I brought to you the metaphor of the pebbles ground, below which surface I invite you to look.

(Of course, this works better when said in a Romance language like Portuguese, because from the Latin word BASSUS we inherited a root that is present in lots of polysemic words and expressions. Translated into English, Portuguese ‘baixo’ can mean either low or short, even popular, and mean. Combined with a preposition, ‘debaixo, abaixo, em baixo’, it can mean both under, and below.)

I have also chosen this picture as a background image for my slides because in it we can see, simultaneously, all the attributes of the manuscripts I came here to comment: they are banal things, small in size, large in number, pretty natural, and shaped by time.

I’ll give you a short report on the work we are carrying out in Lisbon -- and allow me to thank here this Conference’ organizers for having invited me to Berlin and make exactly such a report.
At the Linguistics Centre of the University of Lisbon, a mixed team of Portuguese and Spanish historians and linguists has been preparing a digital edition of a few thousands of ordinary writings of the Early Modern period. The project, in the present form, is called P.S. Post Scriptum. Our online edition aims to publish until 2017 a collection of socially representative personal 7,000 letters, half in Spanish, half in Portuguese, written between the 16th and the early 19th century and kept inside court proceedings of the Inquisition and the Civil Courts as material proof on crimes.

Because of their instrumental value for the judges that assessed the letters’ contents, these manuscripts were archived near the hearings of court witnesses, so there is plenty of material to help us contextualize the letters’ original intervening actors. They were mostly just common people writing about love, hatred, friendship, revenge, or just gossip, poor health, and lack of things.

In order to fit into this short presentation the main lines of our work and the guidelines that inspire us, I would like to proceed with only two questions, which we are finding crucial to this type of research:

First, how do we close the cultural gap between us -- learned scholars -- and them -- the ordinary people in history whose writing behaviour we are studying?

And second, how do we lend some general coherence to the hundreds of themes, thousands of stories, millions of words we are facing when exploring these ordinary papers?

As for the first question, which has to do with the cultural gap between the observer and the observed, it has been an old concern in the social sciences, especially in ethnography. One of the obstacles is this: in the study of non-elite cultures there is the
tendency for researchers, who are themselves learned subjects, to view their object with a patronizing attitude.

Roger Chartier phrased the problem using a well-known aphorism: “Popular culture is a category of the learned” (Chartier, 1995, p. 83). But even more eloquent than Chartier’s formulation, is the line that play-writer Aaron Sorkin made his character Colonel Jessup say when he addressed this issue in the play “A Few Good Men”.

Colonel Jessep is an arrogant marine officer and in this scene he comments on the behaviour of a simple soldier, the private William Santiago. Jessep says in court, as perhaps you can remember since the play became a famous Hollywood movie by film director Rob Reiner: *I'm an educated man, but I'm afraid I can't speak intelligently about the travel habits of William Santiago.*

The context:

**Kaffee**: Is this funny, sir?

**Col. Jessep**: *[face falls to a look of disgust]* No, it isn't. It's tragic.

**Kaffee**: Do you have an answer to the question, Colonel?

**Col. Jessep**: Absolutely. My answer is I don't have the first damn clue. Maybe he was an early riser and liked to pack in the morning. And maybe he didn't have any friends. *I'm an educated man, but I'm afraid I can't speak intelligently about the travel habits of William Santiago.*

Although the audience **watching the play** knows very well that Colonel Jessup is lying, his answer still makes sense because the audience **living in the play** is expected to accept the existence of a cognitive constraint: - in social groups, culturally privileged individuals are pushed to reason inside the logic of their privilege condition
itself. As studied by Social Theory, this is what Pierre Bourdieu called *distinction*, and distinction is a hard limit to fight when we act socially, because culture constantly reproduces the advantages of distinction bearers. (Bourdieu, 1984 [1979]).

Back to the precise problem of learned scholars being or not being able to make sense of insignificant people writing to each other, or of privileged people writing insignificant things, its solution has been pointed out by several thinkers. I’ll choose to echo here cultural historian Michel de Certeau.

Certeau suggested that the first step consists in approaching the ordinary people we study without the armour of scientific discourse. His words, inspired by Wittgenstein’s writings, were a call to “bring together scientific practices and languages back toward their native land, everyday life.” (Certeau, 1984 [1980], p. 6)

If we want to follow Certeau’s advice, it means – and here I’m making a parallel with the example in the film “A Few Good Men” -- to **stop trying to look intelligent yourself** and **start by trying to describe the intelligence of the ordinary persons themselves**. Of course I mean ‘intelligence’ in a neutral way, a synonym of ‘conscience’ or ‘knowledge’.

The practical way of applying this professional ethics leads us, first, to dismiss its radical form, because that would be equal to taking a jump from the edge-of-cliff where the historian stands, in “unstable ground”, “between science and fiction”, like pointed out by Roger Chartier when he dismissed the importance of the ‘crisis of history’, as announced by post-modern thinkers (Chartier, 1997 [1987])

Nevertheless, it’s still an inspiring ethics, if it only means that the language and history of ordinary writings should be firstly approached at the same level as the
testimonies themselves. It’s just going back to the old point that Malinowsky wanted to make in the 1920s, when arguing for Participant-Observation in ethnography.

Above you have a visual flash of linguistic ethnography as it was practised in the early 20th century, when researchers couldn’t help distancing themselves from dialect speakers, only socializing with the local elites, dressing differently and behaving superiorly on the field (in the 1930s, you have Portuguese ethnographer Leite de Vasconcellos posing at Barrancos, then you have Spanish linguist Anibal Otero and a non-identified fieldworker, both of them collecting material for the Linguistic Atlas of the Iberian Peninsula, ALPI). At the bottom you have current linguists who work as participant-observers in the project ‘Enduring Voices’ for the Institute for Endangered Languages. They sit on the ground, dress informally and interact friendly with participants they have known for some time before the inescapable recording of their endangered languages (Institute for Endangered Languages, 2007).

Linguists and ethnographers like these ones, who study disappearing languages, they stress the social importance of such languages and put all their strength into recording and studying the few ones who still speak them, before an announced high-speed extinction. In much the same way, cultural historians and historical linguists try to devise a method of showing the dangers to knowledge, and to democracy, of dismissing everyday uses of language as if it was garbage.
But how can we approach in a way parallel to the ethnographers’ the common individuals in history that wrote the everyday writings we want to study?

We can’t level with those distant persons, of course, but … we can level with their most obvious descendants, the ordinary people of today, which we call ‘the general public’. In order to do so, we have to be able to show that letter writers and readers of the past, in much the same measure as us, today, were complex characters, living unthinkable stories, saying memorable things and using language, moment by moment, rooted on circumstances, in a creative way.

And we can only reach that general public if we use a register they know, which brings me to the Internet, and to digital editions.

I’ll use the rest of presentation to show how we are performing such levelling without losing the connection to current theory in language studies, and in history studies.

It’s a process that is always presided by the need to shrink the size of our data. This also answers the second question that I said to you earlier I wanted to address: -- How do we lend some general coherence to the hundreds of themes, thousands of stories, millions of words we are facing when exploring these ordinary papers?

And the answer is: -- Start by shrinking the size of the letters population, reducing its variation and connecting its many details.
Here we have what time did to a type of everyday writing such as a love letter, used as criminal proof because conducts of bigamy -- or of sexual harassment by priests giving confession -- were condemned by the Inquisition.

Now we have the summary of the court’s proceedings for one other letter, a sort of a short story we can write based on the case that the prosecutor could assemble at the time. In this case, the prosecuted was a Spanish priest who made many enemies in 18th century Portugal.

Now we have the facsimile of the love letter sent to that priest around 1769.
..and its encodings using TEI markup, the same used in a the Digital Edition of the Canterbury Tales, for instance, just to show that we benefit from the innovations in the field of Textual Criticism, Critique genetique, and Scholarly Digital Editions

Here is, for another letter, the reduction of spelling variation, by means of software developed at the Brazilian University of Campinas, which allows us to keep an alignment, word by word, between the authorial spelling performance and the standard spelling; we also proceed at this level to a manual annotation of the letters’ parts, the classical ones: opener, harangue, narration, peroration and closer.

Then the words are reduced into grammatical categories, just as it is usually done in Corpus Linguistics; here the software comes from Campinas for the annotation of the Portuguese texts, and from Barcelona in the case of the Spanish ones
...And the pairs word + category are reduced to annotated phrases and sentences using software that comes from the University of Pennsylvania (for Portuguese parsing), and again from Barcelona (for Spanish parsing)

The results are two annotated historical corpora that can be automatically searched at four levels:

- original spelling/lexicon/morphology/syntax

and linked to textual parts, to social profiles, and to keywords; these latter point to discursive topics and social issues at stake.

The TEI encoded file, the facsimile, a short story telling the context, the profiles of author and addressee, the text’s modernization, a translation into English and the set
of keywords are published online. So will be the annotated corpora and the linguistic information we can extract from it.

Ex:

1) all occurrences of null subject in sentences

2) all negative phrases vs 3) all negative sentences


