

Working on Language Change by using Historical Corpora: Problems and Methods

Linguists working on language change work in many different ways and on very different material bases. Some are working on languages which have been attested for a long period of their history, others work on languages which are oral, and have never been recorded, implying that earlier stages of these languages can only be reconstructed. In still other cases, only modern periods of the language have been attested.

I am working on Romance languages, in particular on French. These languages are all very well attested, as we have written texts covering more than two thousand years. One might think that it is easier to work on attested languages than to do reconstruction and this may be true, but in a way it is easier to reconstruct hypothetical stages than to work on heterogeneous data – and heterogeneous data are indeed what you find when working on what we call “text languages”, i.e. languages that we can only grasp through written texts. In the following, I will first present the heterogeneity of historical corpora, i.e. collection of historical texts, and problems about hypothesis-testing on these corpora.

We know that copyists of the Middle Ages were working on one or more original (as seen in the illustration taken from a manuscript).

Professor Lene Schøsler

Department of English,
Germanics and Romance Studies,
University of Copenhagen, Denmark
schoesl@hum.ku.dk
CAS Fellow 2004/2005



Let us just consider the simple case of one copyist working on one original and copying for a public speaking his own dialect. The text of the original may be written in a dialect that is identical to or different from the copyist's dialect and it may have been written recently or it may be old. If the text was written recently in the same dialect or in a very close variety, the copyist has an easy task of transmission. But if, on the other hand, the text of the original is written in a different dialect, or stems from an earlier stage of the same dialect, the copyist will have to decide whether he is going to modify his text or not, as these texts are normally in rhymed verse.

He may decide to copy the text as accurately as possible – and in that case he is a so-called “mirror-copyist”. Or he may decide to adapt his original to his own dialect or to the dialect of whoever has ordered the copy. In the case of adaptation, the copyist will have predictable problems stemming first and foremost from the fact that dialects have different vowel developments of the same word. Another source of adaptation may be special wishes from the person who has ordered the copy to be made. We know of cases where the future owner wants references to be inserted on the noble origin of his family, for example or more importance shown to courtly love, to religion, etc. But let me exemplify the case of adaptation due to language change.

If the language of the copyist has undergone vowel change, the copyist may decide to respect the original verse-final rhyme by replacing the unfamiliar dialect form of a given word by a regular form in his dialect which fits into the assonance or the rhyme. This can often be done without seriously affecting the sense of the passage, but in other cases we can see that the text has been seriously modified. Consider a simple case of well known vowel change from [o] to [u], spelled *ou*, and further to [ø], spelled *eu*, which took place in French during the 13th century. If we take an Old French text from the 12th century, e.g. *Charroi de Nîmes*, its manuscripts are all later, and they show that the vowel change is well on its way. Some copyists just discard the original rhyme by adapting the original unfamiliar form to the corresponding normal form of their dialects, this holds for the copyist of A1, A2, A3, A4, and C, adopting a spelling that reveals the [u] or the [ø] pronunciation of the word ‘neck’ that was originally *gole*, pronounced with an [o]. But the copyists of B1 and B2 respect the original vers-final rhyme by replacing the original word by another word, having the same vowel, meaning a piece of clothing. Such an adaptation might seem innocent, but if you want to study e.g. the system of punishment in medieval times, it makes quite a difference to have a death sentence or to have your clothes damaged!

Charroi de Nîmes verse 963, O / U / Ø assonance

ms. A1 *perdi les eulz et pendi par la **goule*** [u]

'His eyes were torn out and [he] was hung by his neck'

ms. A2 *perdi les eulz et pendi par la **goule*** [u]

ms. A3 *perdi les eulz et pendi par la **goule*** [u]

ms. A4 *perdi les euz et pendi par la **gueule*** [ø]

ms. F lacuna

ms. B1 *batus en fu, deschirée ot sa **gonne*** [o]

'[He] was hit and his clothes were damaged'

ms. B2 *batu en fu, s'ot desciré sa **gonne*** [o]

ms. C *les iels perdi et pendi par la **goule*** [u]

ms. D lacuna

Now, considering that the transmission of texts is either mirror copying or adaptation of different stages of different dialects, the language of the extant manuscript will probably come out as a chronological and dialectal mixture of features. The traditional policy of edition has followed two well known main directions : either to look for the original, by means of the method of the common error (Lachmann) or to publish one single manuscript, as it is supposed impossible to reconstruct the original (Bédier). I want to call attention to the fact that for most literary vernacular texts, both editorial policies (and especially the first one) are wrong, as they do not provide the scholar with the relevant material basis. What the scholar needs is a trustworthy version of the text tradition she or he wants to study. A Lachmanian critical edition is a construct, something that never existed. It is based on the wrong assumption, typical of the nineteenth century, that a copy is a deterioration of a (perfect) original that needs to be repaired. In bad cases the result is a fictitious text, a result of emendations and regularisations. Even Bédierien editions often correct texts, in order to provide homogeneity. But heterogeneity is inherent in Old French texts. In sum, text languages, i.e. written texts reflecting older stages of languages, are heterogeneous for two reasons: firstly because of the text transmission itself, and secondly because of the absence of a language norm.

Traditional philology from the 19th century and in the early 20th century based its editorial policy on a different, a romantic conception of the author, and they accordingly modified the texts that they published, many of which are still used in teaching and for research. This illustrates in an interesting way, I think, of how our material basis can be altered on the basis of previous scientific theories that we now believe to be wrong.

Let me briefly illustrate the difference it makes to a scholar to work on the basis of manuscripts instead of a traditional critical edition. In my study of the case system of Old French, I made two parallel studies of the declension of proper nouns in *Le Charroi de Nîmes*, one according to a critical edition and another according to the manuscripts. The study based on the edition showed that the declension system was well preserved, 82,7 % of all proper names having the expected nominative form in the functions of subject, subject predicate, apposition etc. with the exception of a few names that were mostly used in the accusative form. The study based on manuscripts showed a quite different pattern. Firstly, proper names tend to be abbreviated in a form without declension markers. Secondly, proper names of *Le Charroi* divide themselves into three

types: (1) those that are almost never spelled out and which are consequently without declension, (2) those that sometimes are found spelled out and that have a tendency to be in a fixed case, irrespective of their function, and (3) those – in fact only five proper names – that sometimes are found spelled out and that have a genuine case marking when spelled out. Let me add that this is only one case among many others where problems arise from choosing editions as the basis of research. Similar problems arise each time a hypothesis is tested on historical data. Things being like this, how can we proceed in research on text languages? We have to choose between two possibilities: either to work exclusively on manuscripts – which would of course be the best thing, but it is not possible, as mss are not immediately available, and certainly not in electronic format, or to work on editions – well knowing their weaknesses – and consulting mss whenever possible. Moreover, we must constantly bear in mind that text languages are of an inherent heterogeneous nature, due to the absence of a standard and due to the particular transmission of these texts. This implies that it is difficult to describe these languages and in particular it is difficult to test non-trivial hypotheses concerning these languages. There is simply too much noise! If you just look for an isolated form or construction, you can find almost everything. It is very important to have accurate documentation and accurate description for each type of variation i.e. for each diasystem.

What I propose is to combine micro- and macro investigations and in particular to draw upon the results of modern corpus linguistics and variational linguistics concerning e.g. medium, register, text type, social variation, and other sources of variation, when we test hypothesis concerning language change. This implies that we should propose hypotheses for e.g. language change in such a way that they can be appropriately tested on a well defined sub-corpus, in order to avoid irrelevant variation. The difficult task, then, is of course to identify the relevant parameters for testing – but this is a task common to all branches of science.

References

- Schösler, L. “Historical corpora. Problems and methods”. *Linguistica computazionale XX-XXI, Digital technology and philological disciplines* pp. 455–472. Andrea Bozzi, Laura Cignoni, Jean-Louis Lebrave (eds.). Istituti editoriale e poligrafici internazionali, Pisa, Roma 2004.
- Schösler, L. “Scribal variations: When are they genealogically relevant – and when are they to be considered as instances of ‘mouvance?’” *Studies in Stemmatology 2* pp. 207–226. Pieter van Reenen, August den Hollander & Margot van Mulken, with the assistance of Annelies Roeleveld (eds.). John Benjamins Publishing Company, Amsterdam/Philadelphia 2004.