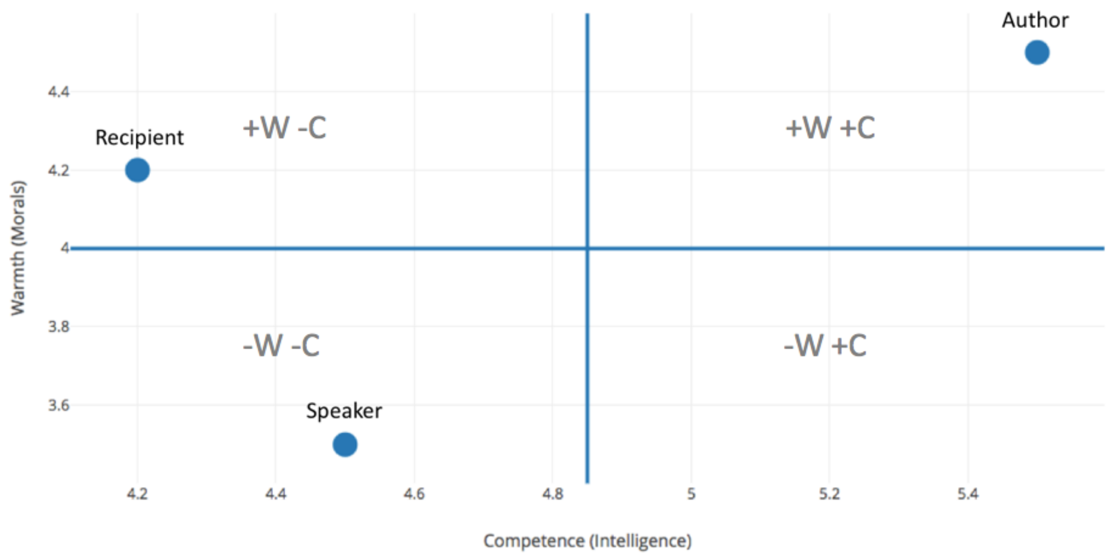


Figure 19. Perceived warmth (W) and competence (C) of the characters of EBYT, as judged by the participants

The answers of the participants were not random but remarkably consistent. Most of the participants rated the author as being very good and very smart, the speaker as very bad and stupid, and the recipient as good and very stupid, in relation to each other.

This shows that making an explicit distinction between author and speaker really radicalizes the ways in which readers judge these characters in systematic ways. Moreover, the stereotypes in accordance to which people judge these characters seem to be effectively consistent with our hypothesis, with the way in which the double meaning of the text is distributed: love-intention (positive meaning) attributed to the author (+W+C) versus stalking-intention (negative meaning) attributed to the speaker (-W-C). Also, this interpretation places the recipient as the object of both love and stalking, which expectedly makes the participants judge it in a mixed way (positive-negative), which was expressed as high in warmth and low in competence (+W-C).

We also know that the W&C model allows us to predict emotions from these stereotypes. Are these predictions also applicable for characters? In a further question, the participants were asked what emotions they felt towards each of these characters. The list included admiration, pride, contempt, disgust, pity, and envy. People were able to chose more than one emotion for each character. If the W&C model was also valid for the perception of characters, then the emotions felt by the participants had to match the expected ones in relation to the stereotypes they attributed to each character. That is, the author (+W+C) had to elicit admiration and pride, the speaker (-W-C) had to elicit contempt and disgust, and the recipient (+W-C) had to elicit pity. And that is exactly what happened:

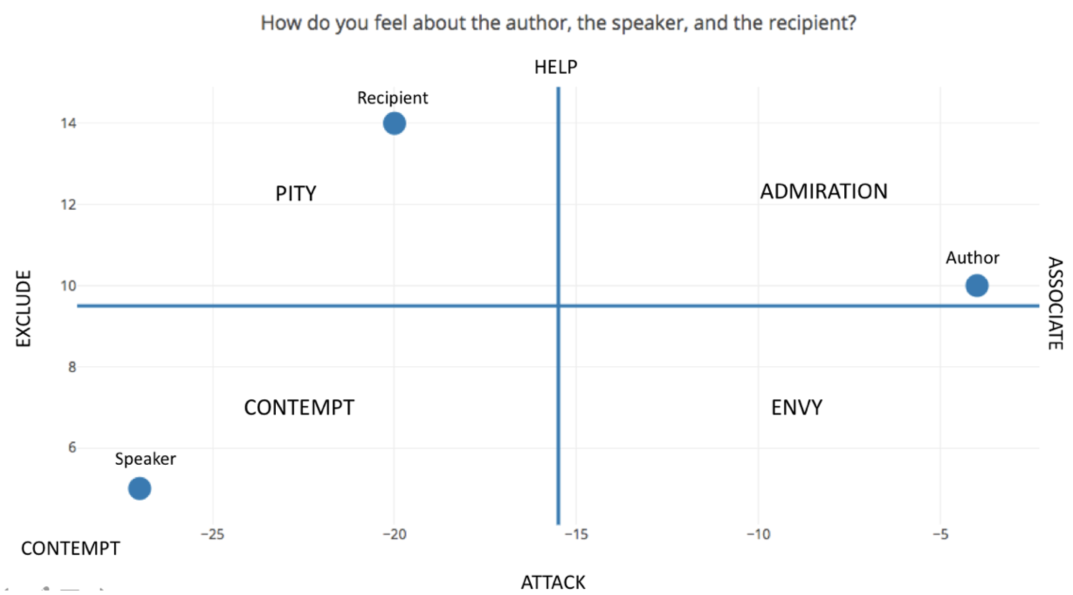


Figure 20. Emotions that the participants declared feeling with regards to the characters of EBYT

As it can be seen, in terms of the elicited emotions, each character was placed in the space that our hypothesis anticipated. The dots are not in the exact same points that the ones in the previous graph –for example, the admiration people expressed for the author is closer to envy than expected in reason of he's perceived warmth, and the pity for the recipient is closer to admiration than expected in reason of his/her perceived competence–, but still, each emotion was successfully predicted.

This is also consistent with our hypothesis of how the double meaning of the song is distributed across the characters: The author (love-intention) produces admiration, the speaker (stalking-intention) produces contempt, and the recipient (object of love and stalking) produces pity.

In the end, EBYT seems to be a song about the dark side of love: about the love drive (*love*) that, when apart from its object (*death*), can turn into obsession (*surveillance*) or worse (*stalking*). The archetype seems to be displayed across the characters' roles, as an implicit plot: the brokenhearted author who copes with his anguish by embodying it in its excess (the stalking speaker) and overtly addressed to the absent object of his love (the implicit recipient). Love, stalking, control, and death… Heartbreaks, indeed, seem to be related with all these topics. Heartbreaks can also make us experience diverse emotions; but, more than anything, heartbreaks are experiences of sadness – which this is precisely the emotion that people reported feeling the most with regards to the lyrics as a whole (s. Figure 21):

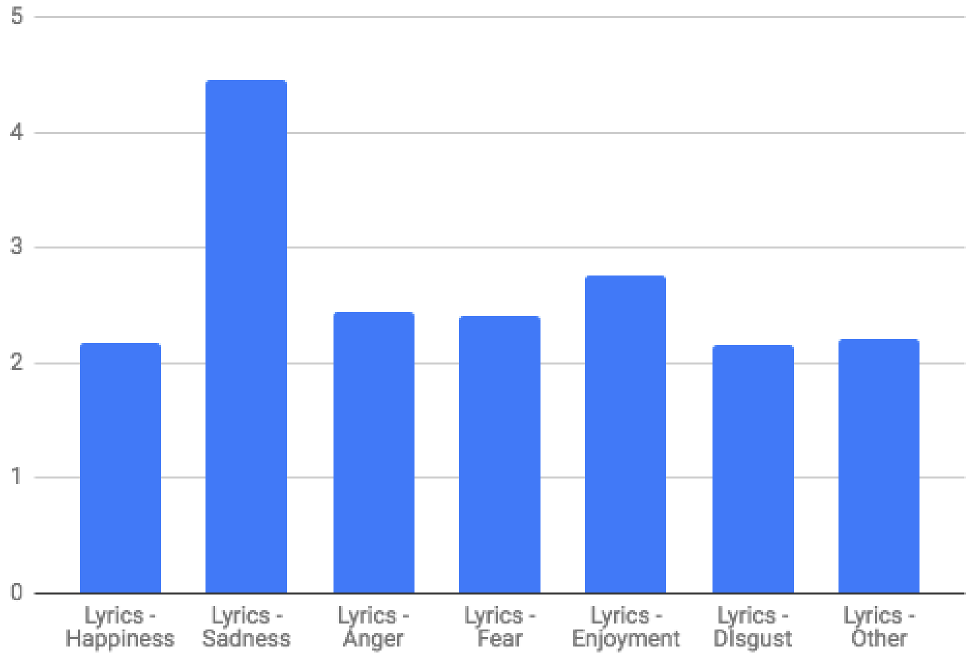


Figure 21. Emotions that the participants declared feeling with regards to the lyrics of EBYT as a whole

A COGNITIVE-DIGITAL LITERARY INTERPRETATION

Let us sum up here what we have discovered about how people interpret the lyrics of this song along these last two chapters. We have engaged in what I would like to call a cognitive-digital literary analysis: the use of digital tools and cognitive models to discover how a literary text is perceived by a society, how a group of people is likely to respond subjectively to particular linguistic stimuli.

First, we have explored how people interpret the lyrics of "Every Breath You Take" (EBYT) (by The Police, 1983), what they think it talks about. We have performed a topic modelling of the comments made by people around the world in the last 10 years to the video of EBYT (+500,000 words). On the basis of these results, we formulated hypotheses about the topics that were likely to be attributed to the text of the lyrics. Through a survey, we verified that these statistical measurements predicted with high accuracy which topics, in which order, and which proportions new readers would attribute to the lyrics of EBYT. Namely, readers tend to interpret that these lyrics talk, 1st about *stalking*, 2nd about *love*, and –to a much lesser extent–

3rd about *state surveillance* and 4th about *death*. We have also seen that these interpretations have a correlation with how much people like the text: namely, people that interpret it talks about *love* like it more, and people that thinks it talks about *stalking* like it less –this correlation is especially marked in female readers.

Secondly, we have explored how readers perceive the rhetorical structure of the meaning of the song. As mentioned, the two main meanings that readers attributed to the song were somehow opposite –in the sense that one was positively correlated with how much the readers liked the song (*love*), and the other was negatively correlated (*stalking*) with like-scores. Nevertheless, most readers gave high scores to both meanings –even if one of them was higher than the other–, that is, they tended to interpret that these lyrics talked simultaneously about a positive and a negative subject. This double meaning was correlated with the double meaning that the readers attributed also to the most frequently commented word of the lyrics: *watching*, which readers judged as simultaneously referring *to monitor* (echoing *stalking*) and to *to care* (echoing *love*). Double meaning, due to its psychological functioning, brought us to consider the intentions that readers' imagined expressed in the text and the subjects that would embody these intentions.

So, thirdly, we have analyzed how people perceived the characters displayed in the text. A subgroup of readers gave high like-scores to the lyrics despite having interpreted that they talked mostly about *stalking* (which, as mentioned, was generally correlated with low like-scores, instead). As a common trait, we have noticed that these readers tended to see the author and the speaker of the lyrics as two different individuals (as measured in the survey). After this observation, it was hypothesized that, when a reader makes this conceptual distinction between the author and the speaker as mental representations of separate individuals, then, in the readers' minds, the double meaning allowed by the ambiguity of the lyrics would get distributed across these different imagined subjectivities, bringing the reader to assign each meaning to the intention of a different character.

The first evidence of this hypothesis was revealed by the participants' answers to the question: "What do think the author meant to talk about in these lyrics?" These same readers that had judged that the lyrics were mostly about *stalking*, answered here, instead that the author had intended to talk about *love*. If they considered the lyrics talked about stalking, but the author had meant to talk about *love*, who was supposed to embody the stalking-intentionality? The answer had to be: in the speaker, the intra-diegetic authorial character built up by the author.

For verifying this, the participants were requested to judge the author, the speaker, and the recipient, as separate characters, with regards to warmth and competence. And, indeed, the results effectively verified that, when forced to make these distinctions among the different characters, people consistently judged the author as high in warmth and in competence (+W+W) (which is consistent with the *love*-intention), the speaker as low in warmth and in competence (-W-W) (which is consistent with the *stalking*-intention), and the recipient as high in warmth but low in competence (which is consistent with the stalking-victim role). This meant that, when distinguishing the author from the speaker, the readers effectively tended to attribute the *stalking*-intentionality to the speaking character, and the *love*-intentionality to the authorial character.

Fourthly, the socio-cognitive model employed allows us to predict emotional reader-responses on the basis of the stereotypes that result from people's judgments of warmth and competence. In accordance to these predictions, the author (being +W+C) would elicit admiration and pride, the speaker (-W-C) would elicit rejection and disgust, and the recipient (+W-C) would elicit pity. These emotions would be precisely consistent with the role-distributions in terms of intentional meaning: the author as the love-intending subject, the speaker as the stalking-intending one, and the recipient as the stalking-victim. These emotional responses were also exactly what the results of the survey showed after the participants were requested to indicate the kind and extent of the emotions they felt to each of these characters, separately presented. The participants effectively reported feeling admiration for the author, rejection for the recipient, and pity for the recipient, which matched the corresponding stereotypes of W&C for each character. After this evidence, a more qualitative kind of interpretation was formulated: "Every Breath You Take" seems to be perceived by the readers as displaying an elaborated form of the idea of a heartbreak: the admired broken-hearted author that channels his anguish by embodying it magnified into a scene of obsession.

What I have tried to show with this case study is that a cognitive-digital approach to literature does not exclude the interpretive task (which is the core of literary criticism). The fundamental novelty of this cognitive-digital approach is that allows us to transform these interpretations in testable hypotheses that can be explored at a large scale. Thereby, this perspective has allowed us to produce a quantitative empirical description of the dynamic effects of a literary work across a society, as well as a qualitative interpretation to be judged in terms of its coherence with the data found. It has allowed us to mapped a segment of our minds and our culture.

AFTERWORD

Descriptive literary studies aim at mapping our minds and culture. Asking what interpretations and emotions a text might elicit in a society is a question about subjective facts, but it is still an objective question, a question that can be valued in terms of truth. For this reason, descriptive literary studies must not only formulate interpretations of literary texts, but also convert these interpretations into testable hypotheses and verify them empirically. In this sense, even though descriptive literary studies differ in their perspective from literary criticism, their task is not opposed to it but complementary. The advantages I see in cognitive science and data science are precisely that they provide us with a framework to formulate these hypotheses and tools and methods designed to empirically test them. The possibility and productivity of such a framework is the first thesis that I have tried to present in this book.

In the first part of this book, we have considered a series of cognitive models that give us valuable insight about the nature of many of our reader-responses, of the ways in which we are moved by literary works and other linguistic stimuli. We have analyzed, from this cognitive perspective, how our minds process words, rhetoric, narrative, characters, and emotions.

In the second part of this book, we have explored the digital tools developed by data science and the possibilities they offer us to study language and people's subjective tendencies at a social scale. We have considered, on this regard, the value of quantitative analysis for cultural research, the literary studies that are performed with text-mining techniques, and even some of the verbal technology that is being produced –as an effective outcome of this kind of research– in the field of artificial intelligence.

In the third part of this book, a case study was presented, where these two perspectives (cognitive science and data science) were applied to the analysis of a literary text. The lyrics of a popular song were taken as an object of study. Then, we performed a quantitative analysis of reader-responses to this song (based on comments to a video). Combining this data with literary interpretation and cognitive models, hypotheses about reader-responses were formulated and then tested against a survey.

The main result measured what topics the readers considered the song talked about (the participants had to rate the pertinence of a series of topics). At the same time, the participants had to make guesses about the interpretations of others (i.e.: "How do you think that most other people will interpret this song?" and they also had to answer by rating topics by pertinence). Some people are more or less skilled in foreseeing the subjective responses of others, especially for something as ambiguous as the lyrics of a song. The remarkable thing is that our model performed 36% better than the human participants —who had used their naked intuitions– in predicting others' interpretations. On the basis of this difference, we can claim that our cognitive-digital model has effectively worked as a powerful tool for upgrading our mindreading skills: just as glasses allow us to better *see* the world around us than our naked eyes, our model has allowed to us to better *see* the world inside (others' minds) than our natural mindreading sense. And this demonstration is capital for the study of culture, since culture is precisely made of our collective minds. Finding new means to increase our vision and measuring capacity with regards to such an abstract and complex sphere as culture is certainly a positive achievement.

This cognitive-digital framework is already being applied for studying culture in many specific disciplines. Fields such as artificial intelligence, behavioral economics, and psychometrics make already productive use of cognitive science and data science in tandem for studying our mental realities in objective manners. What I attempted to show in this book is the interest of applying this framework for studying cultural phenomena of humanistic interest, such as literature.

The interpretation that our case study provided of the song in question, in the end, was based on real data about real people. It was not aimed at creating an elaborate personal speculation on the meaning of the song, but at giving us a better insight about others, about our societies, and about ourselves. In this sense, this research attempts to bring the interest of literary studies to the inquiry of what literature means for our actual fellows. I only saw this point clearly after making a quantitative analysis of this very book (which you can appreciate in the word-cloud of the cover) and realizing that the most frequent word I have used is precisely: “people.” Understanding better ourselves and others has always been the aim of this book, from the very beginning.

As we have stressed in the first part of this book, our minds count with intuitive systems that are evolutionary prior to reason. And these intuitive systems are not directly sensible to abstract logical arguments. We described the nature of these two mental systems by using Jonathan Haidt’s metaphors: the intuitive elephant that rules, and the rational rider that serves it (2012). We have seen, as an example of this, that people make moral judgments immediately and effortlessly; only later people make up rational arguments to justify their intuitive judgments; and the proof that the intuitive elephant is in charge is that refuting people’s arguments (persuading their rational riders) does not make them change their intuitive judgments. Our elephants do not understand logics; what they do understand is metaphors, sensible embodied metaphors. And the elaboration of language in this way –designed to appeal to our intuitive elephants– is what we call *literature*.

An evolutionary psychologist can explain how a cooperative instinct could have evolved in our species, what would its adaptive rationale be. But if he wants to teach his daughter to be cooperative, this explanation would be useless. Instead, he would tell her a fable, he will give her an illustration of that abstract principle embodied in a life-like situation, with life-like characters, to which her intuitive systems can really respond like in real life. We can explain the concept of *good* in terms of analytical moral philosophy; but our intuitive systems experience this concept by embodying it in metaphorical figures such as our *heroes*. You can explain people how funny you are by displaying objective evidence (e.g. a list of all the prices you have won in comedy contests), but you will not persuade anybody of how funny you are until you don’t make them laugh by actually telling a joke. This is also why we use idioms to share abstract ideas and why we can be more moved by the fictitious tale of a suffering character (in a film) than by the factual report of a massive but faceless social catastrophe (in real life).

This is the capital importance of literature and the arts: they allow us to organize and make sense of our experience of the world in ways that fit our intuitive systems, in ways that can be grasped in a sensible manner. If science allows us to *explain* the world (and is valued in terms of its predictive power), literature allows us to *understand* it in our flesh (and its valued in terms of its emotional power).

Wilhelm Dilthey used these terms to refer precisely to these two forms of apprehending the world: explaining (*erklären*) –which he identified with the endeavor of natural sciences (*Naturwissenschaften*)– and understanding (*verstehen*) –which he identified with the historical hermeneutic endeavor of the humanities (*Geisteswissenschaften*) (1914). These perspectives have traditionally been applied to distinct objects: one would try to *explain* things like the stars (e.g. astronomy) but to *understand* things like stories (e.g. literary criticism). However, as I have tried to illustrate in this book, these modes of apprehending the world are not opposing but complementary: every object (natural or cultural) *is* a fact of nature and *means* something to us, at the same time. It is possible, therefore, for us to find beauty in the stars; and it is equally possible to explain the beauty we feel in a story. There can be both a *cultural apprehension of nature* as well as a *scientific explanation of culture*.

The literary record of the humanity –as well as the register of our reader-responses– constitutes one of the most direct windows into the depths of human nature: e.g. each love story contains an intuition-fit description of our mating instincts, as each war story describes an aspect of our tribal instincts, and each joke displays an aspect of our instinctive systems for status ranking. In this fundamental sense, the possibility of a science of literature –enabled by a cognitive-digital framework– is not only great news for literature, but it is also great news for science.

ACKNOWLEDGEMENTS

This book would not have been possible was it not for the aid and funding of the Friedrich Schlegel Graduate School for Literary Studies (FSGS), at the Freie Universität Berlin. I have found here a place both welcoming and challenging that gave me the opportunity to grow intellectually and personally in the exchange with so many interesting people. I want to thank its director, Prof. Jutta Müller-Tamm, the rest of the board, the administrative staff, and all my colleagues and friends for this great experience.

I am especially grateful to my academic supervisors: Prof. Dr. Joachim Küpper (FU Berlin) and Prof. Dr. Dr. Friedemann Pulvermüller (FU Berlin & Humboldt University) for their openness, constructive criticism, and wise guidance during this whole research.

Prof. Küpper honored me by accepting to supervise my research project when it was still a mishmash of ontological and epistemological concerns about the humanities coupled with an obsession for Jorge Luis Borges's literature. His attentive and insightful criticism of my writings helped me to progressively define my position about these topics and to deploy them as a better organized research. He guided the path I have pursued with a great intellectual generosity –often even beyond his agreement. I must thank his advice especially in fostering me not to lose sight of the importance of a sensible approach to literature, which can often be more revealing that artificial measuring methods. The extent to which Prof. Küpper's wisdom has contributed to this book and to my personal intellectual growth is incalculable.

When Prof. Pulvermüller granted me with the privilege of co-supervising my research, my project had become even stranger: I was then an enthusiastic cognitive-science aficionado who was attempting to find some revealing key about literary hermeneutics in the fuzziness of mental categories. I must thank Prof. Pulvermüller for helping me better understand the place of some of these original interests and ideas within a wider and deeper view of the human mind, which was crucial for the elaboration of this book. It was in his lectures on neurolinguistics (and the ones of his colleagues' at the Berlin School of Mind & Brain) that I began to discover the biology of the human brain, the relationship between the body and the mind, and the legacy of Ludwig Wittgenstein. Moreover, the notion of intra-cranial embodied cognition that has been used through this book is directly based on his own pioneering work of research.

The dialogue between science and the humanities present in this book is certainly in debt to the dialogues that I have had the fortune to partake with both of my supervisors. By attentively observing the interplay between their respective fields of expertise (literature and neurolinguistics) I learnt to value the way in which each of these perspectives enlightens the other. I could not have hoped for a better, more complementary, and more enriching advisor team to guide me through this research. *Vielen Dank!*

Part of this research was developed during a stay at the Stanford University in 2017. It was very exciting to discover an environment so striving, forward-thinking, and sunny as that one. I would like to thank many people that I met in this stay. First, Prof. Dr. Hans-Ulrich Gumbrecht (DLCL, and FSGS's international adviser), who so cordially hosted me, guided me through my stay, invited me to participate in the exciting debate sessions of the DLCL, and counseled me regarding my research. Secondly, Prof. Dr. Mark Algee-Hewitt (Literary Lab, CESTA), who invited me to be part of the research team of digital literary studies and to participate in the debates on their projects during my stay. And also Dr. Michal Kosinski (Graduate School of Business), with whom I had the privilege of discussing my research at length, and who gave me methodological advice that was crucial for the design of the case study.

But the person I want to thank the most from Stanford is Jérôme Dockès (INRIA). We shared a home in Downtown Palo Alto, along with other amazing people that made the experience unforgettable (Sasankh, Shini, Samid, Sara, Mon, Sushi, Robyn, and Tuya: What a time, guys!). Jérôme was one of the people that contributed the most to this book: from teaching me the rudiments of coding and statistics to assisting me with the technical processing of the data of the case study. A significant part of the inspiration of the second part of this book comes from all the evenings we spent in the backyard of Webster Inn or cafeterias of the campus imagining the craziest reader-response experiments, planning the design of mindreading-software models, making quantitative analyses of scripts of *The Simpsons*, or thinking of ways of predicting the offensiveness-potential of Twitter posts. It was certainly worth it: *Merci de tout cœur, mec!*

I also want to express my gratitude to other professors who made it possible for me to undertake this PhD research. First of all, I want to thank Prof. Dr. Guy Ducrey (University Strasbourg), my former MA supervisor, for his unwearying support, encouragement, and caring counsel, for his insightful reading and constructive criticism of so many of my manuscripts, for some of the best literary lectures I have ever attended, and for his friendship, which honors me. I am also very thankful to Prof. Dr. Giorgos Freris (Aristotle University of Thessaloniki), who also supported my original application, for his trust and generosity. And also two of my bachelor professors: Prof. Dr. Miguel Veda (University of Buenos Aires), for teaching me historical criticism and mentoring me in my incursion in the academic world; and Prof. Augusto Trombetta (University of Buenos Aires), who first introduced me to the generative grammar, and in whom I encountered the first example of a thinker capable of integrating scientific and humanistic knowledge with such a dexterity and clarity of mind that I still find it inspiring –as this book attempted to testify.

I also want to thank several friends and colleagues who contributed to this research in many different ways. Prof. Dr. Jethro Masís (University of San José), for introducing me to the philosophy of mind and the philosophy of information –besides hosting me for a whole month while I learned how to write proper applications in the *Märchen*-like city of Würzburg with his lovely family: Dany and Andreas. Nicolás Trujillo Osorio (University Diego Portales), for sharing with me his Kant, the enthusiasm for the categories, as well as his Chilean-style scrambled eggs, and the Patagonian feeling; also, for introducing me to Dairon. Prof. Dr. Dairon Alfonso Rodríguez (Industrial University of Santander and Max Planck Institute), for opening my eyes to evolutionary and comparative psychology, for introducing me to the monkeys of Leipzig, and for being one of the friends with whom I enjoy discussing about politics (and infinite other subjects) the most. Jacobo Decamps Mora (Oxford University), for enriching this book by analyzing and criticizing in detail many of its chapters, and for so many (and ongoing) theoretical conversations. Dr. Natalie Phillips (Digital Humanities & Literary Cognition Lab, Michigan State University), for her detailed review of my case study, and for her generous research advice. Paola Ravasio (University of Würzburg), for so much literature, cinema, pasta, and for sharing the Latin-American nostalgia with salsas and tangos. Flor Piluso, for her affectionate and intellectual friendship of so many years, for so much life shared and her unconditional support and encouragement. And to all the guys of *Luthor* (Gustavo, Guada, Eze, Mariano y Lu), for the adventure of literary theory that we shared in El Maestro 5.

I am also deeply thankful to many dear people that has helped me by attentively reading, correcting, editing, and also discussing different parts of this text at different stages of the research: Ailsa May, Mercedes Azar, Natalia García, Michelle Bangura, and Michael Unbehauen. All remaining mistakes are on me!

For the assistance and advice that I have received regarding many aspects of data science, programming, and statistics, I especially want to thank Juan Manuel Benítez Valencia, William Esp, José Torregrosa, Rafael Tascón (Whatxx), Jeff Dean (Google Brain), Maximiliano Bua, Renato Garita, Javier de la Rosa (Stanford University), and Ana María Osiniri –my own mother, who is one of the first (and also one of the smartest) programmers that were formed in Argentina, as well as one of the wisest advisors I know.

Many of the friends that have accompanied and brighten up the days in this period of my life in Berlin deserved to be mentioned, as well: Nicolás Gómez, Diego Sapienza, María Burghi, Dan Avram, Paula García, Ema Avram, Nina Redmann, Natalia Díaz, Belliisi Seenemaa, Giorgi Khomizurashvili, Greg Palmer, Bobby, Miguel Gutiérrez, Mauricio Cardozo, and Emanuel Sotelo. And also many others that have supported me from the distance: The PDG gang (Carlos Díaz Azar, Rodrigo Mallea, Conrado Gumma, and Pablo Azar), the Frezza family (Francesco, Emanuele, Mariapia, and Giampiero), the Constantin family (Ioana, Nicolae, and Michaela), Federico Piedras, Luciano Barreras, Gonzalo Olaguer, Laureano Córdoba, Martín Balboa, and so many others.

I also want to express my debt and gratitude to the friends in my library, a series of renowned scholars in whose work, style, and ideas I have found an inspiration for this book and a feeling of intellectual belonging: Noam Chomsky, Steven Pinker, George Lakoff, John Searle, Daniel Dennett, Matthew Jockers, Jonathan Gottschall, Dan Ariely, Leonard Mladinow, Jonathan Haidt and, always, Jorge Luis Borges. We all stand on the shoulders of giants; these are some of the biggest ones I've known, and whose tradition I have tried to honor.

I want to thank my family in Argentina, for its unconditional love and motivation: Ana María Osiniri, Eduardo Azar, Pablo Azar, Alejandra Núñez, and my dear nephew and godson Agustín Azar, as well as The Osiniris, and Susana & Co. I have missed you so much from this opposite shoreline, but you still held my back, and I am very grateful for having you always pushing me forward. I could not have done this research and this book without your help and example. *¡Muchas gracias a todos!*

Finally, I also want to thank my girlfriend and partner in life: Theodora Constantin. She has been by my side during the whole process of this doctorate, giving me more love, attention, and inspiration than I could have ever asked for, as well as constant intellectual interest and criticism. There is hardly an idea in this book that has not been enriched by my conversations with her. Moreover, the aesthetic of the visuals of this book are her merit – without her designing skills, this book would have probably been much more black and white, both literally and metaphorically. *Mulțumesc!*

To all the people mentioned here, I give my deepest and sincerest gratitude. I hope this work is up to the levels of support and affection I have received while writing it. This book is dedicated to all of you, guys.

BIBLIOGRAPHY

SCIENTIFIC AND CRITICAL LITREATURE

Abbot, H.P.E, 2001. "Imagination and the adapted mind: A special double issue." *Substance*, 30

Allen, J.G.; Fonagy, P., 2006. *The Handbook of Mentalization-Based Treatment*. Wiley

Alvarez, R.M. (ed.), 2016. *Computational Social Science: Discovery and Prediction (Analytical Methods for Social Research)*. Cambridge University Press

American Psychological Association, *APA Dictionary of Psychology*: https://dictionary.apa.org/instinct

Archer, J. & Jockers, M., 2016. *The Bestseller Code. What if a machine could discover the next blockbuster novel?* London: Penguin

Argamon, Sh.; Burns, K.; Dubnov, Sh. (eds.), 2010. "The structure of style: Algorithmic approaches to understanding manner and meaning." Springer Science & Business Media.

Ariely, D., 2008. *Predictibly Irrational: The hidden forces that shape our decisions*. Harper Collins

Ariely, D.; Kreisler, J., 2017. *Dollars and Sense.* Harper Collins Publishers

Augustinos, Walker, Donaghue, 2006. *Social Cognition: an Integrated Introduction*. London: Sage Publications

Aziz-Zadeh, L.; Damasio, A., 2008. "Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology*, *102*, 35-39

Azúa, M., 2011. "Mujer yagana busca preservar lengua a través de diccionario." *La Estrella de Arica*, XXVII, no. 9,999

Bakhtin, M., 1940. *Rabelais and his world* (Original title: *Tvorčestvo Fransua Rable i narodnaja kul'tura srednevekov'ja i Renessansa*). Version: Trans. by Hélène Iswolsky, Indiana University Press, 1984

Barkow, J.H.; Cosmides, L.; Tooby, J., 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press

Baron-Cohen, S.; Leslie, A.M.; Frith, U., 1985. "Does the autistic child have a theory of mind?," *Cognition. Elsevier, 21*, 37-46

Baron-Cohen, S., 1991. "Precursors to a theory of mind: Understanding attention in others," in; Whiten Andrew, *Natural theories of mind: evolution, development, and simulation of everyday mindreading*, Oxford, Uk, Cambridge, Massachusetts, USA: B. Clackwell, pp. 233-251

Bartlett, F.C., 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: University Press

Batson, D.C., 2011. *Altruism in Humans*. Oxford University Press

Beaver, L., 2016. *The Chatbox Explainer*. BI Intelligence

Bebout, J.; Belke, E., 2017. "Language play facilitates language learning: Optimizing the input for gender-like category induction,": https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5318489/

Bergen, B.J., 2012. *Louder Than Words: the new science of how the mind makes meaning*. NY: Basic Books

Berger, P.L.; Luckmann, Th., 1966. *The Social Construction of Reality*. Chicago: Anchor Books

Bergson, H., 1900. *Laughter. An essay on the meaning of the comic* (Original title: *Le Rire. Essai sur la signification du comique*). Version: Trans. by Cloudlesley Brereton, Arc Manor, 2008

Bermúdez, J.L., 2010. *Cognitive science: An introduction to the science of the mind*. Cambridge University Press

Berns, G.S.; Blane, K.; Prietula, M.J., and Pye, B.E., 2013. "Short- and Long-Term Effects of a Novel on Connectivity in the Brain." Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3868356/

BI Intelligence, 2016. "80% of businesses want chatbots by 2020." *Business Insider*: https://www.businessinsider.de/80-of-businesses-want-chatbots-by-2020-2016-12?r=US&IR=T

Blaisdell, A.P., Sawa, K., Leising, K.J. & Waldmann, M.R. (2006) "Causal reasoning in rats", *Science*, 311(5763): 1020-1022.

Bloch, H., 1991. *Medieval misogyny and the invention of western romantic love*, Chicago: University Press

Borges, J.L., 2004. "La fruición literaria," *El Idioma de los argentinos*. Buenos Aires: Emecé

-, 1995. "El arte narrativo y la magia," *Discusión*. Buenos Aires: Alianza.

Boulenger, V.; Hauk, O.; Pulvermüller, F., 2009. "Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension." *Cerebral Cortex, 19*, 1905-1914

Bourdieu, P., 1974. *Esquisse d'une théorie de la pratique*.

Bower, G. & Forgas, J., 2001. "Affective Influences on the Content of Cognition," *Handbook of Affect and Social Cognition*. Ed. Forgas. NJ: Mahwah, 95-120

Boyd, B., 2017. "The evolution of stories: from mimesis to language, from fact to fiction." *Wires Cognitive Science*: https://onlinelibrary.wiley.com/doi/full/10.1002/wcs.1444

-, 2009 *On the Origin of Stories: Evolution, Cognition, Fiction*. Cambridge, MA: Cambridge University Press

Bozovic, N., 2018. "Code Poetry: Poems written in programming languages:" https://www.popwebdesign.net/popart_blog/en/2018/01/code-poetry-poems-written-in-programming-languages/

Brakus, M.; Bluck, J., 2004. "Nasa Develops System to Computarize Silent, Subvocal Speech." *Nasa News*. Link: https://www.nasa.gov/home/hqnews/2004/mar/HQ_04093_subvocal_speech.html

Brandom, R., 2014. "Why is Facebook beating the FBI at facial recognition?" *The Verge*: https://www.theverge.com/2014/7/7/5878069/why-facebook-is-beating-the-fbi-at-facial-recognition

Bratman, 1992. "Shared cooperative activity." *The Philosophical Review*, Vol. 101, Issue 2, 327-314

Breton, A., 1948. *Le Cadavre Exquis: Son exaltation*. Paris: La Dragonne

Brewer, W.F., 1985. "The story schema: universal and cultural-specific properties." *Literacy, Language, and Learning*. Cambridge University press, 167-94

Broek (van den), P., 1990. "Causal inferences and the comprehension of narrative texts." *Psychology of learning and motivation*, Vol. 25: https://www.sciencedirect.com/science/article/pii/S0079742108602558

Brooks, R., 2013. "Whodunnit? J.K. Rowling's secret life as a wizard crime writer revealed." *The Sunday Times* (July 14th): https://www.thetimes.co.uk/article/whodunnit-jk-rowlings-secret-life-as-wizard-crime-writer-revealed-vfxzwf72jzh

Brookshire, R., 2007. *Introduction to neurogenic communication disorders*. St. Louis, MO: Mosby Elsevier

Bross, F., 2018. "Cognitive associations between vowel length and object size: A new feature contributing to a bouba/kiki effect." *Proceedings of the Conference on Phonetics & Phonology*. Berlin: Humboldt University, 17-20

Brown, D., 1991. *Human Universals*. New York: MacGraw-Hill

Bruner, J. 2002. *Making stories: law, literature, life*. New York: Farrar

-, 1994. *Acts of Meaning.* Cambridge: Harvard University Press

-, 1986. *Actual Minds, Possible Worlds.* Cambridge: Harvard University Press

Brunye, T.T.; Ditman, T.; Mahoney, C.R.; Augustyn, J.S.; Taylor, H.A., 2009. "When you and I share perspectives: Pronouns modulate perspective-taking during narrative comprehension." *Psychological Science, 20*, 27-32.

Burdick, A.; Drucker, J.; Lunefeld, P.; Presner, T.; Schnapp, J., 2016. *Digital Humanities*. MIT Press

Calvin, W.H., 1996. *How Brains Think*. New York: Basic Books

Campbell, J., 1949. *The Hero of the Thousand Faces*. Pantheon Books

Call, J.; Tomasello, M., 1998. "Distinguishing intentional from accidental actions in orangutan (Pongo pygmaeus), chimpanzees (Pan troglodytes), and human children (Homo sapiens)." *Journal of Comparative Pyschology, 112* (2): 192-206

Caplan, D., 2006. "Why is Broca's area involved in syntax?" *Cortex, A Journal Devoted to the Study of the Nervous System and Behavior, 42* (4), pp. 469-471

Carole, Ch., 2012. *Author Identification in the Forensic Setting*. *The Oxford Handbook of Language and Law*. Oxford University Press

Carota, F.; Kriegeskorte, N.; Nili, H.; Pulvermüller, F., 2017. "Representational Similarity Mapping of Distributional Semantics in Left Inferior Frontal, Middle Temporal, and Motor Cortex," *Cerebral Cortex*, Oxford, 27: 294-309

Carpan, F., 2017. "The AI Revolution is Underway!" *PM360*: https://www.pm360online.com/the-ai-revolution-is-underway/

Casasanto, D.; Boroditsky, L., 2008. "Time in the mind: Using space to think about time." *Cognition, 106*, 579-593

Chan, T.T.; Bergen, B., 2005. Writing direction influences spatial cognition, *Proceedings of the 27th Annual Conference of the Cognitive Science Society*

Chapman, A.; Barthe, Ch.; Revol, Ph., 1995. *Rencontre avec les indiens Yaghan Relié*. Paris: Editions de la Martinière

Chaplin, S., 2006. "The placebo response: an important part of treatment." *Prescriber*: 16–22

Chomsky, N.; Berwick, R., 2016. *Why Only Us? Language and Evolution*. MIT Press

Chomsky, N., 1975. *Reflections on Language*. NY: Pantheon

-, 1968. *Language and Mind*.

-, 1965. *Aspects of a Theory of Syntax.* Cambridge: MIT

Collins, S.A., 2000. "Men's choices and women's voices." *Animal Behavior, 60*, 773-80

Concordia, 2016. "Cambridge Analytica – The Power of Big Data and Psychographics:" https://www.youtube.com/watch?v=n8Dd5aVXLCc

Crick, F., 1994. *The Astonishing Hypothesis: The scientific search for the soul*. New York: Simon & Schuster

Daly, I., 2018. "You call that poetry?!," *Poetry Foundation*: https://www.poetryfoundation.org/articles/68913/you-call-that-poetry

Dahl, C.D.; Adachi, I., 2013. "Conceptual metaphorical mapping in chimpanzees." Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3798977/

Daniels, P.T.; Bright, W. (eds.), 1996. *The world's writing systems*. New York: Oxford University Press

Damasio, A., 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam
-, 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt
-, 2010. *Self comes to mind: Constructing the conscious brain*. Pantheon
Darwin, Ch.., 1859. *On the origin of the species by means of natural selection, or the preservation of favoured races in the struggle for life*, London: John Murray
-, 1871. *The descent of man, and selection in relation to sex*. London: John Murray.
Dautenhahn, K., 2017. "The origins of narrative. In search of the transactional format of narratives in humans and other social animals." University of Hertfordshire: https://pdfs.semanticscholar.org/bc05/37be6de4c606c314b255ee184a661f096bcb.pdf
Dawkins, R., 1976. *The Selfish Gene*. Oxford University Press.
-, 2013. "Storytelling of Science" (conference). ShirleyFilms: https://www.youtube.com/watch?v=d5KnGaHFF6U&t=220s
Dehaene, S., 2009. *Reading in the brain*. London: Penguin
Dennett, D.C., 2017. *From Bacteria to Bach. The Evolutionary Mind*. London: Penguin
-, 2010. "Darwin's strange inversion of reasoning." (Lecture). Link: https://www.youtube.com/watch?v=DvOEpmZHA1s
-, 1992. *Consciousness Explained*. Little, Brown & Co.
Deutsch, D., 2011. *The Fabric of Reality.* Viking Adult
Dilthey, W., 1914. *Introduction to the Human Sciences* (Original title: *Einleitung in die Geisteswissenschaften*). Version: Princeton Univeristiy Press, 1991
Dissanayake, E., 2011. "Prelinguistic and preliterate substracts of poetic narrative," University of Seattle:
Dolezel, L., 1997. *Heterocosmica: Fiction and Possible Worlds*
Donald, M. 1991. *Origins of the modern mind: three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press
Dreyfus, 1979. *What Computers Can't Do*. Revised Edition. New York: Harper and Row
Dunbar, R.I.M. et al., 2011. "Social laughter is correlated with an elevated pain threshold." Proceedings of the Royal Society B, Biological Sciences: http://rspb.royalsocietypublishing.org/content/279/1731/1161
Dutton, D., 2010. *The art instinct: beauty, pleasure, and human evolution*. Bloomsbury
Dzieza, J., 2015. "The strange world of computer-generated novels." *The Verge*: https://www.theverge.com/2014/11/25/7276157/nanogenmo-robot-author-novel
Ekman, P., 2007. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*, 2nd Ed. Holt Paperbacks
Ensaff, H.; Homer, M.; Sahota, P.; Braybrook, D.; Coan, S.; McLeod, H., 2015. "Food Choice Architecture: An Intervention in a Secondary School and its Impact on Students' Plant-based Food Choices" *Nutrients*, 4426-4437.
Fanning, Ch., 1999. *The Irish Voice in America 250 years of Irish-American Fiction* (2nd Ed.). University Press of Kentucky
Ferrante, J., 2014. *The Political Vision of the Divine Comedy*. Princeton University Press
Fetterman, A. K.; Wilkowski, B. M.; Robinson, M.D., 2017. "On feeling warm and being warm: Daily perceptions of physical warmth fluctuate with interpersonal warmth." Link: 2017: https://www.researchgate.net/publication/316439150_On_Feeling_Warm_and_Being_Warm_Daily_Perceptions_of_Physical_Warmth_Fluctuate_With_Interpersonal_Warmth)
Fillmore, Ch., 1982. "Frame Semantics". In Linguistic Society of Korea, ed., *Linguistics in the Morning Calm*, pp. 111-38. Seoul: Hanshin

Frith, C.; Frith, U., 2005. *Theory of Mind*. Curr Biol. 15 (17); R644-6
Fiske, S.T. & Malone, Ch., 2013. *The Human Brand. How we relate to people, products, and companies*. San Francisco: Jossey Bass
Fiske, S.T.; Taylor, S.E., 2014. *Social Cognition: from brains to culture*. London: Sage
Fiske, S.T; Kervin, N.; and Malone, Ch., 2011. "Brands as intentional agents framework: How perceived intentions and ability can map brand perception."
Fiske, S.T.; Cuddy, A.J.; and Glick, P., 2007. "The BIAS Map: Behaviors From Intergroup Affect and Stereotypes," *Journal of Personality and Social Psychology*, Vol. 92, No. 4, 631-348
Fiske, S.T.; Cuddy, A.J.; Glick, P.; and Xu, J., 2002. "A model of (often mixed) stereotype content: Competence and Warmth respectively follow from perceived status and competition," *Journal of Personality and Social Psychology*, Vol. 82, No. 6, 878-902
Floridi, L. & Searle, J., 2016. "A Discussion of Artificial Intelligence." Link: https://www.youtube.com/watch?v=b6o_7HeowY8
Fodor, J.A., 1975. *The Language of Thought*. NY: Cromwell
-, 1978. *RePresentations. Philosophical Essays on the Foundations of Cognitive Science*. MIT Press
Freund, E., 1987. *The Return of the Reader: Reader-Response Criticism.* Methuen.
Friederici, 2017. *Language in Our Brain: The Origins of a Uniquely Human Capacity*. MIT Press
Gandomi, A.; Haider, M., 2015. "Beyond the hype: Big data concepts, methods, and analytics." *International Journal of Information Management*, vol. 35, no. 2, 137-144
Gavins, J.; Steen, G., 2003. *Cognitive poetics in practice*. London: Routledge
Gazzaniga, M. 2000. *The mind's past*. Berkeley: University of California Press
Geary, D.C., 2010. *Male, female: The evolution of human sex differences*, Sec. Ed., Washington: American Psychological Association
Gersh-Nesic, B., 2016. "Art History Definition: Action Painting." ThoughtCo: thoughtco.com/art-history-definition-action-painting-183188.
Gilovich, Th., 1991. *How we know what isn't so*. New York: Mac Millan
Goswami, U.; Bryant, P., 2016. *Phonological Skills and Learning to Read*. Psychology Press & Routledge Classic Editions
Gottschall, J., 2013. *The Storyteller Animal. How Stories Make Us Human*. New York: Mariner
Gottschall, J.; Norlund, M., 2006. "Romantic love: a literary universal?" *Project Muse*: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.731.5984&rep=rep1&type=pdf
Gottschall, J.; Wilson, D.S., eds., 2005. *The literary animal: Evolution and Literature* 25, no. 2: 314-34
Graefe, A., 2016. *Guide to Automated Journalism*. NYC: Columbia Journalism Review
Grassegger, H. & Krogerus, M. 2017. "The data that turned the world upside down." Link: https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win
Greimas, A.J., 1966. *Structural Semantics: An Attempt at a Method*. University of Nebraska Press (ed. 1984)
Grice, H.P., 1975. *Logic and Conversation*. University College London
Gould, S.J.; Vrba, E.S., 1982. "Exaptation – A missing term in the science of form." *Paleobiology*, 8 (1): 4-15
Gumbrecht, H.U., 2004. *Production of Presence: What Meaning Cannot Convey*. Stanford University Press
Haidt, J., 2006. *The Happiness Hypothesis*. New York: Basic
-, 2012. *The Rightgeous Mind*. London: Penguin

Harlow, H.F., 1958. “The nature of love.” *American Psychologist*, 13 (12), 673-685. Harlow, H.F, & Zimmermann, R.R., 1959, “Affectional responses in the infant monkey.” Science, *130*, 421-432

Harnad, S., 2003. *Cognition is categorization*. UQaM Summer Institute in Cognitive Sciences on Categorization

Hauk, O.; Davis, M. H.; Kherif, F.; Pulvermüller, F., 2008. “Imagery or meaning? Evidence for a semantic origin of category-specific brain activity in metabolic imaging,” *European Journal of Neuroscience*, Vol. 22, pp. 1856-1866.

Hayes, K.J., 1997. *Folklore and Book Culture*. Wipf and Stock

Hebb, D., 1949. *The Organization of Behavior*. New York: Wiley & Sons

Heider, F. & Simmel, M., 1994. “An Experimental Study of Apparent Behavior.” *The American Journal of Psychology*, Vol. 57, No. 2, 243-259. Moving film: https://www.youtube.com/watch?v=sx7lBzHH7c8

Herman, D., ed. 2003. *Narrative theory and the cognitive sciences.* Standford, CA: Center for the Study of Language and Information

Hitt, J., 2004. “Say No More.” *The New York Times Magazine*: https://www.nytimes.com/2004/02/29/magazine/say-no-more.html?pagewanted=all&src=pm

Hogan, P.C., 2003. *Cognitive Science, Literature, and the Arts*. *A Guide for Humanists*. NYC: Routledge

-, 2010. “Literary Universals,” in: Lisa Zunshine (ed.), *Introduction to Cognitive Cultural Studies*. Baltimore: John Hopkins University Press, p. 37-60

Hohwy, J., 2013. *The Predictive Mind*. Oxford: University Press.

Honeck, R.P., 1997. *A Proverb in Mind: The Cognitive Science of Proverbial Wit and Wisdom*. Mahwah, NJ: Lawrence Erlbaum Associates

Hukkinen, J.I., 2012. “Fit in the body: Matching embodied cognition with social-ecological systems.” *Ecology and Society*, Vol. 17, No. 4. Link: https://www.ecologyandsociety.org/vol17/iss4/art30/#ms_abstract

Hutchison, P., 2010. “ Tommy Cooper dominates list of best jokes.” The Telegraph: https://www.telegraph.co.uk/culture/comedy/8078185/Tommy-Cooper-dominates-list-of-best-jokes.html

Huyssen, A., 1986. *After the Great Divide: Modernism, Mass Culture, Postmodernism*

Infobae, 2018. “La abuela Cristina: la última habitante de la etnia del fin del mundo.” https://www.infobae.com/america/mundo/2017/05/12/la-abuela-cristina-la-ultima-habitante-de-la-etnia-del-fin-del-mundo/

Janovik, M.L., 2013. “Communication and shared information.” Online

Jockers, M.L., 2013. *Macroanalysis. Digital Methods & Literary History*. Illinois.

Jockers, M.L.; Witten, D.M., 2010. “A Comparative Study of Machine Learning Methods for Authorship Attribution.” *Literary and Linguistic Computing*, 25.2, 215-224.

Johnson, E.; Goldstein, D.G., 2003. “Do Defaults Save Lives?”: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1324774

Johnson, K., 2017. “Facebook Messenger hits 100,000 bots.” *VB*: https://venturebeat.com/2017/04/18/facebook-messenger-hits-100000-bots/

Jaén, I.; Simon, J.J. (eds.), 2013. *Cognitive Literary Studies*. Austin: University of Texas

James, W., 1890. *Principles of Psychology*. Henry Holt & Co

Jasper, H.; Penfield, W., 1954. *Epilepsy and the Functional Anatomy of the Human Brain*. 2nd edition. Little, Brown and Co.

Josipovici, G., 1976. *The Modern English Novel*. Barnes & Noble Books

Jung, C., 1969. *Archetypes and the collective unconscious*, Collected Works, Vol. 9, Part 1. Princeton University Press

Kant, I., 1781. *Critique of Pure reason* (Original title: *Kritik der reiner Vernunft*). Version: Trans. by Marcus Weigelt and Max Muller. Penguin, 2008

Kelly, J.L., 2011. *Renaissance Futures: Chance, Prediction, and Play in Northern European Visual Culture*. University of California, Berkley.

Kennison, Sh., 2013. *Introduction to Language Development*. Los Angeles: Sage

Koestler, A., 1964. *The Art of Creation*. New York: Dell

Köhler, W., 1929. *Gestalt Psychology*. New York: Liveright

Köhler, W., 1947. *Gestalt Psychology* (2nd Ed.). New York: Liveright

Kosinski, M., 2012. "Private traits and attributes are predictable from digital records of human behavior." Link: http://www.pnas.org/content/110/15/5802.full

-, 2017. "Dr. Michal Kosinski on Facebook, Big Data, and Psychographic Profiling." *Point*: https://www.youtube.com/watch?v=XKNFjVM2SfY&t=720s

Kringelbach, M.L., 2005. "The Human Orbitofrontal Cortex: Linking Reward to Hedonic Experience." *Nature Reviews: Neuroscience 6* (September): 691-702

Krüger, A., 2010. *Wie erzeugt E.T.A. Hoffmann in seinem Nachtstück "Der Sandmann" Spannung?* Gottfried Wilhelm Leibniz Universität Hannover

Lakoff, G., 2014. "Mapping the brain's metaphor circuitry: metaphorical thought in everyday reason." *Frontiers in Human Neuroscience*. Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4267278/

Lakoff, G. & Johnson, M., 1980. *Metaphors We Live By*. Chicago: University Press

-, 1999. *Philosophy In The Flesh: the Embodied Mind and its Challenge to Western Thought*. Basic Books.

Lakoff, G., 1990. *Women, Fire, and Dangerous Things. What Categories Reveal about The Mind*. Chicago: University Press.

LeDoux, J., 1996. *The Emotional Brain: The mysterious underpinnings of emotional life.* New York: Touchstone

Lehne et al., 2015. "Reading a suspenseful story activates brain areas related to social cognition and predictive inference:" https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4422438/

Lemaitre, A.L.; Luyat, M.; Lafargue, G., 2016. "Inidividuals with pronounced schizotypal traits are particularly successful in tickling themselves." *Consciousness and Cognition*, 41, 64-71

Little, A.C.; Jones, B.C., and DeBruine L.M., 2011. "Facial attractiveness: evolutionary based research." Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3130383/

Liyakasa, K., 2018. "The Associated Press uses AI to boost content and video volume." AdExchanger: https://adexchanger.com/publishers/associated-press-uses-ai-boost-content-video-volume/

Lott, R., 2013. "History of Sampling." Joyful Noise Recordings: http://www.joyfulnoiserecordings.com/journal/akai-mpc-2000the-history-of-sampling/

Maas, A.; Russo, A., 2003. "Directional bias in the mental representation of spatial events: Nature or culture?," *Psychological Science* 14, 296-301

Mahy, C.E.V.; Moses, L.J.; Pfeifer, J.H., 2014. "How and where: Theory-of-mind in the brain." *Developmental Cognitive Neuroscience*, Vol. 9: 68-81

Manovich, L., 2001. *The Language of New Media*. MIT Press

Matamalas, R.L.; Ramos May, M.S., 2009. *Marketing Strategy of Supermarkets*.

Marcault, M., 2015. "The History of Text Generation:" http://mathesonmarcault.com/index.php/2015/12/15/

Martinez, M.; Scheffel, M., 2009. *Einführung in die Erzähltheorie*. C.H. Beck
Maurer, D.; Pathman, T.; Mundloch, C.J., 2006. “The shape of boubas: Sound-shape correspondences in toddlers and adults.” *Developmental Science, 9*, no. 3, 316-322
Mayer, R., 2014. “How a California earthquake becomes the news: An extremely precise timeline.” *The Atlantic*: https://www.theatlantic.com/technology/archive/2014/03/how-a-california-earthquake-becomes-the-news-an-extremely-precise-timeline/284506/
McConachie, B.; Hart, F.E., eds., 2006. *Performance and cognition: Theater studies and the cognitive turn*. Londn: Routledge
Mead, G.H., 1934. *Mind, Self and Society*. University Chicago Press.
Merchant, B., 2015. “The Poem That Passed the Turing Test.” *Motherboard*: https://motherboard.vice.com/en_us/article/vvbxxd/the-poem-that-passed-the-turing-test
Merritt, D.J.; Cassanto, D.; Brannon, E.M., 2010. “Do monkeys think in metaphors? Representations of space and time in monkeys and humans.” Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2952654/
Michel, J.B.; Liberman Aiden, E., 2010. “Quantitative Analysis of Culture Using Millions of Digitized Books.” *Science 331*: 176-82
Miller, G., 2001. *The Mating Mind*. New York: Anchor
Mitchel, R.W., 2007. “Pretense in animals: The continuing relevance of children’s pretense.” *Play and development: Evolutionary, sociocultural, and functional perspectives*, 51-57
Mladinow, L., 2012. *Subliminal: How your unconscious mind rules your behavior*. New York: Vintage Books
Molly, R., 2016. “China and the US dominate the world’s book production.” International Publishers Association: https://www.theatlas.com/charts/HkxvNCt_-
Montal, T.; Reich, Z., 2016. “I, Robot. You, Journalist. Who is the author?” *Digital Jounalism, 0*, 1-21
Mooiji, J.J.A., 1993. *Fictional Realities: The Uses of Literary Imagination*. J. Benjamins.
Moretti, F., 2013. *Distant Reading*. Verso.
Nagel, Th., 1974. “What is it like to be a bat?” *The Philosophical Review*
Nass, C. et al., 1994. “Computers are social actors.” *Proceedings of the ACM CHI 94 Human factors in computing systems conference*. MA: Association for Computer Machinery Press, 72-77
-, 1997. “Are computers gender neutral?” *Journal of Applied Social Psychology, 27*, no. 10, 864-76
Nass, C.; Lee, K.M., 2000. “Does computer-generated speech manifest personality? An experimental test of similarity-attraction.” *CHI Letters 2*, no.1, 329-326
Nathan Pepitone, R., et al., 2008. “Women’s voice attractiveness varies across the menstrual cycle.” *Evolution of Human Behavior 29*, no 4, 268-74
Nichols, S.; Stich, S., 2003. *Mindreading*. Oxford University Press
Oatley, K., 1992. *Best Laid Schemes: The Psychology of Emotions*. Cambridge: Cambridge University Press
-, 2002. “Emotions and the Story Worlds of Fiction,” *Narrative Impact: Social and Cognitive Foundations*. Ed. Melane Green, Jeffrey Strange, and Timothy Brock. Mahwah, NJ: Erlbaum, 39-69
Oatley, K.; Keltner, D.; Jenkins, J.M., 2006. *Understanding emotions*. Wiley-Blackwell
O’Connor, T.; Wong, H.Y., 2012. Edward N. Zalta (ed.), “Emergent Properties”. *The Stanford Encyclopedia of Philosophy.* Stanford University

Ouaknine, A., 2018. “Review of deep learning algorithms for object detection.” *Medium*: https://medium.com/comet-app/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852
Pavel, 1986. *Fictional Worlds*. Cambridge: Harvard University Press
Petrie, M., 1994. “Improved growth and survival of offspring of peacocks with more elaborate trains.” *International Journal of Science*: https://www.nature.com/articles/371598a0
Piaget, J., 1923. “Langage et pensée chez l’enfant”. Neuchâtel: Delchaux et Niestlé, p. 43f.
Pinker, S., 2007. *The Blank Slate*. London: Penguin
-, 2007b. “Toward a consilient study of literature (review of Gottschall & Sloan Wilson, *The Literary Animal: Evolution and the Nature of narrative*). *Philosophy and Literature, 31*: 161-177
-, 2002. *The Stuff of Thought. Language as a windows into human nature.* London: Penguin
-, 1997. *How the Mind Works*. London: Norton
-, 1994. *The Language Instinct*. London: Penguin
Pickett, R., 2004. *Sideways*. St. Martin’s Griffin
Pottala, M., 2018. *Artificial Intelligence*. Centria University of Applied Sciences
Povinelli, D.J., 2000. *Folk Physics for Apes*. Oxford University Press
Previc, F.H., 2011. *The Dopaminergic Mind in Human Evolution and History*. Cambridge University Press
Prinz, J.J., 2004. *The Emotional Construction of Morals*. Oxford University Press
-, 2014. Emotions and consciousness. Link: https://www.youtube.com/watch?v=SpxgEQA7pu0
Provine, R.R., 1996. “Laughter.” *American Scientist, 84,* 38-45
Putnam, H., 1961. “Brains and Behavior.” Cal State Long Beach: http://web.csulb.edu/~cwallis/382/readings/483/putnam.brains.behavior.1980.pdf
Puts, N. et al., 2006. “Dominance and evolution of sexual dimorphism in human voice pitch,” *Evolution and Human Behavior 27,* 283-96
Puts, N., 2005. “Mating Context and Menstrual Phase Affect Women’s Preferences for Male Voice Pitch.” *Evolution and Human Behavior, 26*, 388-97
Quartz, 2017. “The story of Replika: The AI app that becomes you.” *Machines with Brains* (S01, E01): https://www.youtube.com/watch?v=yQGqMVuAk04
Ramachandran, V.S.; Hubbard, E.M., 2001. “Synaesthesia: A window into perception, thought and language.” *Journal of Consciousness Studies, 8*, no. 12, 3-34
Raposo, A.; Moss, H.E.; Stamatakis, E.A.; Tyler, L.K., 2009. “Modulation of motor and premotor cortices by actions, action words and action sentences.” *Neuropsychologia, 47,* 388-396
Random International (experimental artist collective), 2017. “Evolution of the mind, consciousness, and AI.” https://youtu.be/o86W0DgrmRc?t=1910
Richardson, D.C.; Matlock, T., 2007, “The integration of figurative language and static depictions: an eye movement study of fictive motion.” *Cognition*, 102, 129-138
Richardson, A.; Spolsky, E., eds., 2004. *The work of fiction: Cognition, culture, and complexity*. Aldershot, UL: Ashgate
Reyes, A., 1929. “Jitanjáfora” (dictionary article), *Diccionario de la Real Academia Española* (23th Ed.). Madrid: Espasa. http://dle.rae.es/?id=MUZjPaR
Robins; Ayede, eds., 2010. *The Cambridge Handbook of Situated Cognition*. NY: Cambridge Press
Rockmore, D., 2006. The Style of Numbers Behind a Number of Styles.” The Chronicle Review: https://web.archive.org/web/20060630113842/http://chronicle.com/temp/reprint.php?id=4fvlt82gn640d1rp48srbpjsvlzhmyrs

Rodríguez, D., 2017. *La cognición como fenómeno biológico*. Ciudad de México: Universidad Autónoma Metropolitana

Romm, J., 2016. “Why *Time*’s Trump Cover Is a Subversive Work of Political Art. Link: http://www.haaretz.com/us-news/1.758226

Rosch, E.; Lloyd, B.L.; 1978. *Cognition as Categorization*. Hillsdale NJ: Lawrence Erlbaum Associates

Russ, S.W.; Wallace, C.E., 2004. “Pretend play and creative process,” *Journal of Play*: http://www.journalofplay.org/sites/www.journalofplay.org/files/pdf-articles/6-1-article-pretend-play.pdf

Ryan, M.L., 1992. *Possible Worlds, Artificial Intelligence, and Narrative Theory*

Sayers Peden, M., 1983. *The Latin American Short Story: A Critical History*. Twayne Publishers

Sato, M.; Schafer, A.; Bergen, B., 2012. “Mental representations of object shape change incrementally during sentence processing. *Language and Cognition*

Schacter, D.L., 1996. *Searching for Memory: The Brain, the Mind, and the Past*. New York: Basic Books

Schaeffer, J.M., 2011. *Petite Écologie des Études Littéraires*. Paris: Éd. Thierry Marchaisse

Schäfer, J., 2006. *Literary Machines Made in Germany: German Proto-Cybertexts from the Baroque Era to the Present*: http://cybertext.hum.jyu.fi/articles/77.pdf

Schleicher, A., 1853. *Die ersten Spaltungen des indogermanischen Urvolkes. Allgemeine Zeitung fuer Wissenschaft und Literatur* (August 1853)

-, 1863. *Darwinism Tested by the Science of Language* (Original title: *Die Darwinsche Theorie und die Sprachwissenschaft*). Version: Trans. by Alexander Bikkers, London: J.C. Hotten, 1869

Schmitz, T. et al., 2009. “Opposing Influences of Affective State Valence on Visual Cortical Encoding.” Journal of Neuroscience. 29: 7199-7207.

Schofield, J., 2014. “Computer chatbot ‘Eugene Goostman’ passes the Turing Test.” *ZDNet*: https://www.zdnet.com/article/computer-chatbot-eugene-goostman-passes-the-turing-test/

Schwencke, K., 2014. *Los Angeles Times* (17th March 2014, 6:33 a.m.): http://www.tronc.com/gdpr/latimes.com/

Schuster, M., 2017. “Machine Learning Behind Google Translate.” (Conference). Google Brain: https://www.youtube.com/watch?v=vQLoIYnhPsY

Schuster, M.; Johnson, M.; Thorat, N., 2016. “Zero-Shot Translation with Google’s Multilingual Neural Machine Translation System.” *Google AI Blog*: https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Scruton, R., 2013. “Scientism and the Humanities” (conference). Bloomsbury: https://www.youtube.com/results?search_query=roger+scruton+scientism

Searle, J., 2015. “Consciousness in Artificial Intelligence” (conference). Talks at Google: https://www.youtube.com/watch?v=rHKwIYsPXLg&t=2s

-, 2005. *Mind: A Brief Introduction*. Oxford University Press

-, 1996. *The Construction of Social Reality: A Treatise on Sociology of Knowledge.* London: Penguin

-, 1983. *Intentionality: An Essay in the Philosophy of Mind*. London: Penguin

Shklovkski, V., 1998. “Art as Technique”. In Julie Rivkin & Michael Ryan, ed., *Literary Theory: An Antology*. Malden: Blackwell

Shore, B., 1996. *Culture in Mind.* NY: Oxford University Press

Sidhu, D.M.; Pexman, P.M., 2015. “What’s in a name? Sound Symbolism and Gender in First Names.” *Plos One, 10*, no. 5.

Simanowski, R., 2010. “Automatisches Schreiben.” XCULT: http://www.xcult.org/helmhaus/simanowski.html
Simonite, T., 2014. “Facebook creates software that matches faces almost as well as you do.” *MIT Technology Review*: https://www.technologyreview.com/s/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/
Solomon, K.O.; Barsalou, L.W., 2004. “Perceptual simulation in property verification. *Memory and Cognition*, 32, 244-259
Sostek, A., 2013. “Duquesne professor helps ID Rowling as author of *The Cuckoo’s Calling*.” *Pittsburgh Post-Gazette*: http://old.post-gazette.com/stories/news/education/duquesne-prof-helps-id-rowling-as-author-695629
Spence, Ch., 2011. “Mouth-watering: the influence of environmental and cognitive factors on salivation and gustatory/flavor perception.” *J Texture Stud.*, 42: 157-71
Spreng, N.; Rusu, A.; Clifford, R.M.; Schacter, R., 2014. “Imagine all the people: How the brain creates and uses personality models to predict behavior.” *Cerebral Cortex,* 28, Vol. 24, Issue 8, pp. 1979-1987
Su, R.; Rounds, J., 2009. “Men and things, women and people: A meta-analysis of sex differences and
Suchman, L., 1987. *Plans and Situated Action*. NY: Cambridge Press
Sulakatko, S., 2014. *Consumer Behavior and Decision Making Process*. Jamm University of Applied Sciences
Taylor, Sh., 1989. *Positive Illusions: Creative Self-Deception and the Healthy Mind.* New York: Basic
Temple, E., 2017. “How many books will you read before you die?” LitHub: https://lithub.com/how-many-books-will-you-read-before-you-die/
Thomas, A. K.; Millar, P. R., 2011, “Reducing the Framing Effect in Older and Younger Adults by Encouraging Analytic Processing,” *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 67B (2): 139.
Tomasello, M., 2010. *Origins of Human Communication*. Cambridge: MIT
Turing, A.H., 1950. “Computing Machinery and Intelligence.” *Mind, 49*: 433-460
Turner, M., 1991. *Reading minds: The study of English in the age of cognitive science*. Oxford: Oxford University Press
Tvarsky, A.; Kahneman, D., 1986. *Rational choice and the framing of decisions*. University Chicago Press
Tynianov, Y., 1977. *Formalist Theory*, translated by L.M. O’Toole and Ann Shukman
Varela, F.J.; Thompson, E.; Rosch, E., 1993. *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge: MIT Press
Visualarte, 2013. “Yagan: Los confines del mundo - Cristina Calderón: Última hablante Yagan.” Centro de Producción Audiovisual Chile: https://www.youtube.com/watch?v=q9XEis5mDJg&t=9s
Vitaliev, V., 2009. “After All. E&T Reviews the first ever novel written by a computer.” *E&T*: https://eandt.theiet.org/content/articles/2009/01/after-all/
Walker, R. et al., 2009. Significant Objects and how they got that way. Link: http://significantobjects.com/
Wansink, B., et al., 2001. “Descriptive menus labels effects on sales.”
Westcott, R., 2006. “Making hit music into a science.” *BBC*
Wernicke, K., 1875. “The aphasia symptom-complex: A psychological study on an anatomical basis.” In: Paul Eling, *Reader in the History of Aphasia*. Amsterdam: Benjamins, Pub, 1995, pp. 69-89

Wilde, O., 2003. *Complete Works of Oscar Wilde*. Collins Classics.
Williams, L.E.; Bargh, J.A., 2008. "Experiencing physical warmth influences interpersonal warmth." *Science, 322*, 606-607
Wilson, D.S., 2003. *Darwin's Cathedral*. Chicago: University of Chicago Press
Wilson, E.O., 1975. *Sociobiology: The New Synthesis*. Harvard University Press
Wilson, N.L.; Gibbs, R.W., 2007. "Real and imagined body movement primes metaphor comprehension." *Cognitive Science, 31*, 721-731
Wimmer, H.; Perner, J., 1983. "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception." *Cognition, 13*, 103-128
Wittgenstein, L., 1953. *Philosophical Investigations*, trans. by G.E.M. Anscombe. Oxford University Press.
Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; Klinger, J.; Shah, A.; Johnson, M.; Liu, X.; Kaiser, L.; Gouws, S.; Kato, Y.; Kudo, T.; Kazawa, K.S.; Kurian, G.; Patil, N.; Wang, W.; Young, C.; Smith, J.; Riesa, J.; Rudnick, A.; Vinyals, O.; Corrado, G.; Hughes, M.; Dean, J., 2016. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation." Cornell University Library: https://arxiv.org/abs/1609.08144
Zak, P.J., 2012. *The Moral Molecule: The Source of Love and Prosperity.* Dutton Adult
Zhong, C.B.; Leonardelli, G.J., 2008. "Cold and lonely: Does social exclusion feel literally cold?," *Psychological Science, 19*, 838-842
Zunshine, L., 2010. *Introduction to Cognitive Cultural Studies*. John Hopkins University Press
-, 2006. *Why we read fiction: Theory of Mind and The Novel (Theory and interpretation of narrative*). Ohio State University Press
Zwaan, R.A., 2004. "The immersed experiencer: Toward an embodied theory of language comprehension. In B.H. Ross (ed.), *The psychology of learning and motivation*, *43* 224-245. New York: Academic Press

ARTWORKS (LITERARY WORKS, FILMS, AND OTHERS)

Adelson, E., 1995. "Checker Shadow Illusion" (optical illusion). MIT: http://persci.mit.edu/gallery/checkershadow
Alighieri, D., 1320. *The Divine Comedy* (Original title: *Divina Comedia*). Version: Tans. by C.H. Sisson, Oxford University Press, 1981
Ariosto, 1532. *Orlando Furioso*. Version: Trans. by Guido Waldman, Oxford University Press, 1999
Austen, J., 1811. *Sense and Sensibility*. Thomas Egerton
-, 1813. *Pride and prejudice*. Thomas Egerton
-, 1814. *Mansfield park*. John Murray
-, 1816. *Emma*. John Murray
Baldaci, D., 2013. *The Hit*. Grand Central Publishing
Barsotti, Niccolò de Lucca, 1657. *Cynosura Mariana*. Source: Marcault, 2018 (v. supra)
Ballard, J.G., 1973. *Crash*. Jonathan Cape Ed.
Beckett, S., 1953. *Waiting for Godot: A Tragicomedy in Two Acts*. Version: Grove Press, 2011
Borenstein, G. (coder), 2015. *Generated Detective: A NaNoGenMo Comic* (computer generated comic): https://github.com/dariusk/NaNoGenMo-2014/issues/70

Borges, J.L., 1932. “Narrative Art and Magic” (Original title: “El arte narrativo y la magia”). Version: *Selected Non-Fictions*, trans. by Esther Allen, Suzanne Jill, and Eliot Weinberger, Penguin Random House, 2000

-, 1941. “The Library of Babel,” *The Garden of Forking Paths* (Original title: “La biblioteca de Babel,” *El jardín de los senderos que se bifurcan*). Version: *Collected Fictions*, trans. by Andrew Hurley, Penguin Books, 1999

Brancato, Ch. (creator and producer), 2015-2017. *Narcos* (TV series). Netflix

Brooker, Ch. (writer), “Be Right Back” (episode), *Black Mirror* (TV series – S2E1). Neflix

Brown, D., 2003. *The Da Vinci Code*. Doubleday

Bukowski, Ch., 1979. “Break-In,” *The Bell Tolls for No One*. Version: CityLight, 2015

Burney, S., 1812. *Traits of Nature*. Henry Colburn

-, 1820. *Tales of Fancy: Country Neighbours*. Version. Gale NCCO, 2017

Burney, F. (a.k.a. Mme. D’Arblay), 1782. *Cecilia or Memoirs of an Heiress*. Version: Oxford University Press, 2009

Butor, M., 1957. *Second Thoughts* (Original title: *La Modification*). Version: Trans. by Jean Stewart, Faber&Faber, 1958

Cathcart, A., 1923. *A Story of Modern Life*. Version: Nabu Press, 2011

Chesterton, G.K., 1938. “Second Childhood.” Version: *The Collected Works of G.K. Chesterton (Collected Poetry)*, Vol. 10, Ignatius Press, 2008

-, 1910-1936. *The Complete Father Brown Stories* (comp.). Version: Wordsworth Editions Ltd, 1992

-, 1908. *Orthodoxy*. Version: Moody Publishers, 2009

Cameron, J. (director, writer, and co-producer), 1997. *Titanic* (film). Paramount Pictures

Conan Doyle, A., 1887. *A Study in Scarlet*. Version: Filiquarian, 2007

Connelly, M., 2014. *The Burning Room*. Little, Brown & CO.

Cronenberg, D. (producer and director), 2002. *Spider* (film). Sony Pictures Classics

Dickens, Ch., 1859. *A Tale of Two Cities*. London: Chapman & Hall

Dylan, B. (author, composer, and interpreter), 1955. “Like a Rolling Stone,” *Highway 61 Revisited* (album). New York: Columbia

Dostoevsky, F., 1866. *Crime and Punishment* (Original title: *Prestupléniye i nakazániye*). Version: Trans. by Constance Garnett, Dover Publications, 2001

Duffer, M.; Duffer, R.(creators, writers, and directors), 2016-present. *Stranger Things*. Netflix

Eggers, D., 2013. *The Circle*. Knopf

Ellison, R., 1952. *Invisible Man*. Random House

Faulkner, W., 1929. *The Sound and the Fury*. Jonathan Cape and Harrison Smith

Frost, R., 1923. “Stopping by Woods on a Snowy Evening,” *New Hampshire*. Henry Holt

Fullwood, M. (coder), 2015. *Twide and Twejudice* (computer generated text): https://github.com/dariusk/NaNoGenMo-2014/issues/51

García Márquez, G., 1967. *One Hundred Years of Solitude* (Original title: *Cien años de soledad*). Version: Gregory Rabassa, Harper&Row, 1970

Gillian, V. (creator and producer), 2008-13. *Breaking Bad* (TV series). Sony Pictures Television

Hemingway, E., 1952. *The Old Man and The Sea*. Charles Scribner’s Sons

-, 1927. “The Killers.” *Scribner’s Magazine*. Charles Scribner’s Sons

Hitchcock, A. (director), 1953. *The Birds* (film). Alfred Hitchcock Productions

-, 1964. *Marnie* (film). Geoffrey Stanly Productions

Homer, c. 800BC-700BC, *The Odyssey* (Original title: *Odysseia*). Version: Trans. by Walter Shewring, Oxford University Press, 1980

-, c. 1260BC-1180BC, *The Iliad* (Original title: *Iliás*). Version: Trans. by Robert Fitzgerald, Oxford University Press, 1998
Harsdörffer, G.P., 1651. *Five-fold Thought-ring of the German Language* (volvelle) (Original title: *Fünffacher Denckring der Teutschen Sprache*). Source: Schäfer, 2006 (v. supra)
Humdrum, L. (pseud.), 1820. *Domestic Scenes: A Novel*. Version: Gale NCCO, 2017
Ionesco, E., 1994. *Rhinoceros and other plays*. Trans. by Derek Prouse. Groove Press
James, E.L., 2011. *Fifty Shades of Grey*. Vintage Books
James, H., 1898. *The Turn of the Screw*. New York: The Maximilian Company
Jardine, L., 2010. "C.P. Snow's Two Cultures Revisited." Christ's College Cambridge: 49-57
Jefferson, Th., et al., 1776. *United States Declaration of Independence*. Version: Georgetwon University Press, 2003
Jonze, S. (writer, director, and producer), 2013. *Her* (film). Warner Bros. Pictures
Kafka, F., 1915. *The Metamorphosis* (Original title: *Die Verwandlung*). Version: Trans. by Karen Reppin, Prague: Vitalis Verlag, 2017
-, 1959. *The Trial* (Original title: *Der Prozess*). Version: Trans. by Breon Mitchell. Schocken
Kazemi, D. (coder), 2015. *Teens Wander Around a House* (computer generated text): https://tinysubversions.com/nanogenmo/novel-2.pdf
Lister, Th., 1828. *Hebert Lacy*. Henry Colburn
Melville, H., 1851. *Moby-Dick or The Whale*. Version: Bryant, J.; Springer, H. (eds.), New York: Longman, 2009
Molière, 1664. *Tartuffe or The Impostor* (Original title: *Tartuffe ou l'impostor*). Version: Trans. by Martin Sorrel. London: Nick Hern Books, 2002
Monterroso, A., 1954. "The Dinosaur" (Original title: "El Dinosaurio"). Version: Trans. by Edith Grossman, *Complete Works and Other Stories*. Texas Pan American, 1955
Nabokov, V., 1955. *Lolita*. Olympia Press
Orwell, G., 1949. *Nineteen Eighty-Four*. Secker & Warbung
Pickett, R., 2004. *Sideways*. St. Martin's Griffin
Picoult, J., 2010. *House of Rules*. Atria
Piroyansky, M., 2018. "Autocorrect." UN3TV: https://www.youtube.com/watch?v=SMqkIPd7ubI&index=7&list=PLGIP2SzAkWMatkHO03-mkd-t5w4C0uPbQ
Poe, E.A., 1839. "William Wilson," *Tales of the Grotesque and the Arabesque*. Burton's Gentleman's Magazine
Puff Daddy (pseud.); Evans, F., (authors, composers, and interpreters) 1997. "I'll be missing you" (song, single). Bad Boy Records
Rhodes, L., 2015. *Playing Games: A Billionaire Romance*
Ritchie, K. & B., 2013. *Addicted to You*. K.B. Ritchie
Rowling, J.K. (pseud.: Robert Galbraith), 2013. *The Cuckoo's Calling*. Little Brown & Co.
-, 1997-2007. *Harry potter* (novel series). Bloomsbury Publishing
Saroyan, A., 1962. "Lighght." Version: *Complete Minimal Poems* Ugly Duckling Presse, 2007
Semple, M., 2012. *Where'd You Go, Bernadette*. Little, Brown & Co.
Shakespeare, W., c. 1599-1602. *The Tragedy of Hamlet, Prince of Denmark*. Version: Bevington, D. (ed.), *The Complete Works of Shakespeare*, 5th Edition, Longman, 2003
-, 1597. *Romeo and Juliet*. Version: Bevington, D. (ed.), *The Complete Works of Shakespeare*, 5th Edition, Longman, 2003
Shapiro, J., 2010. *Contested Will: Who Wrote Shakespeare?* Faber and Faber
Shōnagon, S., c. 990-1000. *The Pillow Book* (Original title: *Makura no Sōshi*). Version: Trans. by Arthur Waley, Vermont: Rutland, 2011

Simmias of Rhodes, c. 300BC. “Pelekys” (visual poem). Source: Marcault, 2015 (s. supra)
Snow, C.P., 1959, *The Two Cultures and the Scientific Revolution*.
Sophocles, 429BC. *Oedipus The King* (Original title: *Oedípus Tyránnos*). Version: Trans. by Stephen Berg and Diskin Clay, Oxford University Press, 1988
Stevenson, R.L., 1886. *The Strange Case of Dr. Jekyll and Mr. Hyde*. Longmans, Green & Co.
Scholl, Z., (coder) 2011. “For the Bristecole Snag” (computer generated text, by software *Poetry-Generator* (poetry generator). GitHub: https://github.com/schollz/poetry-generator.
Schwenke, K. (coder), 2014. “Earthquake: 4.7 quake strikes Los Angeles, centered near Westwood” (computer-generated text, by software Quakebot). *Los Angeles Times*: http://www.tronc.com/gdpr/latimes.com/
Sting (pseud. of Gordon Matthew Thomas Sumner) (author, composer, and co-interpreter), 1983. “Every Breath You Take,” in: The Police (int.), *Syncrhonicity* (album). A&M. Video version: ThePoliceVEVO, “The Police – Every Breath You Take,” YouTube: https://www.youtube.com/watch?v=OMOGaugKpzs
Thomas, D., 1947. “Do not go gently into that good night.” Version: *In Country Sleep and Other Poems*, Dent, 1952
Tolkien, J.R.R., 1954-55. *The Lord of the Rings*. Version: 50th Anniversary edition, Mariner Books, 2005
Tolstoy, L., 1888. “Kholstomer: The Story of a Horse.” Version: Kessinger Publishing, 2010
Tennyson, 1798. *The Rime of the Ancient Mariner*. Version: Arcturus Publishing Ltd, 2010
Yeh, M., 1991. *Chinese Contemporary Poetry: Theory and Practice since 1917*. Yale University Press
VV.AA. (coders), 2008. *TrueLove.wrt, An Impeccable Novel*. (computer generated text, by software PC Writer 1.0). Astrel-SPb
VV.AA., c. 1706-21. *One Thousand and One Nights* (Original title: *Alf Layla wa-layla*). Version: *The Book of the Thousand Nights and a Night - A Plain and Literal Translation of the Arabian Nights Entertainments*, trans. by Robert Burton, Karma Shastra Society
VV.AA., c. 4000BC-96AD. *The Bible* (Original title: *Ta biblía*). Version: *Oxford Annotated Bible* (5th Ed.), Oxford University Press, 2018
Waddington, J.R., 1838. *Misrepresentation or Scenes in Real Life*. Version: Rare Books Club, 2012

DIGITAL PLATFORMS (WEBSITES, DATABASES, AND PROGRAMS)

Botpoet - *Bot or Not* (poetry generator): http://botpoet.com/
DeepFace (facial recognition system). Facebook
ELIZA (chatbot), by Weizenbaum, J. (coder). MIT Artificial Intelligence Laboratory
Eternime (chatbot): http://eterni.me/)
Eugene Goostman (chatbot), by Vaselov, V.; Demchenko, E.; Ulasen, S. (coders), 2001
Eurotra (automatic translator). European Commission, 1978
Google Books Ngram Viewer (text-mining software): https://books.google.com/ngrams
Google Forms (surveys creator): https://www.google.com/forms/about/
Google Sheets (spreadsheet): https://www.google.com/sheets/about/
Google Translate (automatic translator): https://translate.google.com/
Google Trends (searches miner): https://trends.google.com/trends/

Metaphor-a-Minute! (metaphors generator), by Kazemi, D. (coder),. In *Twitter*: https://twitter.com/metaphorminute?lang=en

You Must Be (pick-up-lines generator), by Kazemi, D. (coder): http://tinysubversions.com/stuff/youMustBe/

NaNoGenMo (National Novel Generation Month) (internet-based creative code-based writing project): https://nanogenmo.github.io/. Based on an idea by Darius Kazemi (s. https://twitter.com/tinysubversions/status/396305662000775168)

NeuralTalk, 2018 (sentence generator) - https://github.com/karpathy/neuraltalk

PARRY (chatbot), by Colby, K. (coder), 1972. Stanford University

Poem.exe (poetry generator): https://poemexe.com/

Poetry-Generator (poetry generator), by Scholl, Z. (coder), 2011, GitHub: https://github.com/schollz/poetry-generator

Postmodern Generator (paper generator): http://www.elsewhere.org/journal/pomo/

Recurrent Neural Network (image recognition and story generator), by Kiros, R. (coder), 2017

Replika (chatbot). Luka: https://replika.ai/

SCIgen (paper generator): https://pdos.csail.mit.edu/archive/scigen/

Systran (automatic translator). Dr. Peter Toma, 1968

Voyant (web-based reading and analysis environment for digital texts): https://voyant-tools.org/

Words Frequency Data (wordlist based on the Corpus of Contemporary American English – COCA), 2018. Created by Mark Davis, Birgham Young University: https://www.wordfrequency.info/free.asp

APPENDIX 1 - SURVEY QUESTIONS - How do you interpret this song?

Hi,

I invite you to participate in this survey, which is part of my doctoral research at the Friedrich Schlegel Graduate School for Literary Studies (Free University of Berlin). The purpose is to discover patterns and tendencies in the interpretive behaviour of different people.

You just have to read the lyrics of the song "Every Breath You Take" (Sting, 1983), and then fill out the following form.

(Completion time: 4-6 min).

Please, share. Thanks!

Martín Azar
PhD Student - FSGS FU Berlin

* Required

"Every Breath You Take" (Sting, 1983)

Every breath you take
Every move you make
Every bond you break
Every step you take
I'll be watching you

Every single day
Every word you say
Every game you play
Every night you stay
I'll be watching you

Oh can't you see
You belong to me
My poor heart aches
With every step you take

Every move you make
Every vow you break
Every smile you fake
Every claim you stake
I'll be watching you

Since you've gone I been lost without a trace
I dream at night I can only see your face
I look around but it's you I can't replace
I feel so cold and I long for your embrace
I keep crying baby, baby, please

Oh can't you see
You belong to me
My poor heart aches
With every step you take

Every move you make
Every vow you break
Every smile you fake
Every claim you stake
I'll be watching you
Every move you make
Every step you take
I'll be watching you

I'll be watching you
(Every breath you take, every move you make, every bond you break, every step you take)
I'll be watching you
(Every single day, every word you say, every game you play, every night you stay)
I'll be watching you
(Every move you make, every vow you break, every smile you fake, every claim you stake)
I'll be watching you
(Every single day, every word you say, every game you play, every night you stay)
I'll be watching you
(Every breath you take, every move you make, every bond you break, every step you take)
I'll be watching you
(Every single day, every word you say, every game you play, every night you stay)
I'll be watching you

Personal data

Tell us just a little bit about yourself

1. **What's your age?** *

____

2. **Where do you come from? (Country)** *

____

3. **What's your gender?** *
   Mark only one oval.
   - Male
   - Female
   - Other

4. **What's your occupation?** *

____

Questions about the lyrics of the song

Now you get to play the literary critic

5. **What do you think the word “watching” means when used in these lyrics? ***

Rate each option with a value from 0 to 10, 0 being "In these lyrics, 'watch' doesn't mean that at all" and 10 being "It means precisely that."

Mark only one oval per row.

	0	1	2	3	4	5	6	7	8	9	10
To look at (as in “He watches TV”)											
To pay attention (as in “The company watches the progress of the stock market”)											
To care for, to look after (as in “While you're out, I'll watch the kids”)											
To monitor secretly (as in “His phone is tapped, he’s being watched”)											
Other											

6. **What do you think these lyrics are about? ***

Rate the pertinence of each topic from 0 to 10, 0 being "These lyrics aren't at all about this" and 10 being "They are precisely about this."

Mark only one oval per row.

	0	1	2	3	4	5	6	7	8	9	10
Love											
Stalking											
State surveillance and control											
Death and afterlife											
Other											

7. **What do you think most people think these lyrics are about? ***

Rate the pertinence of each topic from 0 to 10, 0 being "Nobody thinks these lyrics talk about this" and 10 being "Everybody thinks they talk about this."

Mark only one oval per row.

	0	1	2	3	4	5	6	7	8	9	10
Love											
Stalking											
State surveillance and control											
Death and afterlife											
Other											

8. **What do you think the author meant to talk about in these lyrics? ***

Rate the pertinence of each topic from 0 to 10, 0 being "The author didn't mean to talk about this at all" and 10 being "The author meant to talk precisely about this."

Mark only one oval per row.

	0	1	2	3	4	5	6	7	8	9	10
Love											
Stalking											
State surveillance and control											
Death and afterlife											
Other											

9. **Do you think the author (the real person who wrote these lyrics) and the speaker (the voice that says “I’ll be watching you”) as being the same person?** *

If you think the lyrics are ambiguous about this, then chose a value within 0-5 or 5-10 depending on which pole you think as more evident or likely.
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Definitely NOT	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Definitely YES

10. **How good and smart do you think the author (who wrote the lyrics), the speaker (who says "I'll be watching you"), and the recipient (the "You" who allegedly will be watched) might be?** *

Rate each one from 1 to 10 in terms of morals and intelligence, 0 being very bad and very stupid, and 10 being very good and very smart. If you think of two of them as being the same person, just give them the same score.
Mark only one oval per row.

	0	1	2	3	4	5	6	7	8	9	10
Author - Morals	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Author - Intelligence	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Speaker - Morals	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Speaker - Intelligence	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Recipient - Morals	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Recipient - Intelligence	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

11. **How do you feel about each of them?** *

Chose the emotion each one produces in you the most.
Mark only one oval per row.

	Admiration	Contempt	Envy	Pity	Trust	Distrust	Inferiority (you feel they're better than you)	Superiority (you feel they're worse than you)
Author (who wrote the lyrics)	◯	◯	◯	◯	◯	◯	◯	◯
Speaker (who says "I")	◯	◯	◯	◯	◯	◯	◯	◯
Recipient (the "You")	◯	◯	◯	◯	◯	◯	◯	◯

12. **With whom do you identify the most?** *

From whose perspective do you read these lyrics? Do you experience the lyrics more as if you had written them, as if you were telling them, or as if they were being told to you?
Mark only one oval.

- ◯ Author (who wrote the lyrics)
- ◯ Speaker (who says "I")
- ◯ Recipient (the "You")
- ◯ None

13. **How do these lyrics as a whole make you feel? ***

Chose a number from 1 - 10 indicating the degree to which you experience each of the following emotions, 0 being "The lyrics don't make me feel this at all" and 10 being "This is precisely what they make me feel."

Mark only one oval per row.

	0	1	2	3	4	5	6	7	8	9	10
Happiness											
Sadness											
Anger											
Fear											
Enjoyment											
Disgust											
Other											

14. **How much do you like these lyrics? ***

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
I hate them												I love them

15. **If you want, you can add a comment about these lyrics.**

It can be whatever you want, from a word to a paragraph.

Powered by Google Forms

APPENDIX 2 - SURVEY RESULTS

Age	Country	Gender	Occupation	Watching - To look at	Watching - To pay attention	Watching - To care	Watching - To monitor	Watching - Other	You - Love	You - Stalking	You - State surveillance	You - Death	You - Other	Others - Love
63	Argentina	Female	Retired	0	0	8	0	8	8	0	0	0	0	8
28	Romania	Female	Architect	0	3	0	8	0	2	8	0	0	0	6
29	Italy	Female	Teacher	6	0	1	8	0	1	8	2	2	0	5
25	France	Female	Student	0	4	3	5	0	4	5	0	0	0	4
32	Serbia	Female	Editor	0	5	5	9	0	5	8	8	4	0	8
28	Romania	Female	Architect	0	0	3	7	0	8	7	0	0	0	10
30	Italy	Female	Teacher	3	5	6	8	1	9	7	0	0	0	10
24	Russia	Female	Student	2	10	6	10	0	7	10	0	0	0	9
34	Argentina	Female	Professor	0	7	5	10	0	5	8	0	0	10	8
31	Argentina	Female	Scritpwriter	1	6	7	4	1	7	6	0	7	0	8
30	Italy	Female	Teacher	3	5	6	8	1	9	7	0	0	0	10
28	Argentina	Female	Professor	0	7	2	10	2	2	10	5	1	1	9
26	Grece	Female	Teacher	6	0	9	10	0	10	7	5	9	0	10
21	Argentina	Female	Student	2	2	1	1	2	0	2	1	1	2	2
22	Argentina	Female	Student	7	0	3	9	0	0	9	9	0	0	10
25	Romania	Female	Student	0	0	0	10	5	10	10	5	0	0	10
29	Argentina	Female	Teacher	0	8	0	10	0	0	10	10	8	0	9
28	Germany	Female	Student	0	0	8	8	0	8	0	0	8	0	8
34	Spain	Female	Marketing	0	0	0	8	0	3	10	0	3	0	10
22	Argentina	Female	Student	0	0	1	2	0	0	2	1	0	0	2
34	costa rica	Female	Student	0	0	2	7	0	8	7	8	7	0	8
35	Costa Rica	Female	Teacher	8	10	0	0	1	4	0	0	0	10	10
32	Argentina	Female	Editor	10	0	0	0	2	0	2	0	10	3	2
29	Argentina	Female	Lawyer	0	5	5	10	5	5	10	0	0	5	10
22	Argentina	Female	Student	0	6	2	4	0	1	9	7	0	0	6
63	Argentina	Female	IT	0	0	10	0	0	10	0	0	0	0	10
22	Argentina	Female	Student	8	9	2	7	0	10	9	9	1	0	10
32	Argentina	Female	Researcher	8	8	0	10	0	0	10	10	0	4	8
30	argentina	Female	Journalist	0	0	0	7	0	0	8	4	1	0	5
59	argentina	Female	Teacher	0	4	8	0	8	8	4	8	0	8	5
28	Romania	Female	Student	6	8	7	3	5	10	3	0	0	10	8
24	Greece	Female	Student	0	0	8	0	0	8	0	2	6	0	8
32	Macedonia	Female	Teacher	4	6	7	4	0	7	3	1	6	0	8
26	croatia	Female	Translator	0	0	1	0	2	2	0	0	2	0	2
34	Argentina	Male	Business	0	4	2	4	0	6	4	0	3	0	8
32	Argentina	Male	Student	10	10	7	8	6	10	9	8	0	1	10
34	Uruguay	Male	Editor	5	8	10	0	0	10	0	0	0	0	10
30	Chile	Male	Student	0	0	1	0	2	10	0	0	4	0	10
32	Argentina	Male	Musician	0	1	1	1	2	1	1	1	2	1	2
35	Argentina	Male	IT	0	2	0	8	0	7	10	0	0	0	9
32	Argentina	Male	Student	8	8	2	9	2	6	9	0	0	1	9
34	Argentina	Male	Assistant	8	6	6	2	0	6	8	2	0	0	8
63	Argentina	Male	Employee	0	0	0	2	1	0	2	2	1	0	2
24	Romania	Male	Student	0	1	2	1	0	8	5	0	0	0	9
24	Spain	Male	Teacher	5	0	10	10	0	10	10	0	0	0	10
28	Romania	Male	Student	0	0	5	10	7	10	10	0	7	0	10
32	Argentine	Male	Musician	0	0	0	0	10	0	8	10	0	0	10
29	Argentina	Male	Student	0	0	6	6	6	7	7	7	0	0	8
35	Spain	Male	Engineer	0	5	8	0	0	10	5	0	10	0	10
28	Uruguay	Male	Filmmaker	10	10	5	0	0	5	2	2	5	0	10
21	Ecuador	Male	Student	0	0	1	2	0	1	2	0	0	0	2
32	Argentina	Male	Artist	0	8	5	10	0	0	10	10	0	0	10
63	Argentina	Male	Dentist	1	10	10	10	0	10	0	0	0	0	10
26	Italy	Male	Student	0	0	10	0	8	10	0	0	0	0	10
32	Argentina	Male	Student	10	10	7	8	6	10	9	8	0	1	9
35	Argentina	Male	Employee	2	1	2	0	0	2	1	0	0	0	2
32	Italy	Male	Student	7	7	7	0	0	10	0	0	10	6	10
34	Argentina	Male	Lawyer	0	0	0	9	8	4	7	10	3	0	5
60	USA	Male	IT	1	1	0	8	0	0	8	0	0	0	5
33	Argentina	Male	IT	2	2	8	3	3	10	0	0	0	0	10
34	Germany	Other	Programmer	2	6	7	2	5	7	5	1	1	1	7
25	Mexico	Other	Student	0	10	0	10	0	10	10	10	10	10	10

Others - Stalking	Others - State surveillance	Others - Death	Others - Other	Author - Love	Author - Stalking	Author - State surveillance	Author - Death	Author - Other	Author=Speaker?	Author - Morals	Author - Intelligence	Speaker - Morals	Speaker - Intelligence	Recipient - Morals
4	1	3	2	8	0	0	0	0	10	6	6	6	6	6
3	0	0	4	3	5	0	0	0	7	5	6	5	6	4
7	4	3	0	7	2	2	5	0	5	1	8	1	7	5
0	0	0	0	2	5	0	0	0	2	5	5	1	2	3
6	6	4	0	6	5	7	6	0	3	5	5	5	5	5
5	0	0	0	10	3	1	1	1	3	5	5	5	5	5
5	0	0	0	10	0	0	0	3	5	5	5	5	5	5
9	0	0	0	10	0	0	0	0	3	2	5	2	5	5
6	1	1	5	9	3	1	1	0	1	2	0	2	0	0
7	3	7	0	5	2	2	5	1	3	2	2	2	2	2
5	0	0	0	10	0	0	0	3	5	5	5	5	5	5
5	0	2	2	8	3	0	4	0	4	6	7	7	5	7
9	7	8	0	10	2	2	6	0	7	5	10	5	5	5
0	0	0	1	0	2	2	1	2	5	2	2	1	1	2
0	0	0	0	8	0	0	0	0	5	0	0	0	0	8
6	4	4	5	10	0	0	0	10	8	10	10	0	10	0
1	1	5	0	0	10	9	5	0	0	10	10	0	2	8
0	0	8	0	8	0	0	8	0	10	4	4	4	4	4
6	0	5	0	8	8	0	5	0	7	6	6	6	6	6
1	1	1	0	1	2	1	0	0	10	1	1	1	1	1
3	3	0	0	8	5	4	0	0	2	4	4	4	4	4
0	0	0	0	10	0	0	0	0	5	1	2	1	1	5
7	7	2	3	4	2	1	10	4	2	5	5	5	5	5
2	0	1	5	10	0	0	0	0	10	5	5	5	5	5
1	1	0	0	2	8	5	0	0	2	1	2	2	1	1
0	0	0	0	10	0	0	0	0	10	7	7	7	7	7
6	6	4	1	9	3	0	5	4	2	7	8	3	9	8
5	5	3	4	0	10	10	0	8	0	9	9	0	8	8
6	2	1	0	7	0	0	6	0	10	2	8	0	6	1
5	5	0	4	8	4	4	8	8	0	0	0	0	0	0
7	0	6	4	10	0	0	1	3	0	6	8	5	5	5
4	4	7	0	7	1	4	7	0	9	7	7	3	3	5
4	2	7	0	7	2	4	7	0	9	6	6	6	6	6
0	0	0	0	2	0	0	2	0	7	1	0	1	0	1
4	1	3	0	8	6	0	5	0	10	7	7	8	6	5
9	4	4	5	10	7	8	0	1	7	8	8	7	7	5
5	0	3	0	10	0	0	5	0	10	6	6	6	6	6
0	0	2	1	10	0	0	1	1	4	5	4	5	4	2
0	0	1	1	1	1	1	1	1	5	2	1	1	1	0
8	0	0	0	8	2	0	0	0	0	5	9	3	2	5
7	2	2	2	10	1	0	0	0	1	5	5	2	5	5
4	2	2	0	8	8	2	2	0	10	8	8	8	8	8
0	0	1	1	0	2	0	2	1	5	1	2	0	1	0
4	3	2	1	10	0	0	4	0	9	3	5	3	4	3
4	0	0	0	10	10	0	0	0	0	5	5	5	5	5
3	0	0	0	10	10	0	0	0	5	10	10	10	5	7
5	0	0	8	0	5	10	0	0	10	0	10	0	10	0
7	6	0	0	8	0	8	0	0	5	6	6	6	6	6
10	0	5	0	10	0	0	10	0	0	7	7	5	5	5
4	4	8	0	10	5	5	10	0	7	5	8	5	5	5
2	0	0	0	2	2	0	0	0	5	0	0	0	0	0
0	0	0	0	0	8	8	0	0	10	5	8	2	8	5
0	0	0	0	10	0	0	0	0	10	7	7	7	7	7
5	0	0	0	2	0	0	0	0	7	2	2	2	2	1
5	3	1	1	10	2	3	0	1	8	9	7	4	4	5
1	0	0	0	2	1	0	0	0	10	2	2	1	1	1
0	0	10	6	8	0	0	8	8	0	5	5	5	5	5
10	9	0	5	5	8	7	6	10	4	5	9	2	9	6
3	1	1	0	0	8	0	0	0	0	8	8	1	5	6
0	0	0	0	10	0	0	0	0	10	6	6	6	6	6
6	2	4	1	7	1	1	5	4	2	6	7	3	5	4
5	5	0	0	9	9	9	0	10	5	5	10	5	10	5

Recipient - Intelligence	Feeligs - Author	Feelings - Speaker	Feelings - Recipient	Identification	Lyrics - Happiness	Lyrics - Sadness	Lyrics - Anger	Lyrics - Fear	Lyrics - Enjoyment	Lyrics - DIsgust	Lyrics - Other	Likes
6	Other	Other	Other	Recipient	0	10	0	0	0	0	0	7
6	Other	Other	Other	None	1	4	1	1	2	0	0	6
1	Distrust	Distrust	Pity	Recipient	2	4	4	5	2	5	0	2
3	Trust	Contempt	Pity	Speaker	0	2	5	3	0	5	0	5
5	Contempt	Contempt	Contempt	Recipient	0	0	0	0	4	0	0	5
5	Pity	Pity	Admiration	None	0	5	0	3	1	0	10	3
5	Admiration	Pity	Pity	Speaker	3	8	0	0	4	0	0	6
5	Contempt	Pity	Pity	Speaker	0	7	3	7	0	3	0	4
0	Distrust	Distrust	Pity	None	0	0	3	1	0	1	0	2
2	Contempt	Pity	Distrust	None	3	5	1	1	5	1	0	3
5	Admiration	Pity	Pity	Speaker	3	8	0	0	4	0	0	6
7	Distrust	Distrust	Pity	None	0	7	8	9	0	7	2	1
7	Contempt	Pity	Superiority	Recipient	0	10	3	6	2	2	1	7
2	Admiration	Admiration	Contempt	None	1	2	1	2	1	0	2	7
8	Pity	Pity	Pity	None	0	10	7	8	0	7	5	1
0	Distrust	Superiority	Distrust	None	0	7	0	0	0	0	8	5
8	Admiration	Contempt	Pity	Author	0	7	10	7	0	10	0	7
4	Superiority	Superiority	Superiority	None	2	5	0	5	4	0	0	6
6	Distrust	Distrust	Inferiority	Recipient	0	0	8	10	0	8	0	2
1	Contempt	Contempt	Contempt	None	0	0	0	0	1	1	1	3
4	Pity	Pity	Pity	Speaker	0	4	0	0	0	0	0	7
5	Trust	Pity	Contempt	Speaker	0	5	5	0	5	0	0	5
5	Admiration	Admiration	Admiration	Speaker	1	8	2	0	8	0	2	5
5	Distrust	Distrust	Pity	None	0	5	9	9	5	10	5	0
1	Distrust	Distrust	Trust	Recipient	1	2	0	2	0	1	0	1
7	Pity	Pity	Trust	None	0	8	0	0	0	0	0	7
7	Contempt	Distrust	Pity	None	7	0	1	0	8	1	0	6
8	Admiration	Contempt	Pity	None	0	0	4	4	0	7	7	3
5	Contempt	Contempt	Pity	None	0	1	8	8	0	7	0	0
3	Pity	Pity	Pity	None	0	0	0	0	0	0	8	2
5	Admiration	Pity	Contempt	Speaker	2	6	0	0	4	0	7	7
5	Admiration	Admiration	Admiration	Author	6	6	0	0	3	0	0	8
6	Pity	Pity	Pity	Author	0	2	0	0	1	0	5	7
1	Contempt	Admiration	Contempt	Author	1	0	0	1	1	0	0	6
3	Admiration	Contempt	Pity	Speaker	0	7	0	8	4	0	0	9
5	Admiration	Contempt	Pity	Speaker	0	6	7	8	2	0	0	8
6	Sympathy	Sympathy	Sympathy	Author	4	7	0	0	6	0	1	7
2	Pitty	Pitty	Schadenfreude	Author	2	9	0	0	0	0	2	5
0	Admiration	Contempt	Contempt	Author	1	1	0	0	2	0	2	8
5	Admiration	Pity	Pity	None	2	0	0	2	9	7	0	5
5	Pity	Pity	Pity	None	5	5	5	7	1	7	5	5
8	Admiration	Pity	Trust	None	3	8	3	5	5	5	0	7
1	Contempt	Envy	Admiration	Speaker	0	1	0	0	0	0	1	8
3	Contempt	Admiration	Contempt	Speaker	1	6	2	0	1	1	1	6
5	Admiration	Admiration	Admiration	Speaker	10	8	0	4	10	0	0	8
5	Admiration	Contempt	Envy	None	4	7	0	3	4	0	0	7
0	Superiority	Superiority	Distrust	None	0	5	0	0	6	0	10	2
6	Admiration	Distrust	Trust	Recipient	5	3	2	2	4	0	3	2
5	Trust	Pity	Pity	Speaker	5	8	1	1	6	1	1	7
0	Contempt	Pity	Pity	Speaker	0	10	0	5	5	7	0	8
0	Contempt	Contempt	Contempt	None	0	0	0	0	0	1	0	0
5	Pity	Pity	Distrust	Author	8	8	4	6	3	3	10	8
7	Pity	Pity	Pity	Author	0	10	7	0	0	0	0	8
1	Admiration	Admiration	Envy	Speaker	2	0	0	0	2	0	0	7
5	Admiration	Pity	Superiority	Author	5	7	1	0	10	2	0	5
1	Admiration	Contempt	Envy	Speaker	1	2	1	0	0	1	0	8
5	Superiority	Superiority	Superiority	None	0	0	0	0	0	0	10	5
4	Admiration	Admiration	Pity	None	6	0	0	0	7	0	0	6
5	Admiration	Distrust	Pity	Author	2	5	3	3	5	4	8	6
6	Admiration	Trust	Admiration	Author	2	7	0	0	6	0	0	9
4	Admiration	Pity	Admiration	Speaker	1	5	2	1	1	1	4	7
10	Admiration	Distrust	Envy	Recipient	9	2	0	0	8	9	10	8