Digital Edition and Graphemic Analysis of the Ebla Texts

The commitment to digitize the texts of the Ebla archive was made very early, during the first meeting, in 1977, of the International Committee for the Publication of the Texts of Ebla, just two years after the discovery of the archive. That was my specific task within the Committee, and the goal was to provide the colleagues with a computerized data base that could be used as an in-house resource. This was in line with a similar research project I had undertaken in 1968 with regard to Old Babylonian letters, and in working on both projects it became clear that graphemic analysis was an indispensable component of this research. On the one hand, it was necessary to develop coding and tagging procedures that would allow for a more coherent definition of the data than was possible with the standard transliteration system (the numerals being a case in point). On the other, it became immediately clear how desirable it was to exploit fully the unsuspected potential of the system as a way to develop a more clearly articulated approach to graphemics analysis as such.

Proceeding along these lines, I supervised, in the first place, the data entry of the texts. But in those years the technical resources were still at a very primordial level. Online exchange of data was of course unthinkable. But it was not even possible to exchange data electronically, since we depended entirely on mainframe computers, accessible through operators, and with electronic data storage being possible only on large tape reels. Thus the only means of exchanging data was through bulky printouts on large fanfold paper. We produced a number of KWIC concordances (Key Words In Context), with indices, which resulted in thousands of pages. The interest in these early phases of the project is purely historical, and the point is not to be elaborated here. But it is worth mentioning because it explains why the results of our work did not come to full public fruition in those early stages of the project.

But there was an important conceptual corollary that emerged from this early work, and this was, as indicated, a clarification of the properly graphemic implications of the enterprise. The complexity of the cuneiform writing system required much thought as to what graphemes and graphemic rules actually are. This is not to say that the assyriological tradition had not already given considerable thought to this. In fact, the standard transliteration system may be considered a jewel of intuitive understanding of the essence of the problem, well before a theory could be articulated to account for its principles. But focusing on the need for maximum transparency in accounting, even more systematically, for a rigorous graphemic categorization led me to focus more specifically on the underlying theory, in an attempt to formulate the basic principles and show their relevance for the practical task in front of us. This resulted in a long article devoted specifically to Ebla[1], and in another article[2] and a volume in a new series I started about this specific subject,[3] where not only was the theory developed, but also extensive use was made of the data newly gathered in the format of choice.

There was also another side to graphemic analysis. The amount of data in our computerized corpus was so vast, and so immediate was the readiness with which different sorts would bring out significant

1 "Studies in Ebla Graphemics, 1," *Studi eblaiti* pp. 5 (1982) pp. 39-74

2 "Comparative Graphemic Analysis of Old Babylonian and Western Akkadian," *Schaeffer Festschrift, Ugarit-Forschungen* 11, Neukirchen, (1979) pp. 89-100

3 "Introduction," in C. Saporetti, *The Middle Assyrian Laws,* "Graphemic Categorization" Vol. 2. Malibu: Undena Publications (1984), pp. 1-20

distributional patterns, that a new sensitivity began to develop in the face of quantification. There were no practical limits to the amount of data we could control, unmatched by any card catalog of the pre-digital era, and made possible not only by the new technical medium but by the inherent methodology of graphemic analysis. We could potentially include the total amount of known texts, and this meant that statements of non-occurrence acquired a much greater weight than ever before, not only with regard to individual signs, but with regard as well to their correlations. We were closer than ever imagined to the ideal of having access to the competence of a proficient ancient scribe, and, through that, to the competence of a living informant for what concerned the language. But it was precisely a proper graphemic understanding that could serve as a bridge between the mute evidence of the written text to the phonemic reality of the language. From a visible, as it were, to a voiced language. It appears then that digital editing of the texts and graphemic analysis went hand in hand from the beginning, by necessity. I plan to come back to this aspect of the research with a follow-up to my 1982 article.

In this light, it will be clear why the initial publication of the data was subsumed under the title "Graphemic Categorization" which I gave to the series in which the various corpora were to appear. The first corpus appeared as one of the very first electronic publications in our field, as a floppy disk.[4] The second was devoted to the Middle Assyrian Laws.[5] The Ebla corpus that appears here is the third. In each case, what was envisaged was more than a concordance as a tool of analysis: it was also a study in the graphemic dimension of the corpus as a window onto the corresponding phonemics.

The intellectual enthusiasm of merging, in such a way, technique and method colored the whole enterprise, and in that I was followed by a group of dedicated students, who applied these principles to specific corpora, resulting in the following doctoral dissertations devoted to Ebla: James H Platt, *Eblaite scribal tendencies: graphemics and orthography* (1993); Joseph M. Pagan, *Morphological and lexical study of personal names in the Ebla texts* (1994)*;* and Terrence L. Szink, *Computer-aided analysis of the Semitic of the Ebla tablets* (2005).[6] Pagan's dissertation was published in 1998 under the same title as *ARES* 3. Szink's dissertation served as the springboard from which the present work has derived. All three (Platt, Pagan and Szink) contributed immeasurably to the establishment of the corpus, especially in those early years when not only the procedures were very cumbersome, but the mentality in general was not particularly conducive to working with computers. The credit I can give them here hardly reflects the time and dedication they devoted to the project, for which I am most grateful.

The role of Lucio Milano was indispensable in bringing this project to a successful issue, from his early stay in Los Angeles, through the entire process of revising the input, and all the way up to securing the means for this publication. A visit he arranged for me in Venice was also very important in that it gave us the opportunity to compare notes in the latter part of the project. Milano also introduced to the project his students in Venice who completed the work that had been started by my students in Los Angeles.

From the start, Paolo Matthiae set store by the idea of a digital edition, long before the term and the reality behind it had become a normal dimension of scholarly life. Since joining the Committee, Alfonso Archi has been just as supportive, in particular with regard to colaltions against the originals. For their interest, support and friendship I am most grateful.

---

4 G. Buccellati, A. H. Podany and O. Rouault, *Terqa Data Base, 1: OB Texts through the Fourth Season* ,Malibu: Undana Publications, 1987. A matching paper volume was never published. The coding of the texts follows the same format as the one used for the Ebla texts. The data are now available online in the Terqa website under TEXTS > FLOPPY DISK, where one will also find a discussion of this early stage of the research.

5 By C. Saporetti, see above, note 3. In this case, only the paper volume was published, but not the electronic version.

6 Other corpora were used for the following doctoral dissertations: Y. Kobayashi, *Graphemic analysis of old Babylonian letters from south Babylonia* (1975); P. Gaebelein, *Graphemic analysis of Old Babylonian letters from Mari* (1976); T. J. Finley, *Word order in the clause structure of Syrian Akkadian* (1979); J. L. Hayes *Dialectical variation in the syntax of coordination and subordination in Western Akkadian of the el-Amarna period* (1984).

The Directorate General of Antiquities and Museums of Syria has always been the ideal intellectual host for archaeology in general, and for our project in particular. In this regard, I would like to remember in a special way those who were with us at the beginning and were so instrumental in guaranteeing a solid footing for the project , Dr. Afif Behnassi, the lamented Dr, Adnan Bounni, and Kassem Touer.

The preparation of the digital edition and our work onj graphemic analysis was made possible over the years through grants primarily from the National Endowment for the Humanities at the beginning and from the Council on Research of the UCLA Academic Senate at several junctures during the course of our work. It is a special pleasure to be able to acknowledge her their enlightened support, especially at the beginning and then over the long period of time that has intervened.

Giorgio Buccellati