Achtung!
Dies ist eine Internet-Sonderausgabe des Aufsatzes
„The Use of Computers in Ancient Near Eastern Studies [Abbreviated Version]“
von Jost Gippert (1996).
Sie sollte nicht zitiert werden. Zitate sind der Originalausgabe in
Akkadica 99-100, 1996, 45-52
zu entnehmen.

Attention!
This is a special internet edition of the article
“The Use of Computers in Ancient Near Eastern Studies [Abbreviated Version]”
by Jost Gippert (1996).
It should not be quoted as such. For quotations, please refer to the original edition in
Akkadica 99-100, 1996, 45-52.

The Use of Computers in Ancient Near Eastern Studies State of the Art and Future Prospects[1]

Jost Gippert, Frankfurt

When I was invited by the organizers of the XLIII. RAI to read an introductory paper about the use of computers in the field of Ancient Near Eastern Studies, I gladly accepted although I am not an expert in the field. The reason was that there are quite a lot of overlappings between the topics I was expected to deal about on this occasion and the projects I have been involved in within the area of Indo-European Studies during the past ten years.

If I had had to talk about the use of computers in a philological subject like Ancient Near Eastern Studies some ten years ago, I should mainly have talked about the problems we all had to face at that time, namely problems concerning the integration of diacritics and other special transcriptional signs into our manuscripts of books and articles. Of course there exists no final solution of this question yet insofar as there is still no universally adopted international encoding standard for letters such as dotted *s* (*ṡ*) or *h* with a half-circle below (*ḫ*), and even the so-called Unicode, which is sometimes regarded as the basic encoding system for the next generations of all computers, does not contain them. But the creation of the necessary characters has become a minor problem in recent years because font developing has turned into an easily available feature of all systems meanwhile, and camera-ready manuscripts containing the special transcriptional characters can be produced comfortably with nearly every text processor nowadays.

Nevertheless there has been a development recently which makes reconsideration of the problem of encoding original language material worth while again. Eversince internet access has been establishing itself more and more as a means not only of exchanging mailings but also as a basis of storing and retrieving informations and data material, we have been facing the problem of representing "special characters" on the screen and on printers again. It is not only in my personal view that the internet is likely to develop itself into a primary tool of publication, offering more comfort and more prospects than any other means of publication did before, and so we should take this question seriously. This is why I decided to concentrate my paper on this topic.

It is indeed a remarkable fact that in a field as "exotic" as the study of ancient Near Eastern languages, more sites seem to have been established where data can be retrieved from via the internet than for some much more widely spread subjects such as Modern Italian or Polish studies. Maybe, this is only my personal impression, caused by the fact that I have hardly time enough to just "browse" through the "World Wide Web", "surfing"

---

[1] This is an abbreviated, text-only version of my paper read at the XLIII RAI in Prague. The full text will be printed in Studia Mesopotamica, Anatolica et Iranica [SIMA] 3, 1996.

in search for sites that do not interest me very much; and it is true that today, only some three years after its creation, the "WWW" has become so "overcrowded" that it is not easy any longer to find a way through its labyrinth. In the field of ancient Near Eastern studies, we are in an excellent situation, however: Looking for web sites relating to a certain topic, we would normally have to rely upon the results that can be achieved by using one of the so called web searchers, that is sites named Lycos, Yahoo, Altavista, Inktomi, Opentext and so on, where pages published in the net are indexed according to key words they contain. Doing this e.g. for Sumerian, we would indeed be provided with a list of between six and more than 200 addresses of pages that concern topics relating to this language, its mythology or other matters Sumerian. This procedure has a disadvantage, however, in that the data we can retrieve from web searchers is not arranged in any kind of order with regard to contents, and you can never be sure to what extent the lists are complete. So we have to be extremely grateful that for all kinds of cuneiform studies and related subjects, we are offered a special information board on the server of the Chicago ABZU project.

The ABZU server, organized by Charles E. JONES, not only provides a description of the project itself as presented on the homepage, but it contains a collection of relevant net adresses, arranged in several indexes. By today, there exist an author index, a regional index for Egypt and Mesopotamia, a subject index for archaeological sites, a project index and some others.

It may be helpful here to give some explanations for all readers who are not familiar with the internet, of how to use a WWW page. If you are equipped with a net access and a so called net browser such as Netscape (this has developed as the market leader in recent days although there is no market — web browsers are usually freeware), you can access any site entering its unique account name as its "location"; the one of the ABZU homepage is http://www-oi.uchicago.edu/OI/DEPT/ABZU/ABZU.HTML. Every item on a page that is marked by underlining can be used as a so-called "link": If you place the mouse cursor on the screen upon such an item and click upon it, the adress it hides will be accessed automatically, no matter whether this is another paragraph of the page you are already reading, another page on the same server, or even a page on a server in a different continent. Starting like this from ABZU’s "table of contents" and choosing the regional index, you will enter into several lists containing explanatory text as well as further references such as the list of "Languages - Philology - Texts - Translations" (http://www-oi.uchicago.edu/OI/DEPT/RA/ABZU/ABZU_REGINDX_MESO.HTML)this, in its turn, leads you to several sites outside Chicago such as the server of the "Center for Advanced Research on Language Acquisition (CARLA)" at the University of Minnesota, where information is available about institutions in North America that "offer instruction in less taught languages" like Akkadian, Old Persian, or Sumerian (http://carla.acad. umn.edu/lctl/languages/Akkadian.html). Such collectional information sites are an interesting new tool for scholarly exchange of their own; but of course their usefulness depends on their entirety, and this depends on what informations people involved in the field are willing to submit. Everybody who makes use of the Chicago ABZU server for retrieving informations of this kind should keep in mind that cooperation is needed in order for it to develop into a first hand source of information.

A much greater advantage for our scholarly work is provided by a different kind of pages though. What I mean is internet pages that contain primary linguistical and graphical data of the languages we deal with. Several attempts have been made in recent years of establishing sites, i.e. servers, in the internet as a storage basis where original data can be retrieved from, thus providing a new type of scholarly publications. One such site is the one the organizers of the XLIII. Rencontre have provided. On their server named ENLIL (http://anes235-1.ff.cuni.cz/home.htm) — they not only publish an electronic "Journal of Arabic and Islamic Studies", an electronic discussion forum about "Computers and Ancient Languages", as well as all kinds of information material as necessary for participants of the RAI conference, but they also provide a "Library of Ethiopian Texts" named "Corpus Textorum Aethiopiae". This may of course be regarded a little bit farther off the scope of ancient Near Eastern studies, but we can in the same context refer to some other servers that have recently been established as a basis for storing and retrieving textual data pertaining to this field of interest. One of these is the archive of "Proto-Cuneiform Texts from Archaic Babylonia" organized by Robert ENGLUND from Berlin in cooperation with Hans J. NISSEN, Peter DAMEROW and the Computer Center of the University of Lüneburg (http://fub46.zedat.fu-berlin.de:8080/~uruk/Welcome.html). Although it is but a prototype that has been made available so far — it contains "the files of ten texts and a sign glossary with just seventeen signs" — it may well be used to demonstrate what should be aimed at when using the fascinating new internet technology for a publication of cuneiform text material. Choosing the link to the "Catalogue" from the lower part of the homepage, you will be offered a list of texts (http://fub46.zedat.fu-berlin.de:8080/~uruk/Catalogue). From here, you can easily jump to the description of any one of the tablets which contains **first** a set of data such as measures and bibliographical information, **then** pictures of the tablets — both photographs and drawings — **and** the presumed readings arranged by lines (e.g., http://fub46.zedat.fu-berlin.de:8080/~uruk/W20/DIR274/NR3DATA). On the other hand, you can choose to enter the "sign glossary" (http://fub46.zedat.fu-berlin.de:8080/~uruk/Zeichenliste). From this list, access is available to several pages where single signs are described, each containing a picture of the character itself and a list of tablets and lines where it is met with (e.g., http://fub46.zedat.fu-berlin.de:8080/~uruk/G1096 for the sign read *bar*). These latter pages can in their turn be used as a link to the description of the tablets again, thus representing what we might call an information circle.

By the way, all the linkage procedures as demonstrated here are a typical feature of the so-called "hypertext" method: It consists in the possibility of linking different elements of what could be regarded as a set of informations one with each other — without depending on a linear text structure[2]. This feature is in my view the most remarkable difference of hypertext publications as against normal printed ones, and it is the great advantage of internet applications of the type discussed here that they can make use of hypertext procedures even without having to care for the elements to be stored on one "physical"

[2] The basics of hypertext in connection with cuneiform studies were discussed in the paper read by Jürgen LORENZ on the RAI ("Keilschrift und Hypertext"; the paper will be printed in SIMA 3). At the same occasion, a Sumerian hypertext project was introduced by Miguel CIVIL from Chicago.

server. As far as the Berlin project is concerned, I am sure that its organizers will be aware that the facilities offered by hypertext have not yet been exhausted in the way referred here; of course, it would be useful not only to have a link from the sign glossary to the description of the tablets containing the sign but also vice versa — from the textual interpretation of the tablet to the sign glossary.

A project quite similar with the one I have just described is the "Edinburgh Ras Shamra project" run by Jeff LLOYD. Entering its home page (http://www.ed.ac.uk/~ugarit/home.htm) we are not only offered access to the main web sites of interest with respect to the City of Edinburgh and Scottish Highland Malt Whisky but you can also visit an increasing number of Ugaritic texts arranged according to their KTU numbers — http://www.ed.ac.uk/~ugarit/texts/ktutexts.htm gives a description of the whole scope of the project which is still far from being completed. Finally, the texts will be provided with photographs, transliterations and translations as is demonstrated in further pages on the server (http://www.ed.ac.uk/~ugarit/texts/ktu1.htm etc.).

Although this server is still developing, it can easily be seen that here, too, one of the main principles of a computerized hypertext arrangement will be prevalent in due time, namely the extensive linkage of textual and graphical data. Of course there is still a great difference of quality between real photographs and the highly compressed digitalized images that you can retrieve via the internet at a reasonable speed. But the advantage of the latter can easily be accounted for if one considers the price of printed publications containing coloured photographs with the costs of an internet "trip", and the loss of quality that is connected with the process of compressing graphical digital data is continuously decreasing. As far as digitizing cuneiform tablets is concerned, the main problem consists in edges being inscribed as well as surfaces; several attempts at finding solutions for this problem are being made at present[3].

As for the Edinburgh Ras Shamra project, it can only be hoped that it will not keep restricted to a few texts from Ugarit as it is now but that one day it will offer a complete set of the material available. In this respect it would be worth while considering a cooperation with the other institutions that deal with Ugaritic as, e.g., the Prague Institute of Ancient Near Eastern Studies, the "Banco de Datos Filológicos Semíticos Noroccidentales" at Madrid or the Ugaritologisches Institut at the University of Münster. All these institutes have established a data base of Ugaritic texts — and curiously enough, these data bases served as a basis for the preparation of three independent Ugaritic word indexes that appeared recently.

One major advantage the internet structure offers as against all kinds of printed text is that even linguistic tools such as a word analyzer can be made available here; a good example of this is the server of the Perseus project established at Tufts University which aims at providing all kinds of tools for dealing with Ancient Greek (http://www.perseus.tufts.edu). Within the Madrid project named above, a similar tool for an automatical analysis of Ugaritic word forms has been developed; according to its authors, this too will

[3] Within the RAI, Karel VAN LERBERGHE from Leuven talked about new procedures developed in this respect; furtheron, a digitalizing equipment for this purpose was demonstrated by a joint venture of a Czech and a German company.

be accessible via the internet soon[4]. There are two things one has to bear in mind when making such tools available in the net, however: First, the tool must not be restricted to single systems such as UNIX, Windows, or Macintosh if it is meant to be usable for the whole scholarly world. Of course, programming languages for the net are only just developing — maybe JAVA as provided by Netscape will turn into a sufficient basis for this purpose some time. But I do hope that our colleagues from Madrid will take this problem into account so that their data base will no longer be confined to Macintosh users.

The other problem that has to be kept in mind if one wants to provide analysis tools for ancient Near Eastern languages such as Ugaritic is the unavailability of system independent fonts for representing all characters of the languages involved, within net applications. The effects of this problem can easily be demonstrated by looking at the sample text from the Edinburgh Ras Shamra project (http://www.ed.ac.uk/~ugarit/texts/rs000024/00000326/welcome.htm). Here you will see that, e.g., the character *ayn* is replaced by the roof-shaped circumflex mark ("caret", ^) so that the text looks a little bit strange if compared with the usual transcription. The reason for this substitution lies in the fact I have mentioned above, namely that several letters such as *ayn* (ʿ), dotted *s* (*ṡ*) or *h* with a half circle below (*ḫ*), do not form part of the normal encoding system which is used in the World Wide Web, viz. the so-called ISO norm 8859 (ISO is the abbreviation of the "International Standardization Organization"). The same effect can be demonstrated by looking at the WWW pages offered by the Helsinki "Neo-Assyrian Text Corpus project". This project aims at collecting all Neo-Assyrian texts in an electronic data base, as you can read in its homepage (http://www.helsinki.fi/science/saa/). Within its further WWW pages, we not only find informations about the actual exchange rate of the Finnmark and weather report data, but also a description of the project data base itself and a sample page containing a piece of text (http://www.helsinki.fi/science/saa/database.html). As with the Edinburgh text specimen, you will immediately note the awkward and not easily self-explanatory usage of symbols such as the "ampersand" (&), the percent symbol (%), or the "at" sign (@) which is called "Klammeraffe" by German computer freaks.

In which way now could it be possible to make the texts available via the net in a way familiar to all readers without regard of the computer system they use? One way could consist in representing the data in graphical form. This method has disadvantages though, given that the transfer of graphical data is much more time consuming than the transfer of textual data stored in characters, and the "readability" of graphical data depends a lot on the quality of the system the reader uses. Of course I do not mean that graphical data should be dismissed in the field I am talking about — as I said before, it is one of the great advantages of internet publications that they can easily combine graphical and textual data, and it would really be an achievement for future work in the field if we could always have the original texts at hand, reproduced "in a way which allows control of the hand-copies", as Karel VAN LERBERGHE styled it[5]. For all kinds of textual

[4] Jesus-Luis CUNCHILLOS and his staff talked about the tool named SIAMTU on the RAI.

[5] In his paper presented on the RAI.

analyses, however, it is necessary to make the texts retrievable not in graphical form but represented by characters. Petr VAVROUŠEK and I have for long been discussing about how to design a font that would not only meet with the requirements of all languages involved but also with the peculiar structure of the internet. In Table 1 you can see what we offer as a basis for discussion now. This font contains **a)** the so called "ASCII" set of characters which is used in all existing computer systems nowadays, i.e. letters from A to Z as well as numbers, various punctuation marks and symbols (byte values between 32 and 127); **b)** Latin characters with diacritics that are necessary for a transcription of the languages involved in Near Eastern Studies, that is, Semitic, Indo-European, Sumerian, and others (values between 128 and 229); **c)** superscript and subscript variants of several letters and the numbers which allow for a unique, one-to-one transliteration of cuneiform characters (values between 230 and 252); and **d)** a set of marks to be used for the separation of text elements belonging to different languages but occurring side by side within a given text as, e.g., Luvian text within Hittite, Sumerian words within Babylonian, and the like (values between 1 to 31). Unfortunately, the font cannot be made usable for the net as it is represented here, given the restriction that in a UNIX environment as well as on Macintosh or Windows systems a character set to be displayed and printed can only contain some 192 characters, not 255 as in our font (usually, byte values below 32 are not freely usable). This means that as long as the "special" characters we need are not integrated into Unicode and as long as Unicode itself with its 65,536 characters is not used as the basis of all systems, we will have to dismiss some of the characters. I do hope though that we shall soon be able to finally decide what encoding system we shall be using in near future for internet applications concerning ancient Near Eastern languages. Once a usable font of no more than 192 characters has been designed, it can easily be adopted to any one of the systems used nowadays, and if we find a common solution (everybody is invited to make suggestions), we can even try to convince the Unicode consortium that the special characters we need should be adopted into their encoding standard.

There is one more question that deserves to be discussed, and I should be glad if we could arrive at a solution of it soon. For about ten years now, I have been busy myself establishing a data base that is designated to cover all relevant textual material from Ancient Indo-European languages; and it has become more and more realistic that the first aim of this project (named "Thesaurus Indogermanischer Sprach- und Textmaterialien", abbreviated as "TITUS") can be achieved within this century still, viz. entering the texts and preparing them structurally for an electronic analysis (information about TITUS is available under http://titus.uni-frankfurt.de/ and http://titus.uni-frankfurt.de/texte/texte.htm). It goes without saying that within such a data base, the material from Indo-European Anatolian languages (http://titus.uni-frankfurt.de/texte/texte.htm#anat) forms one of the most important parts. As for Hittite and its sister languages, I was extremely grateful when Petr VAVROUŠEK offered his support in this respect: Together with his colleagues in Prague, he has been busy entering the texts that are available in edited form, either in transcription or in autographs, and his collection is rapidly increasing. Additional work has been done by H. Craig MELCHERT, Christian ZINKO, Johann TISCHLER and Sigurdur PÁLSSON so that several single corpora such as a Luvian

or a Lydian one have been completed. Although we would intend to make this material available to the public, we cannot do so now because we have been warned by the "Akademie der Wissenschaften und der Literatur" in Mainz that this would imply an unlegal attack against copyright laws. I do not want to get into detail here as far as the correspondence I have had about this topic is concerned. Instead, I should like to raise some questions publicly, hoping that they might invoke some discussion. My first question is, who is the copyright owner of a Hittite text like Anitta's historical report of the defeat of Neša? Is it Anitta himself or somebody who prepared a printed edition of the text? What are the editor's rights if somebody does not take his wording as the basis of a digitalized electronic version but if he tries to establish a new interpretation on the basis of the published autographs (cf. to the sample page provided by Petr VAVROUŠEK which is printed in Table 2 below)? What rights are at all involved if the scholarly world attempts at exploring new means for their scholarly purposes? Has the so-called copyright of an editor to be ranked higher than the right of free scholarly activity (which is guaranteed, e.g., in the German constitution)? And how does the status of a public institution such as the Mainz Academy agree with the fact that it tries to prevent scholarly activities rather than supporting them? Of course I am not a lawyer, and I am not interested in entering into a court trial about these questions; instead I do hope that an agreement may be found soon. As far as my own point of view is concerned, let me just state some theses as to the questions raised above. In my opinion, everybody is free to prepare and publish, by any means, what he or she reads in a cuneiform tablet that is made accessible as a photograph or as an autograph. If the reading acquired like that, is identical with a reading established in a printed edition, this speaks in favour of both the autographers' and the readers' skill, and it does not involve any kind of copyright. If the printed edition is collated for the purpose of establishing a new reading, this is a normal and necessary procedure in philology, and it does not involve any copyright conflict either.

The material we have to deal with was collected using financial support of national or international institutions, and it must not be treated as a personal property of anybody but as a part of what we call all-human heritage. This heritage must not only be preserved, but it must be worked on in order to reveal the informations it conceals with respect to a very important stage within the history of mankind. It is our common task to do this work, and whether we use computers for it or just pencils, is not a matter of legality but of effectiveness.

	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
000	ᛊ	ᛊ	ᚾ	ᚾ	ᚺ	ᚺ	ˋ	¨	ˆ	˝	̣	˙	̯	⊢	⊣	ᴾ	ᑫ	⊦	⫞	
020	¯	§	⸢	⸣	˅	˄	⸤	⸥	I̯	U̯	⟦	⟧	̑	!	"	#	†	°	+	’
040	(	)	*	+	,	-	.	/	0	1	2	3	4	5	6	7	8	9	:	;
060	〈	=	〉	?	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
080	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_	‘	a	b	c
100	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
120	x	y	z	{	|	}	~	≈	á	à	ã	â	ä	ā	é	ē	ê	ě	è	ẽ
140	î	ì	ĩ	í	ī	i̯	ó	ô	ö	õ	ō	ú	ù	ũ	û	ū	ü	u̯	ŭ	æ
160	ǣ	ə	Á	À	Â	Ā	Ä	Í	Ì	Ī	Î	Ú	Ù	Ū	Û	U̯	Ü	É	È	Ẽ
180	Ē	Ö	ḳ	k̲	ğ	g̃	ġ	γ	g̵	ǵ	t̲	ṭ	þ	ϑ	Ṭ	ḍ	ḏ	d̄	Ḍ	c̣
200	č	č̣	c̄	ḇ	β	p̄	ṗ	p̱	ṅ	ñ	ν	ṇ	ṃ	m̃	ḫ	ḥ	ẖ	Ḫ	ṣ	ś
220	š	Š	Ṣ	ẓ	ž	ṣ̌	Ž	ṛ	ḷ	λ	^{1}	^{2}	^{3}	^{4}	^{5}	^{6}	^{7}	^{8}	^{9}	^{0}
240	$_{a}$	$_{u}$	$_{ú}$	$_{e}$	$_{i}$	$_{x}$	^{c}	^{ɔ}	^{a}	^{e}	^{o}	$^{i̯}$	$^{u̯}$	⌐	¬					
	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9

Tab. 1: System independent transcription font (proposal)

A74 nu LÚ ^{URU}Pu-ru-uš-⸢ha⸣-⸢an⸣[x
B16’ *Ù* LÚ.MEŠ ^{URU}Pu-ru-uš-ha-an-da kat-ti-mi hé-en-ku[m
$C_3$3’ LÚ.MEŠ x x x[ ]-mi [
nu LÚ ^{URU}Pu-ru-uš-ha-a(n-da kat-ti-mi hé-en-ku)[m-uš]

A75 šu-mu I ^{GIŠ}ŠÚ.A AN.BAR I PA.GAM AN.BAR
B17’ AN.BAR hé-en-gur ú-da-aš
$C_3$4’ x-]mu I ^{GIŠ}ŠÚ.A AN.BAR I PA.HAL AN.BAR hé-en[
šu-mu I ^{GIŠ}ŠÚ.A AN.BAR I PA.GAM AN.BAR (hé-en-gur ú-da-aš)

A76 ma-a-an a-ap-pa-ma ^{URU}Né-e-ša
B17’ ma-a-an EGIR-pa-ma ^{URU}Né-i-ša
$C_3$4’]EGIR-pa-ma (5)]x-un
ma-a-an a-ap-pa-ma ^{URU}Né-e-ša [ú-u̯a-nu](-un)

A77 ⸢nu⸣ LÚ ^{URU}Pu-ru-uš-ha-an-da kat-tim-mi
B18’]an-da kat-te-mi pé-e-hu-te-nu-un
$C_3$5’ nu LÚ ^{URU}Pu-ru-uš-ha-an-da kat-t[e]nu-un)]
nu LÚ ^{URU}Pu-ru-uš-ha-an-da kat-tim-mi (pé-e-hu-te-nu-un)

A78 ⸢ma⸣-a-an tu-un-na-ki-iš-na-ma pa-iz-zi a-p[
B18’ ma-a-an ^{URU}Za-al-pa-ma pa-x[
$C_3$6’]x-iš-ma pa-iz-zi a-pa-a-ša
ma-a-an tu-un-na-ki-iš-na-ma pa-iz-zi a-p(a-a-ša)

A79 ⸢pé⸣-⸢é⸣-ra-am-mi-it ku-un-na-az e-ša-ri
B19’]az e-ša-ri
$C_3$6’ pé-ra-a-am-mi-i[t]e-ša-ri
pé-é-ra-am-mi-it ku-un-na-az e-ša-ri

Tab. 2: Sample passage from the Anitta text