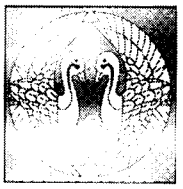
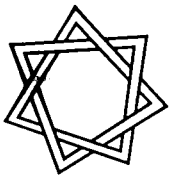
RUSSIAN ACADEMY OF SCIENCES
THE INSTITUTE OF ORIENTAL STUDIES
ST. PETERSBURG BRANCH



Manuscripta Orientalia

International Journal for Oriental Manuscript Research

Vol. 4 No. 4 December 1998



TSESA
St. Petersburg-Helsinki

ORIENTAL MANUSCRIPTS AND NEW INFORMATION TECHNOLOGIES

V. A. Jakobson

COMPUTER ASSYRIOLOGY

1

Assyriology is the branch of Orientalistic which studies the history, languages, and cultures of those ancient peoples which employed cuneiform writing. A specific feature of Assyriology is that it deals with a number of languages not related to one another (with the exception of Urartic and Hurritic, which belong to the same language family). Cuneiform was invented by the Sumerians at the beginning of the third millennium B.C. It was later borrowed from them by neighbouring peoples, either directly or not, who adapted it to their own languages. It is an extremely complex writing system employing several hundred signs, each of which can have several (sometimes dozens) syllabic (phonetic) and up to five or six semantic (ideographic) meanings. It is not difficult to calculate the number of variant readings, which can arise from a combination of only three or four cuneiform signs. The matter is further complicated by the absence of word divisions and punctuation marks in cuneiform — all signs were written one after the other without any spaces, divided only into lines. Fortunately, certain orthographic rules and grammatical features somewhat facilitate cuneiform reading.

Each cuneiform sign represents a combination of several wedge-shaped marks, which are vertical, horizontal or inclined. These marks were impressed by the scribe with a three-sided stick on a tablet of fresh clay which was then preserved by drying or, less frequently, by baking, as with ceramics. On rare occasion, inscriptions were made on other clay objects as well as on metal and stone. In the latter cases, they invariably imitated inscriptions on clay. At present, the world's museums possess approximately half a million cuneiform texts of varying lengths, ranging from a few signs to thousands of lines. The number of these texts continues to grow.

2

Cuneiform texts are published to ensure specialists access to them. The simplest and least labour-consuming method of publishing such kind of texts would be to photograph them. For a number of reasons, however, readable photographs of cuneiform texts are either difficult or impossible to produce. The holographic method is extremely promising, but only the first steps have been taken in this direction. Therefore, cuneiform texts are currently published in the form of so-called drawings, copies made by hand on paper and then reproduced polygraphically. It is a labour-consuming process which requires highly qualified specialists. Ideally, drawings accurately reproduce the text and all of its individual features: the relative size of the signs, the handwriting of the ancient scribe, damaged spots, etc. Transliteration is employed for direct work with the texts: the text is rewritten in Latin letters and equipped with necessary diacritics. Transliteration is, in a sense, a "translation" of cuneiform into another, alphabetic, writing system. Each cuneiform sign is designated by a combination of Latin letters which convey either the most widely used syllabic (phonetic) meaning of a given sign or its most important ideographic (semantic) meaning in Sumerian. Such a transliteration is performed in strict accordance with a well-defined set of rules (algorithms) and should be fully reversible, allowing one to recreate the original cuneiform text, if not its outward appearance. This is ensured by taking into account all possible variant readings which can arise as a result of the polyphonetic and polysemantic nature of each sign, as noted above (section 1).

Thus, texts can be published in the following fashions: a) drawings, b) transliteration, c) drawing in conjunction with transliteration. Texts are then cited in transliterations or transcriptions which convey the actual pronunciation of each word in the corresponding ancient language. Only in rare cases, when handwriting or damage cause doubt about a particular reading, is a passage reproduced in drawing form. The final stage of work on a text is a translation, which is as complete as possible (given the condition of the text) and equipped with all necessary commentary. For this reason, texts are ideally published as drawings with transliteration and translation, introducing them into scholarly circulation, after which the texts can be employed for scholarly problems whatever they may be.

3

Each scholar studies those aspects of texts which interest him. He attempts to extract all relevant information from a concrete group of texts selected on a chronological, geographic, linguistic, thematic, etc. basis. To this aim, each Assyriologist draws up a substantial file or files and then augments or restructures them throughout his scholarly career. Such files, drawn up to answer specific questions, inevitably contain an enormous amount of valuable information which cannot be used within a single or even several works. Unfortunately, at the end of their creators' careers, working files are commonly of no practical value: each scholar has his own system of classification and headings, his own system of abbreviations and references, and, finally, his own (too frequently illegible) handwriting. In order to make use of such a file, another scholar would be compelled to expend a great deal of time and effort without any guarantee that they will pay off. Sometimes, it is much simpler to work through the texts anew. To cite an example, the Research group of ancient Oriental philology at the St. Petersburg Branch of the Institute of Oriental Studies holds a vast file of the late Prof. A. P. Riftin (1900—1945), an outstanding Assyriologist. His numerous works were based on enormous information extracted from cuneiform texts. To great regret, it is impossible to be used. And each Assyriologist cannot help but mourn the fact that his own card-indices, which cost him such effort and contain so much information that he was not able to use, will one day be waste paper. Up-to-date computers, with their large memories and high processing speed, allow us to contemplate the possibility of creating was a computerised Assyriological file, rather, a set of interconnected thematic files, copies, transliterations, and translations of all extant texts and archeological materials, which can be a general Assyriological database accessible to all.

Of course, the creation of such a generalised database remains a task for the future. We must start with individual databases drawn up and organised by specific and uniform rules which permit them to be united in the future. The rules should be flexible enough so that changes and improvements can be made if necessary, including retroactive adjustments capable to alter the entire database.

The invention of writing made it possible for mankind to create a collective memory, imperfect and susceptible to damage though. This memory contains only the results of scholarly work. Generally, the preparatory stages of such work (the "rough drafts") are shadowed, albeit they are frequently capable to grant information, the importance and value of which was not always evident for the scholars themselves. Most often such information is lost forever; only in rare cases is it rediscovered, with inevitable losses though. The creation of a general computerised database, or at least a group of individual databases, eliminates this difficulty, making it possible to preserve the totality of valuable information and intriguing process of scholarly thought.

4

At the same time, to create solely databases would mean to use the computer merely as a typewriter or powerful calculator. The capacities of up-to-date computers permit not only to perform logical operations, but to reveal also hidden information, as well as to facilitate making new conclusions on the basis of available information.

Just to cite an example, cuneiform texts such as letters, administrative, commercial, economic and juridical documents mention various individuals. Such references frequently indicate their official positions and genealogical ties. They name also, directly or indirectly, their place of residence, contain information about their material well-being, commercial deals, legal matters, administrative directives relating to them, etc. It is possible to draw up, on the basis of known texts, an individual "dossier" on each of the individuals mentioned therein. It demands first of all identifying these individuals with reasonable accuracy as well as separating namesakes. Although the number of possible names in ancient Mesopotamia was quite large, each historical epoch had several dozen extremely popular names. The figures found in cuneiform documents are usually identified on the basis of name and patronymic, but this is a) not always true and b) people with the same names and patronymics occur, too. To obtain accurate "individualisation", one must turn to secondary factors, such as age, official position, address, names of wife and children, names of relatives, etc. A computer can perform such identifications quickly and accurately, taking all information into account and pinpointing problematic cases.

Especially important and interesting would be the compilation, where possible, of genealogies encompassing two, three, or more generations on the basis of ties enumerated in the texts. This would permit the subsequent creation of "family dossiers" which account for the material well-being of a given family or clan, changes, the inheritance or non-inheritance of official positions, individual careers, moves, life spans, and a great deal of other information. Obtaining of such genealogical information would be surely a success, the data collected enable us to have more vivid picture of real life in the ancient East. A computer can significantly speed this work and remove the possibility of errors stemming from distraction, oversight, and misunderstanding.

Certainly, the study of large texts presents a much more complicated case. Here one deals with the necessity of revealing and analysing their logical structure, extant variants, and individual characteristics (texts frequently exist in several copies, which differ from one another). The existence of overlapping copies allows one to recover damage-induced lacunae, sometimes in part, and sometimes, with a bit of luck, in full. No less important is the lexical analysis of such texts, and especially the terminology they contain. Finally, it is possible to discover links between certain texts in the form of citations, paraphrases, and direct references.

5

Making of data-entry programs is now a quite routine procedure. Entering drawings (copies) presents no problems too and can be done with a high-resolution scanner. As for entering transliterations and commentaries, it is a fairly complicated procedure. The difficulty is that a number of alphabets — Cyrillic, Latin (with diacritics both above and below the line), and, preferably, Greek and "square" Hebrew — are to be utilized. Though programs for each individual alphabet exist, using them all in a single text presents certain difficulties. These become even greater, if one has to deal simultaneously with three of the alphabets, which are written from left to right, and Hebrew, which is, like Arabic, written from right to left. The problem of ensuring quick and simple toggling between different writing systems in all of their variations has not yet been resolved.

Unfortunately, programs for textual analysis do not yet exist; their creation is the primary aim of the project under discussion here. Here, we will have to advance step by step, proceeding from simple to more complex tasks and bearing in mind that each step forward can reveal new perspectives, which were previously inconceivable or dimly imagined. In this process, as was noted above, it would be desirable to be able to change and augment our programs as we work, and to do this in such a way that changes and additions affect already collected data.

6

It seems reasonable here to give an example of a relatively standard legal text. The text is representative and gives one a sense of the minimum amount of data which must be entered into the computer, as well as a sense of the data which the computer can systematize itself and then produce at user request.

a) Each text has its own "passport" which should indicate the following information:

Table

YOS 8, 120

Rim-Sin 40, 20/IX

Adoption	
(1) *^{m}Ì-lí-gi-im-la-an-ni muni*	**(1)** (A person) by the name of Ili-gimlanni
ki ní-te-na	from himself
^{m}Ši-ip-^{d}Sin	Shep-Sin
nam dumu-ni-šè šu-ba-an-ti	adopted.
(5) *nam-ibila-ni-šè in-gar*	**(5)** He made him his heir.
u$_4$-kúr-šè u$_4$-nu-me-a-ak	In the future, at whatever time,
tukun-bi	if
^{m}Ì-lí-gi-im-la-an-ni	Ili-gimlanni
nam Ši-ip-^{d}Sin ad-da-ni	to Shep-Sin
(10) *ad-da-mu nu-me-en*	**(10)** "you are not my father"
ba-na-an-dug$_4$	says,
ku-šè ba-an-šum-mu-uš	they will sell him for silver,
tukun-bi	(but) if
Ši-ip-^{d}Sin	Shep-Sin
(15) *nam I-li-gi-im-la-an-n[i]*	**(15)** to Ili-gimlanni
dumu-mu nu-me-en ba-na-an-dug$_4$	says "you are not my son,"
é-nig-ga-ra bar-ra-éd-a	he will lose house and property.
mu ^{d}Nannar ^{d}Šamaš	In the name of the god Nannar, the god Shamash
ù ^{d}Ri-im-^{d}Sin lugal	and Rim-Sin the King
(20) *in-pád-meš*	**(20)** they swore [to this].

List of witnesses, date, seal of the first witness.

— name of the edition in full or in generally accepted abbreviation and the number of the text according to this edition (line one, heading);

— museum number of the text (here omitted);

— date of the text in Mesopotamian system of chronology (year in the reign of this or that ruler in accepted abbreviation, date and month according to the Mesopotamian calendar);

— short description of content (standardised).

It is particularly necessary to obtain the following information from the computer:

— all texts of a particular origin, relating to a particular reign or period within that reign;

— all texts stored in a particular museum;

— all texts which mention certain individuals in one capacity or another.

One must be able to obtain from the computer the text in any form, as well as scholarly commentaries on the text. Finally, one must be able to obtain from the database all texts which contain a specific term or word (ideally, even a cuneiform sign or selected group of signs).

b) When a new text is entered, the lists of names and genealogy should be automatically updated. Each name should be linked to the following data: genealogical ties, official positions, place of residence, passports of all texts in which he/she is mentioned and the capacity in which he/she is mentioned (seller, buyer, creditor, debtor, official, judge, witness, slave, lord, etc.). Terms or selected words contained in the text should be entered into lists of terms with an indication of passports for the corresponding texts. One must be able to obtain from the database all of the information enumerated above in any configuration.

7

Since up-to-date computers are not yet able to work with natural language, one must develop a language for the formal description of texts. It is clear that each word in such a language will represent a certain concept, that is, the result of an initial interpretation not only of a given text but of an entire group of texts relevant to a certain question. The entirety of such formal descriptions (formal texts) will form a conceptual memory on the basis of which logical operations are performed. The words (concepts) of this formal description are, as a rule, interconnected; this interconnectedness, in conjunction with the semantic volume or level of each concept, assumes the form of a "concept tree" or "matryoshka" of concepts. Here are several examples (see *figs. 1, 2, 3*).

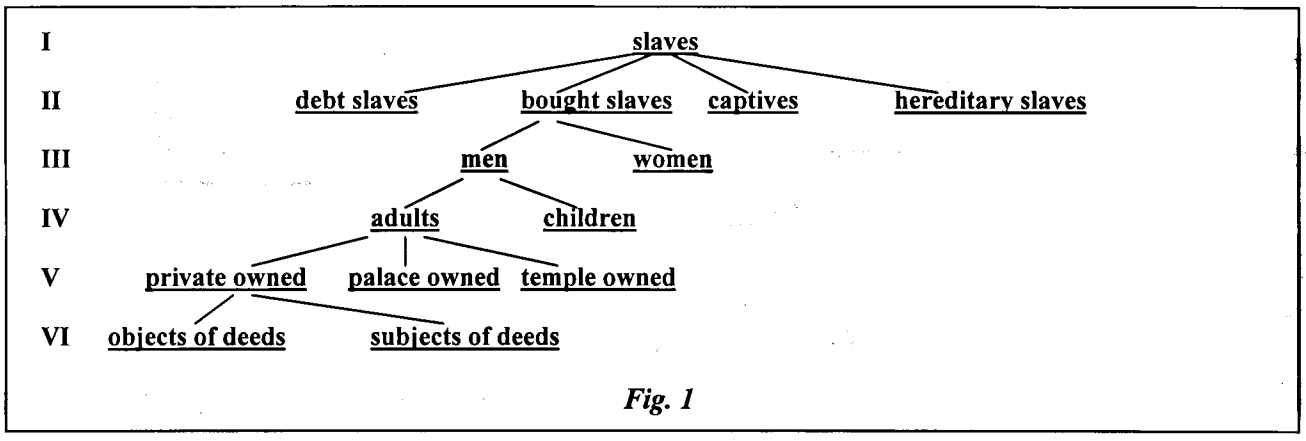


Fig. 1

Obviously, level III—VI concepts and higher are related to level I—II concepts. We see from *fig. 2* that, first, trees can "grow together" (the concept of "slaves" recurs) and, second, that the levels are not equivalent. Thus, if levels II—IV concretise the object of a transaction, levels V—VII do not relate to all concepts present in levels II—IV — the tree can by asymmetrical. In this fashion, nearly all concepts are in one form or another interrelated. The "trees" or *matryoshka*s are themselves concepts, and for this reason they can be reviewed (restructured). The program must make this possible.

It is not accidental that the examples cited here relate to legal texts: such texts have a clearly defined formal structure and are characterized by a fairly limited set of terms and usages. At the other end of the spectrum, literary texts present much greater complexities. All other cuneiform texts lie somewhere between these two poles.

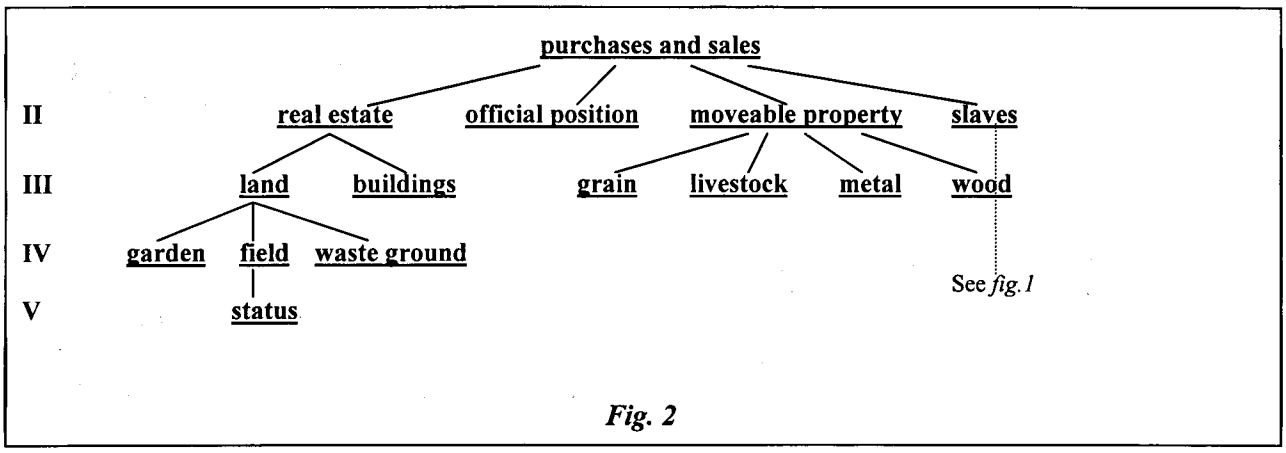


Fig. 2

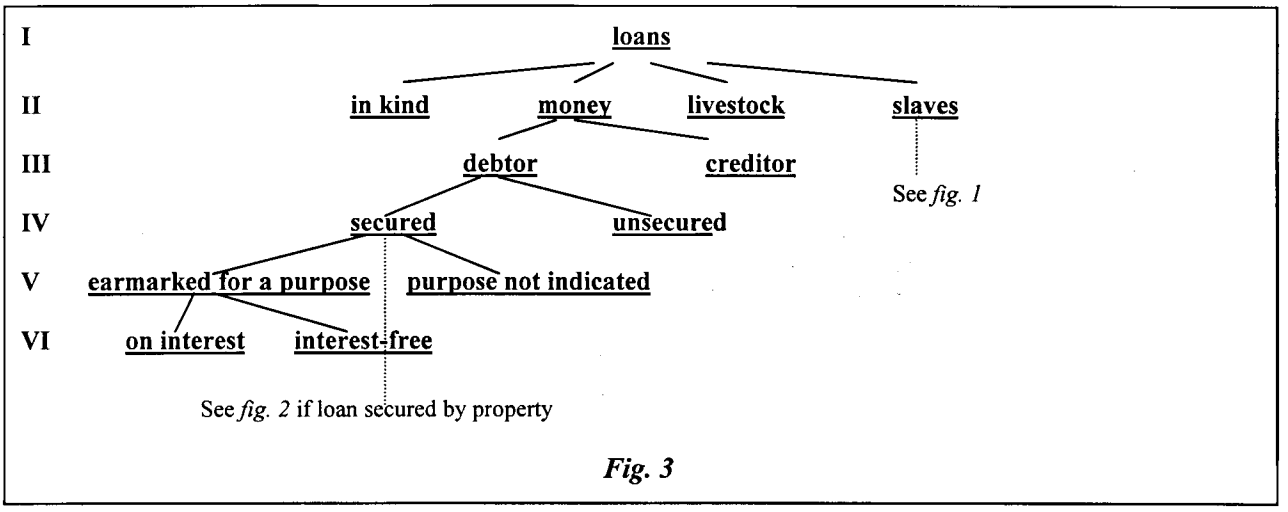


Fig. 3

8

In our view, one of the main problems is the amount of work required to develop formal concepts and enter them into the database. Obviously, this must not significantly exceed the amount of work needed to create an ordinary file, or no one will undertake the task.

The foregoing concerns the formal analysis of texts and is evidently realisable in the near future. As for the analysis of their contents, we remain at the level of feeling our way toward new approaches.