WP4 report (Stuttgart, 1-2 July 2008)

The integration of the watermark repertories on the web in the frame of "Bernstein" will soon be accomplished: the long-lasting dream of those who need to date a graphic document will come true: they will be able to examine the watermarks for which they are looking inside an unique set, and find the dated specimen which is the most similar to the one they have right in front of them.

These new possibilities that we obtained thanks to "Bernstein" constitute a fundamental and decisive step forward in comparison to the previous situation, even though it is clear that all the problems won't be resolved inside the frame of the project: the classification paths of the watermarks lead to watermark groups which aren't always small enough; and the methods we could propose in order to make them more effective ("component model", "landmarks model") imply conceptual problems and above all they are also too expensive: all the watermarks should be previously described, and at the moment it is impossible to do so.

However, the research on watermarks and on the sieve graved by the mould on the paper sheets should not be seen just as an auxiliary technique for historical disciplines. Paper has got a history too; a history which deserves to be studied, because the production and the diffusion of this material is tightly linked to the development of written culture and memory, both at the technical and at the quality level. At the technical level, we can observe that because the production, the productivity and the range of paper typologies had to adapt progressively to the growth of the demand and to new types of paper use. At the quality level, the ability to produce best quality paper at the lowest cost was decisive in the frame of the competition among European printers. In this context, it is important to know when, where and why new paper mills where created; which was their commercial importance; where the paper they produced was used; which was the origin of the paper used in a certain town, or chancellery, or typography; which are the privileged commercial paths and which are the most important centers of exportation. In a more or less direct and explicit manner, Bernstein constitutes a large-scale historical challenge, which deserves to receive the same attention as matters such as the dating of manuscripts or of single editions.

Since the LAMOP is a laboratory which deals with the history of the Middle Ages, it consequently focuses its work on the historical part of "Bernstein". This work is a part of a larger range of activities, but I will shortly report here only the work which really concerns the project, which means both paper itself and the biggest paper consumer at the end of the Middle Ages: incunabula.

The available data permit us to make monographic research: for example, the complete analysis of the watermarks helps us following the production of a paper mill year-by-year; but the most interesting developments take place in the field of quantitative history. It is very simple to enumerate the ingredients we need to make quantitative history; but it is difficult to implement them. Actually, we only need a very complete set of homogeneous and previously classified data; a means which permits us to describe and analyze their statistical characteristics; another means –the GIS- which enables us to visualize the spreading of the phenomena in the space and their evolution across the time.

So, we tried to lay the necessary foundations for reaching this aim.

The GIS

The system of georeferences

First of all, I will spend some words on the system of georeferences which constitutes the very base of the GIS, and in particular on the historical problems brought up by the acquisition process of the georeferences. At the beginning, we had decided that the GIS should be linked to the following sources: Piccard Online, the printed Piccard catalogue, Briquet, WZMA, WILC, IBPH. Recently, thanks to Alois Haidinger's work, we have been able to add the places mentioned in Lykhacev's repertory. However, since the GIS should be able to work on other sets of data concerning written culture, we added those coming from the incunabula repertories (printing places, localization of the copies).

During the preparation phase of the GIS, we faced enormous problems of identification and disambiguation, due to the age of the two main watermark repertories -Piccard and Briquet- and to the fact that their authors didn't identify rigorously the places where paper was used. So, much paper came from places which don't exist anymore, like castles in ruin; many other places changed their name, like all the Prussian towns and villages which belong to Poland and the Czech Republic since the end of the Second World War; and so on.

Furthermore, there are in Europe a huge number of synonyms among the denominations of towns and villages. To disambiguate them, it is necessary to find information which come directly from the source. Concerning paper, these refer to places and collections where the document has been conserved and, for letters, to the identity of the author. But this type of information is not always available and even when it is available, it should undergo a careful examination in order to avoid huge mistakes. A deep knowledge of European geography and history would be ideal for this aim, but it is really difficult to obtain. All these difficulties have made the task quite hard to accomplish, together with the idea that there's a big difference between creating a database for one's personal use, where a reasonable error level can be accepted, and creating a database for general use, where no error can be tolerated.

So which are the results? If we take into consideration all the sources, at the beginning we had a list of 13 000 denominations, which actually correspond to a little more than 7 000 different places. Globally, 87% of all these places could be completely identified or disambiguated. The sources have very different failure rates. The average rate (13%) could seem to be high, but in fact its impact is quite smaller, because the less known places are also those where less paper was used. For example, if we calculate the failure rate (not identified+ not disambiguated) for all the records of POL, we obtain as a result less than 3%. This rate is almost equal to zero for WILC and GW. Clearly, these results concern exclusively the sources which have been examined up to now. Any georeference coming from a certain source can only be used if it has previously been controlled and recognized.

2 Preparation of the data which will be processed by the GIS

First of all it is important to make clear that the GIS shall be able to work at the same time on different databases, whose data are more or less directly linked to the history of written culture. For this reason, we are working in the frame of another project on the creation of a database concerning only incunabula; and in the frame of "Bernstein" we try to implement a database of general biographical notes for authors of books printed before 1501; this database will be completed by the end of the project. Finally, we have already completed a chrono-geographical database of pest epidemics from 1347 to 1600, which contains almost 6 000 records. It is actually known that pest epidemics had a huge impact on book production, and surely on paper production too.

The integration of paper and incunabula

Since many of the watermarks which we can find in repertories come from printed books, and especially from incunabula, one of the "natural" aims of "Bernstein" has been the linkage between the watermark databases and the incunabula repertories. For this purpose, KB and ISTC have created a "crosslink" between every single watermark and the corresponding image in ISTC. In the same perspective, we went through the Heitz repertory of watermarks coming from incunabula from Strasbourg, and we added the reference to the ISTC number of the edition. This has been possible for 114 identified editions. Actually, POL contains more than 4 000 watermarks coming from incunabula. We were able to distribute them among 568 editions that we have identified and we have added the correspondent references. It is also foreseen to add the ISTC references for the Briquet repertory, following the list that Allan Stevenson made in 1968. Except from the Spanish incunabula whose watermarks have been recognized by Van Thienen but haven't been described yet, about 2 500 editions will soon be referenced in our watermark databases. We should also remember that these are only links: nothing has been foreseen yet for a common analysis of the paper and the books' characteristics, which would presuppose the possibility of making common queries inside the two types of databases.

The GIS, like any statistical software, works only using standardized data, which have been classified according to characteristics which correspond to specific fields and where every single field contains a series of mutually excluding values. This implies that inside a certain database, every single value has absolutely to be coded in same way. The situation becomes more complicated when different databases are involved in a common search, because in that case the integration has to be performed at the infrastructural level: not only the fields have to be the same, but also the coding of the values.

It is relatively easy to respect these conditions working with recent databases ; even if they are not always conceived using the same criteria, they nevertheless follow rigorous ones. The situation is very different if we consider old repertories –those by Briquet and Piccard- which have been created day-by-day without establishing previously strict criteria, at a time in which informatics

didn’t impose an absolute coherence from the beginning. Afterwards, these repertories have been digitalized exactly as they were on the cards or on the printed version; it has to be said that re-working them would have hardly been possible, because of the need for the presence of a paper historian, and the task couldn’t have been accomplished in time.

Since the people responsible for POL have been so kind as to give to the LAMOP the complete text of the database, we could successfully accomplish the unification and classification work. The details will be discussed tomorrow morning, when we will give a brief general presentation. We can foresee to gather all the data now available (and others which could be added in the future) in an unique “superbase” concerning quantitative history of paper. I’d like to avoid any fear and misunderstanding: we don’t want to make the existing databases obsolete. On one hand, the standardization work could be successfully used inside the existing fields; on the other hand, the aim and the way this new “superbase” would work are completely different from those of the repertories which are now online, whose aim consists in finding a watermark by choosing it inside a series of images with a short description. On the contrary, the “superbase” is not adapted for this kind of searches, which will always remain the principal one, because it has been worked out in order to follow the synchronic and diachronic evolution of macroscopic phenomena concerning whole populations of paper sheets; it hasn’t got the aim of individualizing, but of gathering. So, there isn’t any concurrence: the two types of database are complementary.