HOW WE READ CUNEIFORM TEXTS*

ERICA REINER
UNIVERSITY OF CHICAGO

Bist du beschränkt, dass neues Wort dich stört?
Willst du nur hören, was du schon gehört?
Faust II

The student who wants to teach himself to read cuneiform texts written in the Assyro-Babylonian language sets to his task with a handy stack of books at his elbow; they are: a sign list or 'syllabary'[1], a grammar, and a dictionary. Any other books, such as chrestomathies, reading exercises, or the like, are of incidental importance beside these basic tools.

The sign list is there to identify the material marks on the clay or their transposition into ink on paper and, once identified, to equate such

* Abbreviations used in this article include: LAA = Reiner, E., A Linguistic Analysis of Akkadian. (Janua Linguarum, Series Practica, 21.) The Hague, Mouton. 1966.

1. The latest such book is W. von Soden and W. Röllig, Das akkadische Syllabar, 2nd ed., Rome, 1967. The following remarks have been stimulated by reading and rethinking this book and the problems it deals with. Examples quoted from the Syllabar are intended as illustrations of theoretical points and in no way as criticisms of the particular example or indeed of this work as a whole which is consistent and in keeping with its own premises. If the premises are changed, however, another picture emerges. The analysis offered here, so far as it differs from that of the Syllabar, differs in being based on a change in premises. These premises, and the resulting corollary statements I was fortunate to be able to discuss with a number of colleagues, Assyriologists and linguists, especially Gene Gragg, A. L. Oppenheim, J. Renger, and A. Westenholz, who have read and criticized the manuscript in its various stages of completion. B. László not only supplied much of the linguistic background and methodology, but also gave generously of his time to supervise the details of the formulations. I need hardly add that the responsibility for these statements rests solely with me; I believe it is, however, an indication of the surge of interest in questions of structure and of the variety of possible approaches to them that substantially the same concerns could be stated in widely differing ways by those of us who cared to formulate them, as exemplified by the article of Renger cited note 24.

marks with one or more letters of the (Latin) alphabet, that is, to make a transliteration;[2] the grammar and the dictionary serve just as much to segment this transliteration into meaningful units — words or morphemes — as to find their equivalents in another language.

The concept of the sign list is as old as cuneiform writing itself. While the layout and arrangement of dictionaries have undergone substantial changes, the layout and arrangement of syllabaries have not diverged much from those of the sign lists compiled by the ancients and, perhaps for this reason, are taken for granted. What are, basically, the ancient sign lists? Tablets with cuneiform signs which are accompanied (usually on the left side) by one or more cuneiform signs, either of the same size or smaller. The list may contain each cuneiform sign once or more than once — in the latter case the sign may either be repeated or the space under it left blank — and correspondingly one or more than one group of signs to its left. The signs to the left have been found to be reading glosses for the sign they accompany, and usually do not exceed a repertoire of about one hundred different signs; their transcription into, e.g., the Latin alphabet would, as a rule, represent a syllable consisting of an obligatory syllabic component (a vowel), and an optional nonsyllabic component (a consonant) preceding and/or following it. These tables in which a syllable is assigned to each sign have been named syllabaries.[3]

The signs provided by such glosses in the syllabaries are usually ordered according to their shape; and in rare cases, by their syllabic equivalents, arranged first according to the consonants, then the quality of the vowel in the syllable, in the sequence *u-a-i*, i.e., *bu*, *ba*, *bi*, *ku*, *ka*, *ki*, and the like. In essence, then, these ancient syllabaries are tables of cuneiform signs and identify the signs through their phonemic content.

2. One may also start with the transliteration, namely with the sequence of alphabetic characters or groups of characters separated from each other by blank space or by some punctuation mark such as a period or a hyphen, and transpose some of these groups of characters into other groups of characters, in other words, change — usually in order to improve — the transliteration, e.g., change *ina pī lab-bi* 'from the mouth of a lion' to *ina pī kal-bi* 'from the mouth of a dog'. This transposition may seemingly be done directly from one group of alphabetic characters into another, on the basis of a correspondence table (*lab* = *kal*); in reality, it is done by mentally transposing the alphabetic characters into the corresponding sign, and by retransposing that sign into one of its equivalent alphabetic correspondences.

3. "Dans un grand nombre de cas, la lecture que fournit la première colonne est précisément la valeur phonétique du caractère dans les textes assyriens. Aussi crut-on d'abord — et c'est ce qui ne pouvait manquer d'arriver tant qu'on ne connaissait que des inscriptions assyriens — que ces tables de caractères avaient été dressées pour former les étudiants de l'Assyrie à la lecture de leur propre langue. C'est alors que l'on adopta la désignation de *Syllabaires*, considérant que la première colonne renfermait l'indication de la valeur phonétique et la troisième l'explication idéographique des signes figurés dans celle du milieu." F. Lenormant, Les syllabaires cunéiformes (Paris, 1877), p. 9.

A similar function is performed by modern syllabaries, and their layout is not unlike that of the ancient ones. Thus, a modern syllabary consists, essentially, of a list of cuneiform signs and their Romanizations, with appropriate documentation of the string or strings in which their Romanization, called *value*, can be inserted. The signs are 'alphabetized', i.e., arranged in a conventional sequence based, first, on the direction, then, on the number of the wedges that compose them. First come the signs which begin with one horizontal wedge, then those with two, three, and four horizontals; there follow the signs which begin with one, two, etc., so-called wedge-heads, and finally those which begin with vertical wedges. Each group is subsorted according to the same sequence. The signs may be reproduced in some stylized manner based on the form they have in a given period of cuneiform writing, or they may be given a name[4] as identifier; this name is normally a distinguished value taken from among the syllabic values of the sign; it may, however, be a name devised in some other manner, such as a Sumerian value, or a combination of the names of its component parts. The signs also usually have an identifying serial number. For instance, the first edition of von Soden's Syllabar identified the sign by a serial number (column 1) and its name (column 2); the second edition contains all three identifying properties through the addition in column 2, beside the name, of a drawing of the sign. The value or values of each sign are listed in column 3; the documentation is given in column 6. The occurrences are more or less amply documented depending on the certainty of identification, that is, well-established values are not documented at all, whereas rare or newly suggested ones are usually documented fully.[5] If there are several dissimilar values of a sign, the sequence in which they are listed is arbitrary. (So is, of course, the system of diacritic marks that appears on the Romanized values.) It is important to note that although it appears in the form of a prescriptive chart, the list of values in the syllabary — ancient and modern — is merely a set of descriptive statements, noting and documenting actual occurrences.

Let us pause to ask how such a syllabary actually enables one to read cuneiform texts. For an answer, we can try a little experiment. Let us

4. This is not the Akkadian or Sumerian sign-name, usually a descriptive name in either language, that some ancient lists also include.

5. In this respect, too, the modern syllabary parallels the ancient ones; items into which the sign enters are not illustrated if they are common, but are illustrated if an uncommon value is to be substituted in them; such illustrations appear in the form *ša* ITEM $(= \text{sign}_i + \text{sign}_j) = \ldots$, in which sign_i or sign_j exemplify the sign, and the right-hand side of the equation identifies the item by its translation into Akkadian, e.g., *ša* GA+EŠ$_8$ *ga-iš-šu* as illustration of the value *eš*$_8$ of the KASKAL-sign. The so-called *diri*-type lists are made up exclusively of such items (without the introductory *ša*).

resort to the fiction of programming an information transducer, a machine to read cuneiform texts. While so far only human beings have learned it, it is equally possible, and may one day be tried, to teach this skill to a machine. Since a machine lacks human intelligence and intuition, it will have to be given a formalized description of the language to be read (in this case, Akkadian; in actual fact, preferably one dialect of it), and an algorithm, i.e., operational rules that it can apply to the input in order to produce a correct reading of it. Let us then simulate the machine for a moment, to achieve a better understanding of the operations of our own biological transducer, i.e., the human mind.[6]

Since, for all machine operations, data are encoded numerically or, with the addition of letters, alphanumerically, we may immediately conceive of the text — the string of cuneiform signs — to be fed into the machine as sets of alphabetic and/or numerical symbols. Our symbols may, for instance, be serial numbers given to cuneiform signs in some prearranged sequence, or alphabetic symbols, or alphabetic symbols combined with numbers, which may also be conceived of as indexes. Let us stipulate the use of alphabetic symbols functioning in the same way as serial numbers. Since there are more cuneiform signs than letters of the alphabet (e.g. the Latin alphabet), there will be cuneiform signs that have to be identified by means of more than one letter of the alphabet: namely, after we have exhausted the alphabet once (after 26 cuneiform signs), we will have to give the subsequent signs groups of symbols, such as aa, bb, cc, . . ., aaa, bbb, . . ., or aa, ab, ac, . . ., aaa, aab, . . ., or some such combinations. Since the alphabetic symbols we purport to use are arbitrary anyhow, we may also denote each sign by a combination of letters which are pronounceable like syllables. Further, again as a convenient mnemotechnic device, we may as well pair each cuneiform sign with a group of letters that we otherwise know to represent a syllable associated with that sign. There is, in other words, nothing to hinder us from giving the code name AB to label the cuneiform sign with a value *ab* and not to some other, the code name BA to label the sign with a value *ba*, and so forth.

Since a cuneiform sign may have as values, that is may be represented by, not only one syllable as defined above p. 5, but other syllables which are either partially similar or totally dissimilar to the first, this information, too, must be given to the machine. The cuneiform signs thus encoded constitute part of the vocabulary of the machine; they have to be supplemented by operational reading rules, i.e., rules for selection from this vocabulary of possible values.

6. Such an experiment was conducted with a human performer under conditions simulating a machine, and the reading rules suggested here have been developed and modified on the basis of the simulated machine performance.

The human reader uses the same initial data: the signs and their corresponding values, and performs the same operation of selecting appropriate values as the machine. He uses his intelligence and intuition to internalize the algorithm which he does not need to be aware of step by step. Therefore an analysis of the machine's performance makes explicit the process involved in reading cuneiform, and the set of instructions — called the program — given to the machine is sufficient for the human reader as well.

For illustration of this process, let us take a few lines from the Flood Story (from Tablet XI of the Gilgameš Epic), a story which appears in most chrestomathies for learners of Akkadian. (This text is used in preference to another 'classic', the Code of Hammurapi, or to an Old Babylonian letter, because it illustrates a greater variety of points.) The text is encoded with the distinguished value of each cuneiform sign, as explained on p. 5.

Sample Text

(Epic of Gilgameš XI 9–20)

1. # LU–UB–TE–KA–AN–IZ–ṬU–BAR–A–KUR–NI–MUŠ–TI #
2. # Ù–PI–SAG–TA–ŠÁ–AN–MEŠ–KA–A–ŠÁ–LU–UG–BI–KA #
3. # URU–ŠU–RI–IB–ḪU–URU–ŠÁ–TI–DU–ŠÚ–AD–TA #
4. # [ ]–ÍD–BU–RAD–TI–SAG–NU #
5. # URU–ŠU–Ú–LA–BIR–MA–AN–MEŠ–GIR–BU–ŠÚ #
6. # A–NA–ŠÁ–GAN–A–BU–BI–UB–LA–ŠÀ–BA–ŠÚ–NU–AN–MEŠ–GAL–MEŠ #

. . . . .

12. # A–KUR–SU–NU–Ú–ŠÁ–AN–NA–A–A–NA–KI–IG–KI–ŠU #

The scanning is envisaged from left to right; the code syllables are connected by hyphens, except those which end or begin a line on the clay tablet or have a word-divider mark between them. Thus the only segmentation available to the machine is the one that is physically perceivable on the original text.

I. ALGORIT IM

The algorithm consists of four parts:

(1) Basic value table look-up.
(2) Finding the ultimate value.
(3) Segmentation.
(4) Morphosyntactic analysis.

1. *Basic value table look-up.*

The first instruction having been, "Read the text," we start the look-up from left to right. It will suffice as an illustration to take only the first four signs, represented, for practical reasons, by the four code-syllables LU–UB–TE–KA–.

The next step is: For each successive code syllable, find all corresponding strings (basic syllabic values, for short: values) from the appropriate table:

$$\begin{Bmatrix} lu \\ ṭip \end{Bmatrix}\text{–}\begin{Bmatrix} up \\ ar \end{Bmatrix}\text{–}Ti^{7}\text{–}Ka$$

In our example, the table contains two basic values each for the first two code syllables, and one each for the third and fourth; if there are more values in the table, the procedure is the same. The values found through this operation, in contradistinction to the undifferentiated discrete units of the text which are code names for signs, actually consist of strings of lower level units, of vowels and consonants: in particular one vowel, and two, one, or no consonants.

2. *Finding the ultimate value.*

For any consonant contained in the basic value, check the appropriate table if applicable and substitute its ultimate values, if any, to obtain the ultimate, maximally differentiated value of the string.

One of the ultimate values of the basic value is the one chosen to represent the basic value (e.g., *up*). To find additional ultimate values different from the one which was chosen to represent the basic value, look up, if applicable, the following table and substitute for the consonant of the basic value, e.g., *p*, *g*, *k*, *q*, the adjacent, immediately neighboring consonant or consonants, e.g., *b* (for *p*), *q* (for *g*), *q* (for *k*), both *g* and *k* (for *q*).

p		*b*
t	*ṭ*	*d*
k	*q*	*g*
s	*ṣ*	*z*

If the basic value contains a consonant which, although it appears in the table, is not to be looked up there, it is capitalized (e.g., *Ti*). The ultimate value is then obtained by decapitalization, i.e., by replacing

7. The reasons for treating TE as *ti* are given on pp. 45ff. The capitalization of T and K is discussed under (2) below.

the capital by a lower case letter (e.g., *ti*). Capital letters will be found only in values of the shape CV, e.g., *Ti*, *Ka*.[8] The table is thus applicable to all basic values VC, CVC, and lower case (not capitalized) CV.

The criterion of adjacency stipulated, although obvious to the human reader, is — if we, temporarily, disregard the deficient top row — in practical terms equivalent to the use of three overlapping tables, corresponding to right-adjacency, ambi-adjacency, and left-adjacency, depending on which column of the table contains the consonant of the basic value, the left, middle, or right. Two of the tables, namely

ṭ	*d*		*t*	*ṭ*
q	*g*	and	*k*	*q*
ṣ	*z*		*s*	*ṣ*

consist of three rows of consonants (dental stops, velar stops, and sibilants) of two members each; one member belonging to the emphatic column, the other to the respectively remaining, non-emphatic column. The emphatic column is the intersection of, that is, the column shared by the two tables; the respective non-emphatic columns are the non-shared or differentiating columns, in other words, those that characterize the tables as distinct from each other: the one whose non-emphatic column is unvoiced, the other whose non-emphatic column is voiced. Consequently, one may be called for short the *voiceless subtable*, the other the *voiced subtable*. The third table is the union of the voiceless and voiced tables. All three tables are characterized by the fact that they deal with an emphatic column, but it is the third table that is the most complete emphatic table, and since the other two are called, respectively, voiceless and voiced, the name remaining for it is *emphatic*, short for 'complete emphatic', table. The division of the table into subtables is indicated in the diagram by the interrupted and dotted lines surrounding each subtable:

p	*	*b*
t	*ṭ*	*d*
k	*q*	*g*
s	*ṣ*	*z*

The symbol for the emphatic consonant is the corresponding voiceless consonant of the row, with a dot underneath: *ṗ*, *ṭ*, *ḳ*, *ṣ*. For typographic and other reasons, since the beginning of this century *ḳ* has been re-

8. The capitalization, i.e., a special mark, for values to which a certain operation is not applicable (instead of marking the converse case, when the operation is applicable) is in accord with the theory of markedness. We call *marked* a member of a set which is exceptional, less general, more restricted. In the cuneiform system of writing, basic values which are identical to ultimate values form the exception, and are found with some frequency only among CV values.

placed by the letter *q*,[9] which we have used in the table; since *ṗ* is not phonemic (or, there is no *ṗ* different from *p*, or, no labial was pronounced emphatic), we have emphasized this fact by putting an asterisk at the intersection of the labial row and the emphatic column instead of the expected *ṗ*, and in the subsequent discussion whenever the term 'emphasis' is used, it also refers to *p* in the labial series.

Since the instructions for using the table are couched in terms of adjacency, consonants contained in the basic values must be so chosen that instructions may be properly executed. Thus, final consonants will be spelled with the emphatic member of the row, e.g., *aṭ*; initial consonants will be spelled emphatic if the sign has three ultimate values, e.g., *ṭip*, voiced if the sign has two values: voiced and emphatic, and voiceless if the sign has two values: voiceless and emphatic. The letter-symbols chosen for the basic value differ from those customary in Assyriology, where preference is given to the voiced member if there is an ultimate value containing it. In this practice, the possible choices have to be individually specified for each sign.

After the table look-up, all ultimate values of the four basic values to which the table is applicable are:

lu	*up*	*Ti*	*Ka*
ṭip	*ar*		
ṭib	*ub*		
dip			
dib			
tip			
tib[10]			

The values for which the table is applicable share one phonological feature and differ in one phonological feature; in this case the table which provides values for reading (although reading rules do not need to have a phonological justification) reflects the phonological distinction — voice, lack of voice, emphasis — that pertains to some types of consonants, namely to those which participate in the opposition emphatic *vs.* non-emphatic, i.e., stops and sibilants (the non-emphatic members share the opposition voiced *vs.* voiceless).

In the sample text, other basic values that are subject to the look-up are *maṭ*, *laṭ*, etc., *riṣ*, *uq*, and so forth, for their final consonant, and *qir*, *qan*, and so forth, *ki*, *bu*, and so forth, for their initial consonant.

9. In fact, as with the letters used for the vowels (see p. 45), the former notation made the structure of the phonological system more explicit.

10. For the possibility of values *ṭVp*, *dVp*, etc., i.e., with the vowel unspecified, see p. 31.

Note that the values *Ti* and *Ka* are not subject to further table look-up, since their initial consonant is capitalized. These values would have to be spelled differently, with a lower case initial, for some dialects of Akkadian, such as *ṭi* for Old Akkadian, *ti* for Old Babylonian, the respective lower case letters indicating that the basic value has three, respectively two, ultimate values.

3. *Segmentation.*

The purpose of this operation is to isolate shorter strings in the stretch of text between line-beginning and line-end. Formally, it consists of eliminating hyphens where possible by:

(1) marking the presence of a boundary by replacing the hyphen by a + −sign (the + −sign marks morpheme boundary or the higher word boundary, since word boundary is at the same time morpheme boundary, but not the converse);

(2) marking the absence of a boundary by copying out the item without the hyphen and closing up the space and, whenever this closing up brings together two identical vowels, deleting one of them. However, the original form of the item copied is not erased but preserved as a possible alternate segmentation (e.g., *lup*, *lu-up*).[11]

In the remaining cases, boundaries cannot be established at this level, but by further, higher level analysis.[12]

The operations can be performed or not, depending on the vowels and consonants that end and begin a value. The following table illustrates the operations:

11. A further step that might be executed at this stage is the insertion of a special mark, e.g., the aleph-sign, or H (see p. 27), between two identical vowels to indicate that the deletion of one of them does not apply, e.g., *luʾup*. This sign, if inserted before each vowel preceded by a hyphen, such as C-ʾV, V_a-ʾV_b, is useful for certain segmentation procedures, see p. 27. For a fuller motivation, see my article in Studies Presented to A. Leo Oppenheim (Chicago 1964), p. 169 and note 4. However, the insertion of such a sign is simply an obligatory spelling rule which has no impact on the dictionary look-up, because in the current Akkadian dictionaries the aleph sign is ignored in the alphabetization of the entries. (e.g., *buʾārum* and *buārum* are two consecutive entries.)

12. The situation is similar to the 'total rhyme' pairs, such as the French alexandrins

Galle, amant de la reine, alla, tour magnanime,
galammant de l'arène à la Tour Magne à Nîmes

which are read as:

/galamã dəlaʀen alatur maɲanim/
/galamã dəlaʀen alatur maɲanim/,

i.e., which can be segmented into morphemes only at a higher level analysis.

letters on either side of hyphen		absence of boundary	undecided	presence of boundary
$C_{incomp}-C_{incomp}$	(two incompatible consonants)[13]			$C_{incomp}+C_{incomp}$
V_a-V_a	(two identical vowels)	V_a	V_a-V_a	
C–V	(consonant and vowel)		C–V	
$C_{comp}-C_{comp}$	(two compatible consonants)		$C_{comp}-C_{comp}$	
V–C	(vowel and consonant)		V–C	
V_a-V_b	(two different vowels)		V_a-V_b	

Turning to the example, at this point it is segmented, using the basic values, as

ṭip-up-Ti-Ka ṭip-ar-Ti-Ka lu-up-Ti-Ka lu-ar̤-Ti-Ka
lup-Ti-Ka

Each of these strings is an abbreviated notation for a number of strings which result from optional hyphen-elimination, e.g.:

ṭipupTiKa, ṭipupTi, ṭipup, ṭip.

In the following, we omit the capitalization of T and K as irrelevant for the discussion.

4. *Morphosyntactic analysis.*

This operation consists of finding the morpheme structure of the text's segmented strings which are not found in the dictionary. The analysis aims at isolating: (1) discontinuous morphemes, and (2) continuous morphemes, i.e., affixes, which amounts to breaking down the string into substrings, and repeating the steps of the analysis.

The string *luptika* has the canonical structure CVCCVCV, i.e., it consists of the interdigitation of a consonant structure C*CC*C* and its negative, a vowel structure *V**V*V (the asterisks serve as place-holders or slot-markers). The particular vowels (slot-fillers) constitute the vocalic pattern, and the consonants the corresponding consonantal root. It is more economical to begin with the analysis of the vocalic pattern, since this is the morpheme that represents the grammar (inflection, derivation), a closed list which is finite, while the consonantal root belongs in the lexicon, an open list, and is infinite in the mathematical sense, or indefinitely large. Elements remaining after the identification of the vocalic pattern, i.e., the grammatical category, and of the consonantal root, i.e., the lexical item, belong to the affixes.

The vocalic pattern of the string is C*u*CC*i*C*a* (or **u**i*a*, not *uia* by itself, because the relative position of the vowels among the neighboring

13. For incompatible consonants, i.e., a list of non-occurring clusters, see LAA 4.3.

consonants, indicated by the asterisks, is part of the pattern). This represents either one morpheme or more than one morpheme.

If C*u*CC*i*C*a* is one morpheme, the corresponding four consonants cannot represent a quadriliteral root, because there are certain constraints on the distribution of consonants in quadriliteral roots. The vocalic pattern under examination is part of the grammar only if the two contiguous consonants (between the first and the second vowel) are identical (pattern C*u*C_iC_i*i*C*a*); in that case it represents the imperative of the D-stem, e.g., *purrisa*. In other similar patterns the first consonant is an affix, *š* or *m*, and has to be separated from the string. The pattern C*u*CC*i*C*a* is found in a table of patterns, also called paradigm if the C-s are replaced by actually occurring root consonants, as in the cited *purrisa*. It is, of course, also possible to construct the table by using the letter C, as above (or asterisks, etc.), by which procedure we obtain a table of (vocalic) root patterns (a term used by Giorgio Buccellati, A Structural Grammar of Babylonian [in MS] p. 90). Such a table contains the root pattern C*u*C_iC_i*i*C, without the final vowel; the final vowel (*a*, but *i* may also occur in this position) is identified as the affix (*a* = plural suffix, *i* = fem. suffix).

Since CC in the above string is not C_iC_i, we cannot use this pattern and proceed to cut. We cut off, from the left, C*u*, i.e., the first consonant and the following vowel, the latter as a check for the consonant's occurrence with such a vowel. The segment C*u* is an affix only if C is one of the consonants *m*, *n*, *š* (with *s* as its morphophonemic alternant), *t*, *l*, *k*,[14] as it appears from a table of affixes (all prefixes and suffixes except the last two: *l* is only a prefix, *k* is only a suffix). *m*, *n*, *š* and *t* appear in the table of affixes of derivation, and as such may be followed by *u* or *a* (in this table, *m* and *n* are in complementary distribution, *m* occurring before roots which do not contain labials, and *n* in those which do, and thus belong to one morpheme). The table in which these appear contains both the prefix and the pattern of the stem, e.g., *mu*CC*i*C, *šu*CC*i*C, etc.

The consonants *n*, *t*, and *l* appear in the table of affixes as verb inflection prefixes; *n* and *t* representing actor prefixes of the 1. pl. and 2. respectively, and *l* the prefix of the optative 1. and 3., and as such may be followed by *u* or *i*. Thus, the consonants *n* and *t* belong to both derivation and inflection.

Since Akkadian has as continuous morphemes not only prefixes and suffixes, but also infixes of the shape *t*, *tan*, a segmentation of the infix is also a possible operation under certain morphophonemic constraints.

14. Only the consonants *m*, *n*, *š*, *t*, *l* enter into Akkadian prefixes; except for *l*, the same consonants plus *k* are the only consonants that enter into suffixes. See LAA 5.0 (where, however, *l* is not included in the list of consonants that enter into affixes).

If C*u* does not represent an affix, either it represents a stem, and CC*i*C*a* (here *ptika*) must be looked up in the table of affixes, or a further cut must be made, e.g., C*u*C-C*i*C*a* (*lup-tika*), C*u*CC*i*-C*a* (*lupti-ka*), and so forth.

The remaining segment of the string, CC*i*C*a* is looked up in the table of vocalic root patterns, and will be found there with the identification 'preterite.' We identify the consonantal root *ptk*, and look it up in the root dictionary. The dictionary may list either the roots (such as the older dictionaries based on the Hebrew pattern, or the list appended to an alphabetically arranged dictionary, as AHw.) or the stem in some conventional form, more or less redundantly, e.g., the stem and the pertinent stem vowel or vowels, e.g., **ptik*, or the infinitive and the pertinent stem vowels, e.g., **patāku* (*i*), as in the current Akkadian dictionaries.

If the segment C*u* did not represent a prefix, the alternatively cut segment C*u*CC*i* (*lupti*) would be segmented into the vocalic pattern and the consonantal root *lpt*. This root does occur in the dictionaries, and is cited currently as *lapātu* (*a/u*), the notation *a/u* standing for the stem vowels which are characteristic of the present and perfect (*a*) and the preterite and imperative (*u*) respectively.

Since in CC*i*C*a* the C-s cannot be substituted by root consonants (radicals) that occur in the dictionary, the string is further segmented into CC*i* and C*a*. First, the segment C*a* is looked up in the table of affixes. If C is *k*, it appears in the table of affixes as nominal possessive suffix and verbal suffix both; if C is *š/s*, it appears in the table of affixes as prefix and as nominal possessive suffix.[15] If the vowel is not *a*, but *i* or *u*, different entries in the same tables apply. If C is *m*, the affix is listed in the table as a clitic. The identification of an affix is determined by syntagmatics, that is, to what grammatical category the table of affixes applies: the clitics apply to any word, the nominal suffixes to nouns, the verbal suffixes to verbs (usually divided again according to transitive and intransitive verbs).

The remaining segment, CC*i*, is found as *pti*, or as *petû* (*i*), in the dictionary as verb stem, specifically, as 'weak' verb, i.e., a verb which contains less than three consonantal radicals, either two consonants and a vowel, semivowel, or other 'weak' radical, or one consonant and two vowels, semivowels, or other 'weak' radicals. In the inflection of such verbs, the vowel, semivowel and 'weak' radical are realized as length; consequently, we place length on the vowel and transcribe *luptīka*.

The direction of segmentation is irrelevant for the result of the analysis so long as no segment remains unaccounted for. If the segmentation had

15. The morphophonemic alternant (allomorph) /*s*/ of /*š*/ appears only in suffixes. The environmental conditions for this alternation are definable (after dentals, sibilants, and *š*), as in the string A-KUR-SU-NU of the sample text line 12, segmented as *a-mat+su-nu*.

proceeded from the right, and the first cut made before C*a*, the remaining segment C*u*CC*i* is identified from the table of patterns either (a) from the table of root patterns as 'imperative' + fem. suffix, and the root as *lpt*, or in the conventional form *lapātu* (*a*/*u*); the imperative affix (*a* or *i*) has been discussed above; or (b) from the table of derivational patterns C*u*CC, which identifies nouns. The segmentation of C*u*CC*i* as 'imperative' is rejected because imperatives do not co-occur with 2nd person suffixes, namely the following segment C*a*, above identified as *ka*, a nominal possessive suffix or verbal suffix of the 2nd person (masc. sg.). If C*u*CC is a noun, the following vowel is found in the affix tables for nouns; one such is the possessive suffix table, where *i* occurs (as 1.sg.) as well as the above identified *ka* (2.masc.sg.); and the other is the nominal declension table, where it is identified as genitive. This theoretically possible segmentation *lupt-i-ka*, i.e., noun + genitive + possessive suffix, is grammatical only if the string is preceded by other strings which require a following genitive; the alternative, noun + 1.sg. possessive + 2.masc.sg. possessive, is excluded because two possessive suffixes are mutually exclusive. Similarly, other segmentations, possible according to the grammatical tables, may be excluded on the basis of syntagmatic incompatibility (e.g., the verb being intransitive). Moreover, since both nominal and verbal suffixes are subject to syntactic concord, the final segmentation only appears after the larger context is examined.

It can be seen that all the essential information pertaining to grammar has to be available for the analysis of the morphological structure of even one word. The rest of the sample text may be similarly segmented, but the operations required will be mostly repetitions of the ones described above, and the purpose of this paper is not to present the whole grammar.

The lexical items isolated either are associated as derivatives with a radical, i.e., a group of consonants or combination of consonants and vowels with a semantic field, or are free morphemes whose semantic sphere is not shared by other items containing the same radicals, such as the preposition *ina*, the loanword *paššūru*. Moreover, the Akkadian lexicon includes foreign words (Sumerograms), to which the discontinuous morpheme analysis is not applicable, and which are to be looked up in a dictionary of foreign words. Foreign words contained in Akkadian texts may normally occupy not more space in an Akkadian dictionary than a glossary, similar to the pink pages of the Larousse dictionary.

The operations of the aforegoing machine-like linguistic system have shown that the reading of a cuneiform text is based on information which includes all three components: syllabary, grammar, and dictionary, as indicated in the beginning, and also that the three components are interrelated (this fact is also revealed by the cross references of von Soden-Röllig Syllabar, GAG, and AHw. to one another); or, in Saussure's words, "tout se tient" in the system that we call language.

The simplest dictionary to provide the necessary information would be one to list all forms of all words and bound word groups, so that any form in the text can be found by a simple search. Such a dictionary would be, of course, very voluminous, and therefore it is not customary to compile dictionaries of that form, for any language. A more analytic dictionary, such as those we commonly visualize, lists the free morphemes, and possibly also the bound morphemes, i.e., affixes.[16] For Akkadian, a more analytic dictionary would list both discontinuous and continuous morphemes (such as *p-r-s* and *i-i*, *ina*, *paššūru*), and contain references to rules of grammar.

The grammar would contain various rules, including tables such as referred to under section (4).

The syllabary can again be envisaged as the simplest: it would contain all ultimate values that could be substituted in a text; such a syllabary can actually be compiled, because the list of items is smaller than that of the comparable simplest dictionary. If the syllabary, however, were constructed as dictionaries are, more analytically, it would have to include generalizations comparable to those a dictionary includes and contain the basic values along with references to rules for converting basic values to ultimate values.

It is this latter solution that we should investigate further. I will do that by discussing in greater detail the steps illustrated above, first the rules for segmentation and selection (pp. 17ff.), then those regarding alternations, trying in the process to clarify the term 'morphophonemics' (pp. 33ff.), and only after that take up the question whether phonetics is part of the information necessary to reading cuneiform texts (pp. 39ff.). Finally, I will show what other form an index to a syllabary may take (pp. 55ff.).

Since we have used the fiction of programming a machine to read cuneiform, a general point should be made here. Although the number of table look-ups necessary to arrive at the correct reading of some string is, to be sure, very large, the time necessary for these operations is negligible if they are in fact performed by a machine due to the speed with which it performs any single operation. Moreover, it is not necessary to store the data in the machine in an alphabetical order as here assumed; they can be arranged according to frequency. Such an arrangement would actually simulate the human memory, which stores the most frequent items closest to recall. This is why the human reader of the text will pick out the 'familiar' items without numerous look-ups (for instance,

16. Note that the CAD does not list affixes, and the AHw does, but not exhaustively; for instance, the first entries, under A, are five bound morphemes (*-a*, *ā* I, *ā* II, *-ā* III, *-â*), but not *a-* (the 1.sg. actor prefix), while under E five bound morphemes, including the prefixes *e-* and *ē-*, are listed with a cross reference to GAG.

lupti- before *luarti* or *lubti*, etc.), and will resort to a look-up, i.e., dredge his memory or consult a list (syllabary, dictionary, or grammar) only if the first identified item fails to make sense in the larger context.[17] In actual fact, the human mind can compete in speed with a machine because it need not follow through all look-ups — and for the same reason, it is also susceptible to error, by neglecting alternate possibilities.

II. DISCUSSION OF THE ALGORITHM

1. *Basic value table look-up.*

The items comprising the basic value table are essentially identical to the items that also appear in the ancient syllabaries; they are the main entries in the modern syllabary, which, as mentioned on p. 5, contain in column 2 the code syllable, and in column 3 the basic values (though often this column already contains the ultimate values as well). Modern syllabaries, of course, often contain additional information, not directly pertaining to the values. For instance, in the von Soden-Röllig Syllabar, there appears (in column 4) a notation by means of symbols to indicate whether the value is a Sumerian value, has been derived from a Sumerian value by some process such as changing its vowel or one or more of its consonants (see n. 36), or is based on the Akkadian translation of the Sumerian word denoted by the cuneiform sign, and the like. On the other hand, the numbers which (in column 5) indicate for each value in which historical period of the language it was current, inherently belong to the value and might have been attached as some type of index[18] to the value itself, thereby adding a diachronic dimension to the list. Theoretically, other contextual specifications are possible and have been used elsewhere: Labat's Manual d'épigraphie akkadienne indicates by means of symbols in what text genre (medical, astronomical, etc.) a given value is attested.

The Syllabar contains, in addition to the list of values (*Zeichenliste*), a second list which serves as index to the first. This second list (*alphabetisches Register*) is arranged alphabetically according to the value, and refers to the sign which is identified by its code name (a distinguished value, see p. 5) and serial number both. The alphabetical arrangement is modified, however, by making use of some of the alternations that we have used in the table on p. 8 (see p. 21). The list of values, which can be specified as directed from the sign to the value, enables the user

17. In fact, only failure to notice that the item selected makes little sense is the cause for most erroneous interpretations in Assyriology.

18. The subscript position is already claimed by another index.

to 'pronounce' the signs, and in that sense serves as a guide to phonetics. Phonetics, we may recall, owes its origin as a discipline to the concern with orthoepy, the science of pronouncing correctly what is written; rules for such pronunciation became necessary for languages in which, in the course of historic development, pronunciation and orthography ceased to cover each other.[19] The complementary, alphabetic value-to-sign, or rather, value-to-(code) name list, apart from serving as index to the reading list, provides the means for transposing one basic value into another via the corresponding sign or the code name common to both values (see note 2). In the case of a still living, or at least still written language, this second list would serve as a guide to orthography, by indicating which signs to use if one wished to write a message in cuneiform. In itself, the list would nevertheless not be sufficient, but would have to be accompanied by a set of writing rules which, beyond the mere selection of signs, would also provide rules for their tactics.[20] Since there is little need for writing in cuneiform, except for pedagogical purposes (in which case one is anyhow better advised to select actual text segments) or as a facetious or cryptographic device, no tactical writing rules have been developed. Were they developed, they would be more or less the counterpart of the segmentation rules and incompatibility rules discussed above pp. 11f. under section (3). Since the alphabetization of the value-to-sign list takes into account the table of orthographic neutralizations, it could be further reduced and rearranged; this will be discussed on pp. 55ff.

The basic values of a sign — and the more basic values there are, the less this is the case — permit no unequivocal choice in any given case. The values rather are strings or partial strings which by themselves or in combination with the surrounding strings refer to entries in the dictionary in which a particular word or affix may be looked up.[21] Consequently, the syllabary does not point directly to any particular dictionary entry; it does not dispense with the sometimes very numerous

19. For Latin, such pronunciation guide is given by the *Appendix Probi*, a 3rd century A.D. Latin work listing 227 entries of correct versus "vulgar" forms, in the arrangement "*x* non *y*" in which *x* indicates the correct or classical Latin pronunciation, *y* the Vulgar Latin one, as in *coqui non coci*, *tabula non tabla*, *labsus non lapsus*, etc.

20. Such writing rules have been written for Sumerian by G. Gragg, in Toward a Syntax of the Sumerian Verb: The Dimensional Infixes (Diss. Chicago, 1966) in MS. See notes 25 and 34.

21. The dictionary itself provides such cross references, sometimes for the initial value, e.g., '*bu-* see *pu-*', because *bu* and *pu* are two values of one sign, and sometimes for the initial consonant only, e.g., under the letter *d* 'see also under *ṭ*'. The traditional Akkadian dictionary is in fact a combination of a 'pronouncing dictionary' and a spelling dictionary, in the sense that words are cited as they are assumed to have been pronounced wherever comparative linguistics or other extraneous information solves the ambiguity of the writing system, but are cited as they are written in other cases, e.g., *b/puzu*, the entry indicating that the ambiguity of the writing with the initial having the two ultimate values *bu* and *pu* has not been resolved.

look-ups necessary to find an item, as has been illustrated; moreover, often all the references of a syllabary cannot dispel the ambiguity which only a larger context — for instance a sentence — or extraneous information can resolve. For instance, in line 1 of the sample text, the basic values for the group A-KUR are:

$$a\text{–}\begin{Bmatrix} kur \\ maṭ \\ naṭ \\ laṭ \\ \dots \end{Bmatrix}$$

More than one of these strings, once their ultimate values have been found, yield strings that appear in the dictionary, such as *amat*, *alad*; a choice between them can be made only in the light of a larger context. On the other hand, in line 3, the sequence ŠU-RI-IP-ḪU will be found to yield both *šuriḫu* and *šurippaq*, and, substituting ultimate values for the last two signs, also *šuribbaq*, *šuribbag*, *šuribbak*, etc. The choice between these is given in the dictionary, but the dictionary entry is based on information that lies outside this text segment: there is an entry '*šurippak*' illustrated by the spellings *šu-ri-ip-pak*, *šu-ri-ip-pa-ak*, etc. Without the existence of such an illustration as the second one somewhere in the corpus, the ambiguity cannot be resolved at all. Note that a syllabary with illustrations (such as Labat, von Soden) selects as its illustrations precisely those spellings which resolve such ambiguities; the difference between them and the above hypothetical dictionary entry (of a type which is found in all existing Akkadian dictionaries) is their distribution: the dictionary lists these spellings under one and the same entry, while the syllabary lists one under ḪU and the other, e.g., under PA.

Moreover, the syllabary, since it gives references for dictionary look-up, is by nature open-ended and not finite. This is exemplified by the growth of the number of values in the syllabaries, and by the possibility, built-in through the tables which direct to ultimate values, of increasing these when new texts, or new dictionary entries, are discovered.

In effect, then, the syllabary does not provide reading rules, or, if it does, only the context-free, not the context sensitive ones (for these terms, see p. 29); it is rather both an illustration of the spelling and a device to direct the user to the proper place in the dictionary and the proper grammatical form. While there is nothing wrong with using the syllabary as an illustration for all possible ways of spelling paradigmatic forms,[22] such a use burdens the syllabary with information that is

22. "Ein Syllabar muss . . . die Lautwerte enthalten, die eine auf möglichst exacte Wiedergabe grammatischer Formen ausgehende Umschrift braucht" Syllabar, p. xviii.

predictable, in other words, it necessitates the repetition of the same information in connection with a great number of signs, when this information could also be provided by a single rule — or a table — applicable to an entire class of signs. By a more analytic syllabary, we would not only achieve the same goal, but also gain an insight into the structure of the writing system which now has to be gathered from widely scattered statements. If we left in the syllabary only information that is not predictable (the basic values), and collapsed the predictable information into a small number of rules, the number of entries in the syllabary would be reduced without information being lost; the only thing that might be lost in the process might be the illustrations, but these illustrations properly belong in the dictionary as alternative spellings — and are in fact listed there — or in a spelling course where repetition is a pedagogical requirement (*repetitio mater studiorum*), and not in the syllabary, whether as 'reading rules' or as 'writing rules'. On the contrary, the information presented in the form of rules will also point to illustrations not now appearing in the syllabary; in other words, it will predict spellings.

In addition to the examples used in connection with the text segment presented above p. 7, I would like to further illustrate such a less redundant syllabary and its concomitant reading rules without, however, rewriting the Syllabar whose open-endedness[23] has just been pointed out and has been illustrated by reviews which for the most part consist only of adding further illustrations to the tacit rules included in it.[24]

Selection rules: The basic value table look-up corresponds to selection rules. They provide the data from which one of several values of the same sign is selected.[25] Different values of the same sign are of two types: either they are partially similar to each other, or they are totally dissimilar. The partially similar ones form the class of predictable ultimate values of the same basic value; the totally dissimilar ones are not otherwise predictable and must be given in a table, as illustrated above, p. 8.

23. This open-endedness, mentioned before (p. 19), is actually the strength of the Syllabar, since it proves that the predictions that are made for signs belonging to a certain class are correct also for other signs belonging to the same class. Had the tacit reading rules been made explicit, there would have been no opportunity for reviewers to increase the number of values.

24. Significant exceptions are the review of J. Renger, "Überlegungen zum akkadischen Syllabar," ZA 61 (1971), pp. 23–43, and the review of I. J. Gelb, "Comments on the Akkadian Syllabary," Or. NS 39 (1970), pp. 516–46, which address themselves to methodological questions. Unfortunately, this manuscript was submitted earlier than the publication of either of these two articles, and they could not be considered in the context of the present article.

25. The inverse process of selecting a grapheme to match a segment of a phonological string is called "Replacement rules" by Gragg, op. cit., pp. 44ff.

To some extent, however, at least as probability based on frequency, even for these there can be formulated rules which specify the selection of a value in an environment where it is to a great extent predictable, for instance: *up* → *ar* / *a* _ and *lu* → *tip* / _ {*b*, *p*} (which may be read as '*up* goes to *ar* in the environment "after *a*" ', etc.), that is, select the value *ar* and not *up* (i.e., select the value *ar* from among {*up*, *ar*}) when the preceding value ends in an *a*, and select the value *ṭip* (from among, e.g., {*lu*, *ṭip*}) when the next value begins with a labial, *b* or *p*.

These totally dissimilar, basic, values must be included in the basic table, that is, they must be listed in the syllabary. The Syllabar which differentiates between values by typographic means quite properly marks the basic values, in contradistinction to the predictable ultimate values, by bold face type.[26]

2. *Finding the ultimate value.*

This procedure, illustrated above pp. 8ff., gives rules for converting a basic value into one or more ultimate values by tabulated feature replacement rules. While the selection rules obtain among totally dissimilar values, and hence direct to lexically different items (i.e., words which have to be looked up under unpredictably different places in the alphabet), the choice among partially similar values may involve only a slight correction in the dictionary look-up, unless the feature replacement applies to the initial. The existence of partially similar values is due to the limitations of the writing system which provides no choice, or a minimal choice only, for certain types of values. For instance, the spelling offers no choice for writing a VC or CVC syllable, whether the final consonant is voiced, voiceless, or emphatic, when the consonant belongs to a class with three members which differ only in this phonological feature; the same sign must be used in all three cases. This is equivalent to saying that orthographic neutralization affects all VC values and the final consonant of all CVC values, since in syllable- or word-final position the phonemic contrast, if such exists, between voiced *vs.* voiceless *vs.* emphatic is orthographically neutralized. Due to this neutralization, some part of the corpus must contain a cross reference to, or else list

26. In the Syllabar's arrangement, each boldface type value is followed by lightface values which are partially similar to the boldface one. Some exceptions to this arrangement are found: non-predictable values are not set in boldface, e.g., apart from (presumably less common) signs, such as ŠUN (no. 7), none of the Romanizations of which are set in boldface, there are Romanizations totally dissimilar to the preceding boldface one which are not in boldface, e.g., *tuḫ* (no. 117); on the other hand, some partially similar Romanizations are nonetheless all set in boldface, e.g., both *be* and *bi*4 for no. 42, both *kul* and *qul* for no. 45. These exceptions are due to the fact that von Soden defines the purpose of the boldface differently: set in boldface are values that are most common in all or most periods "deren Kenntnis also unerlässlich ist" (p. xxxviii).

side by side, all items which differ from each other only in this feature. If this is done in the dictionary, either the items *abtu* and *aptu*, (both spelled *ap-tu*), e.g., will be listed in both forms under each entry as in CAD: '*aptu* (or *abtu*)' and '*abtu* see *aptu*', or there will be a generalized cross reference, such as '*ab* see *ap*', this to be repeated for each item or for each similar syllable. If this is done in the syllabary, one possibility is to list all pertinent items, e.g., for the code syllable AB the items {*ab*, *ap*}, and in the case of a three-way neutralization, such as the code syllable AD, the items {*ad*, *at*, *aṭ*}; another possibility is to give a rule: (Optional) *ab* → *ap*, *ad* → {*at*, *aṭ*}[27] or *at* → {*aṭ*, *ad*}, or *aṭ* → {*ad*, *at*}, i.e., by listing the values either in alphabetical order or in some other order, for instance an order based on phonological features, as *at* → {*aṭ*, *ad*} where the first member, *aṭ*, is obtained by the rule $\begin{bmatrix} + \text{ obstr} \\ + \text{ tense} \end{bmatrix}$ → flat, i.e., adding a feature (+ emphatic or + flat) to the consonant, and the second, *ad*, by the replacement rule $\begin{bmatrix} + \text{ obstr} \\ + \text{ tense} \end{bmatrix}$ → − tense, i.e., replacing a feature (− voiced or + tense) by its opposite (+ voiced or − tense).

Instead of providing this rule in each individual case, the cross reference rule can be formulated more generally:
$C_{vd} \rightarrow \{C_{vl}, C_{em}\} / V \_\_ \{C, \#\}$, or $C_{vl} \rightarrow \{C_{em}, C_{vd}\} / V \_\_ \{C, \#\}$, or $C_{em} \rightarrow \{C_{vd}, C_{vl}\} / V \_\_ \{C, \#\}$, i.e., any voiced (respectively voiceless, emphatic) consonant may be replaced optionally by the corresponding voiceless or emphatic (respectively voiced or emphatic, etc.) consonant after a vowel and before another consonant or word boundary (i.e., end of the word). (This rule is the conflation of two rules, one for the replacement — in the first rule — of a voiced consonant by the corresponding voiceless, the other for the replacement of a voiced consonant by the corresponding emphatic.) This rule, which may look unfamiliar, can be couched in the form of a table from which the choices can be immediately read off, such as the table on p. 8, but the table form is only a convenient visualization, not a change in any sense, of the above rule.

That such cross reference rules have to be included in the information necessary to read a text is evident not only from the already cited cases, but also from the existence in the system of other types of signs (CV signs) that never (i.e., in no environment) provide any information concerning the voice of the consonant which appears in their basic value, and also from the spelling habits of dialects in which the inventory of signs includes only one CV sign for either two of the three or for all

27. The notation of the Syllabar, e.g., ab^{p}, $ad^{t,ṭ}$ may be considered either an abbreviation of the lists {*ab*, *ap*} and {*ad*, *at*, *aṭ*}, or an idiosyncratic notation of the rule *ad* → {*at*, *aṭ*}.

three phonologically contrasting homorganic stops and sibilants.[28] For example, if a labial stop is followed by the vowel *u*, no dialect can indicate its voice, that is, there exists only one sign to write the two sequences — parts of a syllable — *bu* and *pu*. Again, in all dialects there exists only one sign to write the two sequences *za* and *ṣa*; only two signs — in varying distribution — to write the three sequences *da*, *ta*, and *ṭa*; in some dialects, there exists only one sign to write sequences for which in other dialects there exist three signs, e.g., their inventory includes only *du* which is used where other dialects use either *du* or *tu* or *ṭu*. In such cases, when the writing is incapable of specifying the voice of the consonant, the choice is based on factors extraneous to the writing system; similarly, it is possible and sufficient to indicate the fact that there exists such a choice by a general rule instead of the more redundant way of listing the existing choices for each sign.[29] The general rule is not only more eco-

28. It may be salutary here to wonder why no grammar of Akkadian has yet sought, or deemed necessary, to establish the phoneme inventory of Akkadian on the customary bases, either by distributional criteria, or the establishment of minimal pairs, be it directly from the spelling or secondarily from the text established from a particular spelling, that is, from the available written or morphophonemic evidence. Rather, the phoneme inventory appearing in the grammars is the Common Semitic inventory, with the omission only of those consonants (of Proto-Semitic, i.e., for all practical purposes, Arabic) that are said to be lost in Akkadian, namely the interdental fricatives and the "laryngeals" (i.e., the laryngeal and pharyngeal stops and fricatives ʾ, ʿ, *ḥ*, *ḫ*, *ġ*). Consequently, no one has formally proved the phonemic character and the distribution of the opposition voiced *vs.* voiceless or voiced *vs.* voiceless *vs.* emphatic, or the opposition stop *vs.* fricative, or stop *vs.* nasal in such sets as */p, b, w/, /t, d, ṭ, s, z, ṣ/, /k, g, ḳ/, /ʾ, ḫ/, /b, m/, /d, n/, /g, g/*. This is the more surprising seeing that the writing system fails to indicate some of these oppositions, as just noted, and that certain dialects of Akkadian have for centuries used a writing system which consistently failed to distinguish most of the mentioned oppositions. As the oppositions now regarded as phonemic began to be noted in the writing, and even then only in a more or less systematic way, only around the middle of the second millennium B.C., it is obvious that the phoneme inventory of (Common) Akkadian is the projection of second millennium data into an earlier system, based on and supported by etymology. It turns out, however, that, as our corpus or our understanding of it increases, the tacitly accepted oppositions do not actually correspond to the facts reflected in the writing system. The writing reflects partly less, and partly more oppositions; we can phrase this also by saying there are variations in the spelling of certain lexical items that cannot be explained on the basis of etymology or Akkadian morphophonemics. The suggestions for positing vowels and consonants beyond those accepted in the standard inventory, suggestions discussed on pp. 45ff., are based on this discrepancy; no suggestions have as yet been put forward, however, either questioning, or proving or disproving, the existence of the oppositions taken for granted. A suggestion to be expected on the basis of the available data is, e.g., that the Assyrian dialect (at least Neo-Assyrian) does not know the three-way opposition of stops (and of sibilants?), but maintains only an opposition emphatic *vs.* non-emphatic (see LAA 4.3.1.2).

29. Indeed, if we follow rules and make the appropriate substitutions, we avoid the practice, criticized by von Soden, Syllabar, pp. xvii f., of using always the basic

nomical, but also has, as mentioned on p. 20, a predictive value: even when no illustration has yet been found for one or more of the possible choices, it can be predicted that there may occur some item written with the not yet attested value of the sign.[30]

Similar generalizations apply to basic values of the shape CVC in regard to their initial consonant. For some, all three choices obtain, that is, the initial may be read as voiced, voiceless, or emphatic, and for some, two of the three. In order to apply the same cross reference rule as before, and consult the same table, we can write the initial consonant *p, ṭ, q, ṣ* (i.e., the emphatic member of the set) if all three choices obtain; we may choose another member of the set to denote the initial consonant in the other cases, so that the table given above p. 8 will apply. Thus, e.g., the basic values *ṭuḫ* and *ṭim* will yield the ultimate values *duḫ, tuḫ, ṭuḫ,* and *dim, tim, ṭim* respectively; the basic values *gur* and *dil* yield the ultimate values *gur* and *qur* and *dil* and *ṭil* respectively. Since it would appear that the alternation between voiceless and emphatic alone (i.e., without voiced) only rarely or never occurs, a basic value with a voiceless initial consonant will have no ultimate value different from itself. We can thus express by the basic value *tál* (Syllabar no. 223) or *kur* (no. 211) that the basic value is identical to the ultimate value.

Such a notation, with different initial consonant according to the number and kinds of choices, predicts cases which are not yet attested, i.e., in which the basic value is the ultimate value, either only with a voiced, or only with a voiceless, or only with an emphatic initial, such as *zik* (no. 128), *ṭàr* (no. 72), *tál* (no. 223), and in which the choice for the ultimate value is between voiced and voiceless (but not emphatic), such as *dug* (no. 164), or between voiceless and emphatic (but not voiced), such as *tar* (no. 11). If such cases did actually exist, we would have to forego the spelling proposed for basic values, and instead resort to the device of capitalizing the initial consonant when the choices do not apply, as we have suggested for CV values, and only leave the alternating values in lower case, e.g., write *ṭuḫ, ṭim, gur, dil,* but *Tál, Kur, Ziq,* and *Ṭàr.* However, when we survey ultimate values, we find that cases with two choices (voiced and voiceless, but not emphatic) are rare

---

value in our transliteration, thereby leaving the burden of looking up every pertinent cross reference to the reader, and thus obliging him to retrace step (2), finding the ultimate value, of the operation described on p. 8, whereas the next step, the segmentation, has already been performed in the given transliteration.

30. This prediction has already materialized for no. 322, ṬU, i.e., du_x, with two values, since it must be given the value <*du*> in *na*-ṬU-*te* (on the basis of variants *na-du-te* and etymology) in AKA 298 ii 10 and Scheil Tn. II 24; the same example also serves for $ṭu_x$ (three values): no. 30 TU also appears in *na*-TU-*te* (AKA 240:46), and thus the value *dú* under no. 30 has to be extended to still another dialect beyond those marked in the Syllabar.

enough so that the non-occurrence of the third option is likely to be accidental, and similarly that the cases with no choices are equally accidental. We may therefore be authorized to spell all such values with lower case: with an emphatic initial, e.g., *ṭuq*, to predict that whenever a CVC sign has two ultimate values, the third will eventually be found, and with a voiced initial, e.g., *ziq*, to predict that it will occur with both a voiced and an emphatic initial; this spelling *ziq* of the basic value covers the value *ṣig*, because there is no other way to write this sequence except the less economical one for the scribe to use two signs, *ṣi* and *ig*.

If the syllabary did not contain either a device for obtaining ultimate values of CVC values or a list of them, it would be the dictionary's job, as set forth on p. 21, to provide the cross references, perhaps in some general form, such as '*b*⟶ *p*'. There exists a practical solution for some languages written in the cuneiform writing system, and consequently using the signs *ba* and *pa*, *da* and *ta*, etc., as if the difference in voice were grammatically pertinent in these languages, although it is known to be pertinent in Akkadian only, but not in these languages, where such a difference cannot be proven to exist or in fact can be proven not to exist at all. In dictionaries of such languages, such as Hurrian or Elamite, the solution usually adopted is to alphabetize together *b* and *p*, *d* and *t*, etc., that is, the consonants which are phonetically similar to one another but differing in one feature in Akkadian. This system of alphabetization is used, by the way, in the Syllabar's alphabetic index for the final consonants of the values, see ibid. p. xl.

It is not only the features of voice and emphasis in stops and sibilants that are susceptible of generalization. There are certain basic values which are subject to the alternation of one of their consonants due to the fact that the spelling is sometimes assimilative (as *m* vs. *n* in English *impossible* vs. *intransitive* or in Spanish *embarrar* vs. *enlodar*), and sometimes analogical[31] (as in English *input*, pronounced [imput] and sometimes misspelled as *imput*). This particular kind of analogical spelling may be called *morphophonemic*, meaning that the spelling preserves the phoneme known from the morphology. (For a detailed discussion of morphophonemics, see pp. 33ff.).

This oscillation of the spelling can be illustrated by basic values

31. The terms *assimilative* and *analogical* are those of W. Sydney Allen, Vox Graeca, p. 32 (Cambridge University Press: 1968). John Lotz has coined for the latter (analogical or morphophonemic) spelling the term *morpheme-analytical* ("The Imperative in Hungarian," Uralic and Altaic Series 1 [1960], p. 90). Such a morpheme-analytical orthography is sometimes used by von Soden in AHw in the transcriptions of suffixed words, e.g. *qaqqas-sa* (p. 3a), *qāt-su* (p. 208a), the hyphen indicating morpheme boundary, whereas the CAD uses assimilative (phonological) orthography without morpheme boundary, i.e. *qaqqassa*, *qāssu*. The morphophonemic orthography (an orthography using underlying forms) of these words would be *qaqqad-ša*, *qāt-šu*.

ending in a sibilant or š, an appropriate example being the group PI-SAG-TA, in line 2 of the sample text. Among the basic values of SAG there may be one in the form of *riš*, and step 2 of the operation gives *pi-*$\begin{Bmatrix} sag \\ riš \end{Bmatrix}$*-ta*. This reading is sufficient information if the dictionary contains the entry *pirišt-*. If, however, the dictionary contains roots or stems, it will contain only *prs*, not *prš*. The ultimate value, i.e., the one which directs to the proper entry, is obtained by the application of a rule š → Z/__ {#, *t*}, i.e., substitute Z (namely, one of the three sibilants *z*, *s*, or *ṣ*) for š before *t* or word boundary. This rule can also be formulated as operating in the opposite direction, namely Z → š/__ {#, *t*}. This fact is due to the orthographic neutralization of the feature 'palatalization' in sibilants before *t* (and probably also before word boundary), i.e., of the difference between the sibilants *z*, *s*, *ṣ* and the palatal spirant š. In the case of the word serving as example, the spelling of the middle syllable varies between two signs, *ri* and *iš*, and one sign, SAG. On the basis of this variation, we may choose to give the sign SAG a value *riš*, and transcribe the word as *pirištu*.[32] We know from morphophonemics, however, that the š before the *t* corresponds to morphophonemic *s*. Hence, if we transcribe *pirištu*, we use a phonological notation, and register the phonological, assimilation, rule, that *s* → š / __*t*. On the other hand, we may choose for the sign SAG its ultimate value *ris* (derived from a basic value *riṣ*) and transcribe the word as *piristu*, thereby preserving the morphophonemic information concerning the third radical, *s*. As to the other spelling, with *ri-iš*, we may also preserve its morphophonemic information by giving the sign *iš* a morphophonemic value *is*. Which alternative we choose depends on whether we make a phonological — assimilative — transcription or a morphophonemic transcription. The phonological transcription follows rule Z → š / __ {#, *t*}; the morphophonemic transcription, the opposite rule š → Z / __ {#, *t*}. The table corresponding to both rules is

z š
s š
ṣ š,

to be read from left to right and right to left respectively. Preference for one of the two rules may be based on possible other references to it. Since there are other cases when a rule *s* → š will be applied (see note 15), the first rule, or the left-to-right reading of the table, will have a more general application, in addition to starting from basic values which preserve morphophonemic information. In any case, it is desirable always to use either one or the other in all pertinent cases. This would avoid

32. See Borger, BiOr 14 (1957), 190, n. 1.

the oscillation between phonological and morphophonemic transcription, such as, e.g., transcribing *pi-riš-tu* (listed in the Syllabar under the ultimate value *riš* of SAG) but (morphophonemically) *mi-ḫi-íṣ-tu* (from the root *mḫṣ*, listed in the Syllabar as illustration of the value *iṣ* of the sign IŠ).

Another general rule for converting basic values to ultimate values would read: $m \rightarrow \emptyset$ / CV __ # i.e., a basic value ending in *m* may optionally be read without the final *m*. The rule expresses the fact that although the final *m* in the inflection was lost after the Old Babylonian period, it was preserved, as traditional spelling, in the writing system.[33] As a consequence of the practice of writing CV*m* instead of CV# which once was CV*m* (e.g., in the suffixes *tum*, *nim*, etc.), the possibility enters the system of writing CV*m* instead of CV# even in morphemes which were not formerly written CV*m*, i.e., morphemes which never contained */m/*. Thus a CV*m* value is read either CV, conforming to the morphology of the period, or CV*m*, and constitutes a historical spelling, or, when the restoration of a historical */m/* would yield an ungrammatical form, constitutes a hypercorrect spelling, that is, a spelling that the writer considers correct on the basis of a false analogy. For those occurrences when CV*m* stands for CV in non-final position, which are rare, the pertinent values may be marked in some fashion so as to indicate that the final *m* is to be omitted. A possible notation would be, on the one hand, *rum* with reference to the rule, indicating that the ultimate value is *ru*, and on the other *ga*M, with capitalized *m* to indicate that the rule is not applicable and that the value will not yield an ultimate value *ga*. All other similar values would come under the stated rule and need not be individually marked.

3. *Segmentation*

The procedure for putting in word boundaries in a string[34] has been described above pp. 11f. The procedure of segmentation described there left undecided the presence or absence of a boundary before a vowel beginning a value, and deferred the segmentation to a later, higher level analysis. Thus the algorithm did not operate with the insertion of a special mark (aleph or H) that was suggested in note 11. Were such a mark to be used in transliterations, strings could be segmented between C-V, i.e., between a consonant ending a value and a vowel beginning a value, whenever this mark was not present, such as, in the cited example, between *ṭip* and *up-ti-ka*, *ṭip* and *ar-ti-ka*. However, it is a well-known fact that there are words — belonging to a particular morpho-

33. See von Soden, Syllabar, p. xxviii.

34. Called "Boundary rules" by Gragg, op. cit., pp. 41f., referring to the inverse process of placing grapheme-boundaries in Sumerian phonological strings.

logical class — to which the frequent segmentation 'word boundary between C and V' does not apply; e.g., there is no word boundary in the string *iš-al*. Such spellings may be rewritten so that the segmentation need not be tested. We may, for instance, rewrite the value beginning with a vowel with a consonantal initial, choosing as consonant, e.g., aleph, or H (acrophonic for *hiatus*) and rewrite the cited form as *iš-Hal* which may be read as one word, *išHal*. A similar notation may be used between two identical vowels to exclude the possible contraction V_a-V_a to V_a (exemplified in *lu-up-ti-ka* segmented both as *lu-up-ti-ka* and as *lup-ti-ka*). We may thus write *ma-'al-tu* or *ma-Hal-tu* (>*maHaltu*) but *ma-al-ku* (>*malku*). The conversion of H, a symbol of the writing, into a phonological or morphophonemic symbol is carried out on another level.[35]

Probable (as opposed to predictable) word boundaries can also be established on some frequency basis, and would involve commonly occurring clitics (such as *ma*) and suffixes (*šu*, *ša*, etc.).

35. The introduction of the juncture feature H will prove useful for morphophonemic analysis of Akkadian; rules for its conversion will give the morphophonemic rules pertaining to the roots with laryngeals, since forms spelled with H represent underlying forms and most likely will turn out to be identical to historically underlying forms, similarly to the H's of Proto-Indo-European reconstructed (underlying) forms. Such reconstructed H's or 'laryngeals' were deduced for Indo-European even before the discovery of Hittite in which such 'laryngeals' have correspondences in spelling, by internal reconstruction, for which the following Greek paradigmatic set may serve as an example:

The future passive of the verb *rhḗgnūmi* 'break' is *rhagḗsomaḭ*; the corresponding form of the verb *meḭgnūmi* 'mix' is *migḗsomaḭ*. The future of the latter, *meḭxō*, is obtained, after adjusting the ending, by inserting *e* before the first vowel (the root vowel) of *migḗsomaḭ*, which becomes non-syllabic, as indicated by the diacritic crescent underneath it. Correspondingly, the future of *rhḗgnūmi*, obtained by inserting *e* before the first vowel in *rhagḗsomaḭ*, is expected to be **rheu̯xō*. The reduplicated perfect of the verb *leḭpō* 'leave', *léloiṗa*, is obtained from the future *leḭpsō* by replacing *e* with *o* (and adjusting the ending); correspondingly, the reduplicated perfect of *rhḗgnūmi*, obtained from **rheu̯xō* by replacing *e* with *o*, is expected to be (taking into consideration that the perfect reduplication of initial *rh-* is *err-* as in *érrīga*, the perfect of *rhīgéō* 'shudder') **érrou̯ga*. Greek *ea* (with a syllabic *a*) contracts to *ē* as in the acc. of *Diomḗdēs*, uncontracted *Diomḗdea* contracted to *Diomḗdē*, and Greek *oa* (with a syllabic *a*) contracts to *ō* as in the acc. of *ēṓs* 'dawn', uncontracted *ēóa* contracted to *ēṓ*; the more so *eu̯* and *ou̯* (with a non-syllabic *a* which does not belong in the phonological inventory) contracts to *ē* and *ō* respectively. Hence we expect **rheu̯xō* to yield *rhḗxō* and **érrou̯ga* to yield *érrōga*, and indeed *rhḗxō* and *érrōga* aré the attested forms of the future and the perfect of *rhḗgnūmi*. The reconstructed non-syllabic *u̯* (appearing only in the underlying form but not in the surface form) is a morphophoneme belonging to the deep structure, symbolized by Indo-European scholars as A or H, and is called the hypothetical semivowel, the a-coloring semiconsonant or (by analogy to the Arabic a-coloring pharyngeal ḥ) the a-coloring laryngeal. This morphophoneme has turned out to be a part of historically underlying forms obtained by the comparative evidence of Hittite where this morphophoneme corresponds to a written ḫ.

4. *Morphosyntactic analysis*

The ultimate values discussed so far obtained from the basic values by substitutions make the morphological analysis required in step (4) feasible. However, the string itself in which these values enter is sometimes ungrammatical, i.e., the morphemes identified look 'strange' or the grammatical paradigm in which they are found gives an interpretation that is contrary to the requirements of the syntax. It is the custom of Assyriologists, for pedagogical reasons, to correct the written string. Since often a minor change in an ultimate value will make the string grammatical, some Assyriologists have considered it desirable to add yet other modified values to the syllabary, and such values indeed appear in the von Soden-Röllig Syllabar. However, if we examine the modified values, we find that they do not provide any new information, as I will try to show. The modifications are based, first, on the larger context of which the string is a part, and so we may call them *context-sensitive*, and second, on the information that the Assyriologist has at his disposal about etymology, language history, and other data that the scribe of the cuneiform text either did not possess or chose not to express in the writing (see also pp. 42ff. and note 51).

If we examine ultimate values so modified in the syllabary,[36] we find that the modification consists of changing a vowel or supplying an additional vowel in the basic value, and that the resulting ultimate value directs us to another table in the grammar (a) in the table of affixes, and (b) in the table of discontinuous vocalic morphemes.

A. RULES AFFECTING AFFIXES. (1) *Vowel replacement*, in the context of case endings. The rule is of the form $V_1 \longrightarrow V_2$ (in a specified environment, e.g., in the environment C_*m*), in the syntactic environment or context of a particular case. For example, the basic value *tum* has an ultimate value *tim* in the environment 'genitive', and *tam* in the environment 'accusative'; the basic value *nim* has an ultimate value *nam* in the environment 'accusative', and so forth. The fact that not all CV*m* values have such modified readings in the syllabary is due only to the lack of illustrations, and once such illustrations — attested spellings — are found, predictably similar modified values can be derived from the basic values, with the application of the rule

$$V \longrightarrow \begin{bmatrix} a \text{ / accusative} \\ i \text{ / genitive} \\ u \text{ / nominative} \end{bmatrix}$$

36. Modifications of values are marked in column 4 of the Syllabar under the sigla Ca — Cg, vowel modifications marked by the sigla Cb and Cf, in reference to derivation of one value from another (basic) value; note, however, that these derivations are context-free, not context-sensitive.

Such a rule does not actually provide any new information, but only makes explicit the information about the syntactic function of the word which is gained from outside the writing system.

(2) *Vowel suppletion*, in the context of other suffixes. There is a rule of the form (C)V_1C $\longrightarrow$ (C)V_1CV_1, i.e., to the final consonant of a CVC value, add a vowel identical to the vowel preceding the consonant, in the grammatical environment where a vocalic suffix is required. This rule is applied to (C)*u*C values when the context requires (a) the masc. pl. suffix *u* (e.g., *im-qutu* written *im-qut*), (b) the subjunctive suffix *u* (e.g., *i-man-guru* written *i-man-gur*); to (C)*i*C values when the context requires the 2.fem.sg. ending *i* (e.g., *šum-dili* written *šum-dil*), and to (C)*a*C values when the context requires the 2.pl. or the ventive ending *a* (e.g., *il-laka* written *il-lak*).[37] Just as for the cited rule (1) in the context of case endings, in the context of other inflectional suffixes too the morphological environment of a basic value determines its ultimate value, with the application of the rule

$$\emptyset \longrightarrow V \begin{bmatrix} a \ / \ \{2.pl., vent.\} \\ i \ / \ 2.fem.sg. \\ u \ / \ \{3.pl.masc., subj.\} \end{bmatrix}$$

i.e., add the vowel *a*, *i*, *u* when the inflected form is a 2.pl. or ventive, a 2.fem.sg., and a 3.masc.pl. or subjunctive respectively.

B. RULES AFFECTING DISCONTINUOUS MORPHEMES. (1) *Vowel replacement.* We may take as example two rules:
(a) $a \longrightarrow u \ / \ b\_l$, and
(b) $u \longrightarrow a \ / \ b\_l$;
Rule (a) effects the replacement of *a* by *u* between *b* and *l* (i.e., in the basic value *bal*, no. 8), and rule (b) the replacement of *u* by *a* between *b* and *l* (i.e., in the basic value *bul*, no. 10); these rules are counterparts

37. Matouš points out (ArOr 36 341) that of the CVCV-values postulated for Neo-Assyrian by Deller, von Soden in the Syllabar has not accepted those which are required only as case-ending, "da im Neuassyrischen Auslautvokale schon teilweise abgefallen sind"; this means that in Neo-Assyrian — and in other dialects in which final vowels of the case-inflection are dropped — the environments nominative, accusative, genitive are omitted from the context-sensitive reading rule. Indeed, in Neo-Assyrian in a great number of words (those which have a short /a/ before a single final consonant) the vowel-oppositions of the case-inflection regress to the last vowel of the stem, transforming the Neo-Assyrian case system into an Umlaut-system with such oppositions as *pagul* (nom.) vs. *pagal* (acc.) vs. *pagil* (gen.), similarly *karput* vs. *karpat* vs. *karpit*, a development similar to Modern German *Vogel* 'bird' (sg.) vs. *Vögel* (pl., formerly *Vogeli*), *Hafen* 'mug' vs. *Häfen*, in words ending in *l* and *n*, while the plural ending is retained in *Blatt* 'leaf' vs. *Blätter* (formerly *Blatir*); or to English *goose* [gu:s] vs. *geese* [gi:s].

of one another and in effect neutralize the distinction of the two basic values. The apparent neutralization does not take place, however, because the two rules are sensitive to different contexts: (b) obtains in the string *a-ta-búl-kàt*, and (a) in the string *lu-bal-ta*. The morphological environment of (b) is 'perfect tense of the root *blkt*' or 'discontinuous morpheme *a-a-a*' which requires the modification of the string into *a-ta-bál-kàt*; the morphological environment of (a) is 'stem (lexical item) *lubult-*' or 'discontinuous morpheme *u-u*' which requires the modification of the string into *lu-bùl-ta*.

A further example for a vowel replacement rule is:
(c) $i \rightarrow u$ / *b_š*, which obtains in the string *lab-biš* in the morphological environment 'stative of stem D' which requires the modification of the string into *lab-buš*.

For other values derived from ultimate values by vowel replacement the illustrations are so few that it is questionable whether rules could be formulated for them. The basic value *rad* also appears in the strings *i-rad-da*, *aṭ-rad-su*, *i-nar-rad*, and *ú-ga-rad*. The first three strings are equivalent to morphological *irudda*, *aṭrudsu*, and *inarrud*, i.e., the vowel in *rad* is to be replaced by *u*; the last is equivalent to (the proper name) Ugarit, i.e., the vowel in *rad* is to be replaced by *i*. For such sporadic occurrences, a basic value with unspecified vowel would be a sufficient notation, e.g., *r*V*d*.

(2) *Vowel suppletion.* The rule is of the form $(C)V_1C \rightarrow (C)V_1CV_1$ (the same as above under A 2). This rule obtains in basic values of the shape C*a*C or C*e*C in the grammatical environment 'present tense', or 'discontinuous morpheme *a-a*, respectively *e-e*'.

The present tense of the strong verb has in its second syllable the configuration /C*a*C_1C_1/ (e.g., ta*rakk*as, graphemically <C*a*(C_1)-C_1V> (e.g., <*ra*(*k*)-*ka*>); in the entire string: <(C)V-C*a*(C_1)-C_1V(C)> (e.g., *ta-ra*(*k*)-*ka*(*s*)). If this string is written <(C)VC-C_1V(C)> (*tar-kas*), the characteristic configuration *–rakk–* is replaced by a configuration <CC_1>*–rk–*; the latter is either to be interpreted as a preterit for which this configuration is characteristic (e.g., *taṣbat*), or, if the vowel appearing after such a consonant cluster excludes the interpretation as preterit, it will yield an ungrammatical form (*tarkas* and not the correct *tarkus*). To exclude, on the one hand, an interpretation as preterit, and on the other, to correct the ungrammatical form, the vowel-suppletion rule is introduced for the first syllabic sign. The resulting (C)VCV (e.g., *tara*) value yields the graphic sequence <(C)VCV-C_1V> *tara-ka*(*s*) equivalent to <(C)V-CVC_1-C_1V> (*ta-rak-ka-*), i.e., the non-occurrence of a cluster may be substituted for the graphic sequence <CC_1> (*rk*), i.e., the occurrence of a phonologically impermissible cluster CCC (CC_1C_1) (*rkk*) that seems to have resulted, judging from the purely graphic configura-

tion, from the configuration CaCC. A common example is */tarakkas/* written <*tar-kas*>, rewritten as *tara-kas*; the following values are thus augmented in the Syllabar:[38]

ultimate value	expanded value	ultimate value	expanded value
bal 8	*bala*	*rim* 244, 280	*reme*
il 12	*ila*[39]	*an* 12	*ana*
kil 280	*kele*	*ak* 70	*aka*
tal 61, 223	*tala*	*aš* 192	*aša*
mir 198	*mere*	*tap* 90	*tapa*
par 48, 151, 221	*para*	*taḫ* 119	*taḫa*
tar 11, 72	*tara*		
ḫar 238	*ḫara*		
tan 173	*tana*		

(the last four, *ak*, *aš*, *tap*, *taḫ*, are given for Neo-Assyrian only).

The ultimate values discussed under 'morphological analysis' are all based on the fact that there exists grammatical information that enables the reader to "correct" the written string. This correction is performed on an ill-formed string, for pedagogical reasons; this absence of well-formedness can, however, be of two kinds: the string is strange or unfamiliar, and the string contrasts with another string. For example, in B 1a and b, the basic value would yield the strings *attabulkat* and *lubalta*; while both are 'strange', they do not contrast with (i.e., cannot convey a different meaning from) the "corrected" strings *attabalkat* and *lubulta*, and thus the corrected reading contains no information that the uncorrected reading does not also contain. On the other hand, in the example under B 1c, the basic value would yield the string *labbiš*, which contrasts, as a minimal pair, with the corrected reading *labbuš*, the former representing the imperative (discontinuous morpheme *a-i*), the latter the stative (discontinuous morpheme *a-u*). The correction is based on the identification of the morphological category appropriate in the larger context, the sentence, since without such a larger context either string is grammatical. But given this larger context, it is immediately evident that not the imperative but the stative is syntactically required. If this were not evident from the sentence, the basic value *biš* of the given sign would not have been suspect and replaced, in this particular context only, by the value *buš*.

38. In the Syllabar, these values appear after the basic value. In my examples, I omit diacritic signs to avoid confusion but, in order to facilitate identification of the signs in question, I add their serial number.

39. This is the only example for the expansion $(C)V_1C \rightarrow (C)V_1CV_2$ in the environment 'present tense' (due to the fact that V_1 happens to be a 3. actor prefix and thus not of the same quality as the vowel of the present tense vocalic morpheme *a-a*).

In conclusion, we can say that the context-sensitive rules, i.e., non-automatic partial replacement rules, are necessary only for and arise only from the morphological analysis of a larger text segment. In the syllabary, the modified values gained from such analysis have only the function of illustrating spelling practices, and thus merely repeat information that the dictionary is supposed to, and does indeed provide.

III. MORPHOPHONEMICS

The four steps described above are a necessary and sufficient condition for reading cuneiform texts; their result is a morphophonemic transcription. This term 'morphophonemic transcription' partially overlaps with von Soden's term 'erläuternde Umschrift' (Syllabar p. xvii), also formulated as 'der Etymologie entsprechend' (p. xviii), which he contrasts with 'lautgetreue Umschrift' (p. xvii) or 'Standardaussprache [als] Grundlage der Umschrift' (p. xviii), which I would compare to the term phonemic transcription, using the terms morphophonemic and phonemic as they are employed in contemporary linguistic usage. The two types of transcription should provide different kinds and amounts of information about the language they are applied to; by distinguishing them and using these terms in a strictly defined linguistic sense, we may achieve a more explicit understanding of their range and usefulness.[40] The second type (phonemic transcription) will need a reexamination of questions pertaining to phonetics and phonology, and will be taken up on pp. 39ff. Here I would like to take up first the definition of 'morphophonemic', a term from which most Assyriologists shy away because it is unfamiliar, and by way of explanation give, as before, a single form as illustration.

Let us take four commonly occurring spellings for the same word: (1) <*it-ta-di-šu*>, (2) <*it-ta-di-iš-šu*>, (3) <*it-ta-di-in-šu*>, (4) <*it-ta-din-šu*>.[41] The four spellings can be divided into two groups: one in which there is a basic value containing <*n*> (3 and 4), and another in which there is no such value (1 and 2). In both groups there is one spelling (2 and 3) containing more graphemes (maximal spelling), and one (1 and 4) containing fewer (minimal spelling).

Let us assume that, on the basis of context and various other factors, we have established this word as 3.sg. perfect of the verb root {*ndn*},

40. In von Soden's use of his terms the borders between the two categories are somewhat blurred, both in his statement in the introduction and in the actual choice of values given to individual signs.

41. Notation: < > enclose graphemes; ∫ ∫ enclose connected transcription; { } enclose morphemes; / / enclose phonemic transcription; || (vertical bars) enclose morphophonemic transcription.

infinitive ⌈*nadānu*⌋ 'to give', with 3.sg.masc. object suffix {*šu*} 'him'. On the basis of other 3.sg. perfect forms with suffix {*šu*}, e.g., ⌈*iktaširšu*⌋ (written, e.g., <*ik-ta-ši-ir-šu*>) of the verb root {*kšr*}, infinitive ⌈*kašāru*⌋ 'to repair' (which may be more explicitly noted, with morpheme boundary markers, as {*i*+*K*+*ta*+*ŠiR*+*šu*}), we expect a form *⌈*intadinšu*⌋ corresponding to the infinitive ⌈*nadānu*⌋ on the model of ⌈*iktaširšu*⌋ : ⌈*kašāru*⌋. This expected form *⌈*intadinšu*⌋ contains the full information about (a) the verb root {*ndn*}, also called the three radicals, which are realized in the infinitive as ⌈*nadānu*⌋, (b) the vowel of the verb stem, *i*, which appears between the second and third radicals, (c) the prefix morpheme of 3.sg., {*i*}, (d) the infix morpheme of the perfect, {*ta*}, and (e) the 3.sg.masc. suffix {*šu*}; in explicit notation, with morpheme boundary markers, |*i*+*N*+*ta*+*DiN*+*šu*|. Such a string of would-be phonemes, which contains the full information about its morphological structure, is called *morphophonemic*, and its would-be phonemes are called *morphophonemes*.

However, none of the spellings cited for this word exhibits this expected form. If we convert the first group, spellings (1) and (2), into connected transcription,[42] then in order not to lose relevant information, we must choose the maximally informative maximal spelling (2), in connected transcription ⌈*ittadiššu*⌋, in contrast to ⌈*ittadišu*⌋, as the underlying form to represent both spellings (1) and (2). From this maximal spelling, the minimal spelling (1) is obtainable by the simplest possible operation: deletion, while the converse procedure, i.e., deriving the maximal from the minimal spelling by insertion, requires more involved operations.

If we convert spellings (3) and (4) into connected transcription, both will yield identical information, namely ⌈*ittadinšu*⌋, although in terms of number of graphemes (3) represents the maximal and (4) the minimal spelling.

Of the two forms, ⌈*ittadiššu*⌋ can be derived from ⌈*ittadinšu*⌋ by a simple (assimilation) rule (see below), but not the converse. This derivable form ⌈*ittadiššu*⌋ does not show the morpheme structure shown by |*intadinšu*| but probably represents the pronunciation in some fashion.[43] This form we shall call *phonemic*. The form ⌈*ittadinšu*⌋, as compared to the morphophonemic form *|*intadinšu*| and to the phonemic form */ittadiššu/*, is partially phonemic, partially morphophonemic.

Referring to our model <*ik-ta-ši-ir-šu*>, in connected transcription ⌈*iktaširšu*⌋ (with optional morpheme boundary markers {*i*+*K*+*ta*+*ŠiR*+*šu*}), no rules are needed to derive phonemic */iktaširšu/* from

42. By some such rules as: (1) Delete hyphens, and (2) Delete one of two consecutive identical vowels, see above, pp. 11f.

43. E.g., the 'Standardaussprache' referred to by von Soden, Syllabar p. xviii.

morphophonemic |*iktaširšu*| because the connected transcription reflects both the morphological structure and its phonological realization, and thus the morphophonemic and the phonemic forms coincide. In cases where the two forms do not coincide, it is the morphophonemic form that contains the larger amount of information, for reasons that may be briefly stated as follows.

The phonemic spelling transmits information referring to phonemes, i.e., how the utterance is realized by means of distinctive units, while the morphophonemic spelling transmits information referring to morphemes, i.e., how the utterance is realized by means of meaningful units (stems and affixes). The meaningful units, which consist of morphophonemes (also called systematic phonemes), are ultimately realized on the surface through phonemes. Consequently, the morphophonemic spelling necessarily contains the information concerning the phonemic identity of the morpheme shape, which can be obtained by the application of assimilation or other rules. In phonemic spelling, on the other hand, there remains some uncertainty (if allomorphs exist) concerning the morpheme structure of the utterance, as in the case of the cited form */ittadiššu/*.

Applying the above to the two transcriptions of our example, it is evident that the form |*ittadinšu*/ (i.e., the partially morphophonemic, partially phonemic form) contains information concerning the phonemic shape of the word, on the basis of the rule $n \rightarrow š/ \_\_+š$, i.e., that *n* at morpheme boundary assimilates to a following *š*.[44]

This partially morphophonemic string still does not contain full information on the verb root, since the root is {*ndn*}, and the first *n* is not apparent in this string. There is, however, a certain predictability as to the first radical, once the string is analyzed, like the model |*iktaširšu*|, into ∫*it*+*ta*+*din*+*šu*∫, since there are rules to find out what morphophonemes may result in a particular phoneme.

The *t* which appears before the infix {*ta*} represents a radical which is either *t* itself, or a consonant assimilated to the initial *t* of the infix. The assimilation to *t* neutralizes the opposition among two phonemes, *t* and *n*, and two morphophonemes, a *u*-coloring one or *w*, and a geminating one or consonantal length (:). If no other assimilation rules apply at morpheme boundary before {*ta*}, the string ∫*ittadinšu*∫ is analyzed as {*i*+C+*ta*+*DiN*+*šu*}, i.e., verb root CDN, where C=*t*, *n*, or either morphophoneme, *w* or :.

If the neutralized radical is *t*, the radical has no alternant ("alternates" with itself) when followed by the infix {*ta*}, e.g., *it-ta-kal*, (preterite) *i-tkal*, (infinitive) *takālu* (the model being *ik-ta-šir*, *i-kšir*, *kašāru*).

44. A more precise partially morphophonemic notation would thus be |*ittadin*+*šu*/, since a boundary condition is necessary for this assimilation (see LAA 6.1.2.2).

If the radical is *n*, *it-ta-din* corresponds to a preterite (with another assimilation) *iddin*, and an infinitive *nadānu*. The *u*-coloring morphophoneme radical is elicited from the correspondences *it-ta-šab*, *u*:-*šib*, (*w*)*ašābu*; verbs with a *u*-coloring radical are called historically "*primae w*" verbs. The geminating morphophoneme radical is elicited from the correspondences *it-ta-lak*, *illik*, *alāku* (morphophonemically *:alāku*).[45]

Since the dictionary lists no *wadānu*, *:adānu*, nor (unless the Assyrian dialect is taken into consideration) *tadānu*, but lists an infinitive *nadānu*, on the basis of the model *kašāru* : *iktašir* = *nadānu* : *x*, the morphophonemic form |*intadinšu*| will be selected, from which the morphophonemic (assimilation) rules predict the phonemic string /*ittadiššu*/. This form is actually attested in spelling (2), the maximal spelling.

Just as the mentioned phonological conversion rules, applied to an expected but not actually attested morphophonemic spelling, convert the morphophonemic string into a phonemic string, the same or similar rules will convert into the same phonemic string other, partially morphophonemic spellings, and even so-called hypercorrect spellings, which do not correspond to the morphemic analysis. Such an instance is given for the word under discussion by the spelling <*it-ta-dim-šu*>.[46]

Converting this spelling into connected transcription, i.e., ⌠*ittadimšu*⌡, and applying the same model as before (analysis into 3.sg. actor prefix {*i*}, perfect infix {*ta*}, etc.), and applying the rule, we obtain *it-ta-dim-šu*. Proceeding as before, we obtain the morphological information 'root CDM', where C = {*t*, *n*, *w*, *:*}. Such a root not being found in the dictionary (roots with initial : are to be looked up under initial ʾ [aleph] in AHw.), we search for a further applicable rule. Such a rule is $m \longrightarrow n/\_\_+š$ (formulated in LAA 6.2.1.4 as *m* > *n* before dental and *š*); applying it, we obtain the string |*ittadinšu*/, i.e., the same string that was yielded in the original example by spellings (3) and (4). Since there exist rules applicable to *m* before *š* for converting morphophonemic spellings into phonemic spellings for other items of the lexicon as well, there is no need to create new rules for the case of the spelling with the sign <*dim*>. We may thus retain in our (graphemic and connected) transcription the value <*dim*> and choose rules of more generality. We apply the pertinent phonological rules to <*it-ta-dim-šu*>, which by application of the $m \longrightarrow n/\_\_+š$ rule yields the partially morphophonemic string |*ittadinšu*/ and by application of the second rule yields the

45. Only one other verb apart from *:alāku* shows this alternation, but its infinitive conforms to a different pattern, namely *izuzzu*. In other verbs which are similar to *:alāku* in that they lack an initial consonant other than the morphophoneme length, the initial morphophoneme is not consonantal length but vowel length, i.e., brings about vowel lengthening, e.g., *i*:-*ta-mar*, *i*:-*mur*, *amāru* (morphophonemically *:amāru*).

46. The only attestation, according to von Soden's Syllabar, comes from a peripheral region, Boghazköy, but serves our purpose here as an illustration.

phonemic string */ittadiššu/*, i.e., the same phonemic string that was yielded by the more frequent spellings (1) — (4). When von Soden's Syllabar creates for this particular spelling a value — isolated and clearly created for this particular case — <*di*> for the sign <*dim*>, and transgraphemicizes <*it-ta-di-šu*>, it actually selects of the two possibilities the "lautgetreue Umschrift," i.e., phonemic transcription, rather than the ideal "erläuternde Umschrift," i.e., what we call morphophonemic spelling. The choice of the value <*di*>, once it is substituted for <*dim*>, causes an irretrievable loss of information, while retaining the value <*dim*> conserves more information.

The partially morphophonemic spelling with the sign <*dim*> illustrates one case of graphemic neutralization, in the case of <*it-ta-dim-šu*> that of */m/* and */n/* before a suffix with initial */š/*. The same neutralization in a different environment may now be illustrated on another inflected form of the verb *∫nadānu∫*. The 3.sg. present tense of this verb, earlier both morphophonemic and phonemic */inaddin/*, and later morphophonemic *|inaddin|* and phonemic */inandin/*, with dissimilation of voiced long stops and sibilants, is written, e.g., <*i-na-ad-di-in*>, <*i-na-ad-din*>, <*i-na-an-din*>, but frequently also <*i-nam-din*>. The first two spellings are morphophonemic, and the last two phonemic. These last two differ only in the choice of the grapheme before */d/*. Since the dissimilation of a long voiced stop and sibilant yields the homorganic nasal plus stop (*b*: ⟶ *mb*, *d*: ⟶ *nd*, *g*: ⟶ *ng* — presumably [*ŋg*] —, and *z*: ⟶ *nz*, see LAA 6.2.1.4.5), the articulation point of the nasal is predictable from the articulation point of the following stop or sibilant, and there is no need for the writing to specify one of the two (phonemic) nasals, i.e., the labial nasal */m/* and the dental nasal */n/* are again neutralized in this environment. It is not necessary to give the sign NAM a value <*nan*>, and still less to introduce a (non-phonemic) 'palatal nasal *n*' (GAG § 32, Syllabar p. xxii).[47]

47. The choice of the term "palatal nasal" which normally denotes ɲ (as in Spanish ñ) is unfortunate since clearly von Soden means some other sound, presumably [ŋ]. A similar observation may be made regarding the sign SAG. An ultimate value that must be listed for it is *san* (as it is in Syllabar no. 87). This value, however, is restricted to occurrences under two conditions jointly applying: (1) in Sumerian words or loanwords, and (2) before velars (*g*, *k*; *q* is not included because *q* does not occur in Sumerian). The selection is very much reminiscent of Greek spelling: *gg* for *ng* (i.e., phonetic [ŋg]), *gk* for *ŋk*, and *gχ* for *ŋχ* as in *ἐγγύς*, *ἄγκυρα* and *ἔγχος* (examples from Allen, Vox Graeca, p. 32). The Akkadian spelling similarly probably expressed [ŋg] borrowed from Sumerian [ŋ] as well as from Sumerian [ŋg] (in compounds such as *saŋ-gar*). If this spelling rule is phrased with reference to SAG (i.e., *sag* ⟶ *saŋ*/ __*g*, if non-Akkadian), it is automatically extended to ŠAG, namely after application of the otherwise necessary rule *s* ⟶ *š* /__VC to the SAG-sign; we may therefore transfer the examples SAG-*gi-lu* and SAG-*ga-maḫ-ḫu*, cited under no. 87 under, respectively, the values *sag* and *šag*, to the value *san*, respectively an added value *šan*. Indeed,

I have tried to show, using one example presented in detailed fashion, how not every seemingly normalizing — and even normative — reading rule preserves information about the morphophonemic form which, I agree with von Soden, gives the maximal amount of information to the student of Akkadian, to the linguist, the comparativist, or historian. It is therefore indeed desirable to keep values closer to the morphophonemic than to the phonemic level, and make the accompanying morphophonemic rules explicit.[48]

That the morphophonemic principle in the interpretation of the spelling cannot be carried out without at the same time providing morphophonemic rules may be seen from some examples which illustrate that once the conversion from morphophonemic form to phonemic form has been applied by the ancient scribe, it is only with reluctance that the phonologically oriented modern scholar attributes morphophonemic values to a sign. Take the spelling <*i-ru-um-ma*>, to be converted into connected transcription as *ʃirummaʃ*. From alternations with spellings <*i-ru-ub-ma*> the morphophonemic form |*irub*+*ma*| may be established. This form is established, however, not by giving the sign *um* a value **ub*, as consistency would require, nor similarly elicited values, e.g., the value **ab* to the sign *am* in <*uš-ša-am-ma*> which equals morphophonemic |*ušša**b*+*ma*|, the value **ip* to the sign *im* in <*is-ki-im-ma*> which equals morphophonemic |*iskip*+*ma*|, etc. All these spellings may be rendered in the respective morphophonemic

---

according to the above generalizations, the values *san* (and *šan*) and *šag* are derivable from SAG and thus neither need to be printed in boldface. Since the value *san* (and *šan*) can be, however, formulated in terms of a more strict environment than *šag*, the distinction is perhaps quite properly expressed in the Syllabar by printing *šag* in boldface, but not *san*. The rule '→ŋ/__G, if non-Akkadian' can be extended to (a) signs ending in a nasal (*m*, *n*), such as ŠÀM (no. 125), ŠÁM (no. 127), etc., and (b) omitting the non-Akkadian criterion, to all signs ending in *g* after the OB period, thus describing the phonological rule of dissimilation (or nasalization) of voiced double velars, itself one case of the rule 'dissimilation (or nasalization) of long voiced consonants', discussed on p. 37. More generally, {*g*, *m*, *n*} → ŋ / __{*g*, *k*} which rule also accounts for Assyrian *ra-an-ga* (i.e., [*rāŋga*]) for */rāmka/*, which is written *ram-ka* or *ra-am-ka*.

48. It holds for other writing systems too that morphophonemic information is necessary to pronounce a written form correctly. For instance, English <askance> is not pronounced ['a:skəns] as if it were related to *ask* as <sufferance> [s'ʌfrəns] is to *suffer*, but [əsk'a:ns], similarly to <awry> [ər'ai], because the morpheme boundary is after *a* and not after *ask* as in real derivatives of *ask*, e.g. *asking*. In Hungarian, "sometimes grapheme sequences which appear to constitute digraphs, single or geminate (e.g. *zs* [ž], *ggy* [ďď]), belong to separate morphemes and must be pronounced accordingly (underlining indicates the element which is a part of the separate morpheme), e.g., *község* 'village, community', from *köz* 'general' + *ség* (abstract suffix); . . . *meggyón* 'confess', from *meg-* (converb for completion) + *gyón* 'he is confessing'." See John Lotz, "The Conversion of Script to Speech as Exemplified in Hungarian," The Linguistic Reporter, Volume 1 Number 5 Supplement 23 (October 1969), pp. 20f.

spellings plus the rule $\{b, p\} \rightarrow m/\_\_+m$ (formulated in LAA 6.1.1.1 as $C+m=/m:/$, where $C=p, b, n$). Both from the fact that no rule exists to be applied to $m+m$ and that no other rule results in /*mm*/, in the case of a surface string containing /*mm*/ the question must be asked — and it is the only question that is to be asked — whether /*mm*/ represents morphophonemic $|mm|$ or morphophonemic $|\{p, b, n\}+m|$. The answer is usually given by the dictionary.

IV. SYLLABARY, TRANSCRIPTION, AND PHONETICS

The text segment, when the procedure described on pp. 7ff. is applied to it, will appear partly in morphophonemic, partly in phonemic transcription, as these terms were defined on pp. 33ff. While the morphophonemic transcription is easily convertible into a phonemic transcription, if desired, no provision was made for the converse case, to provide an algorithm (step by step instructions) for the actual pronunciation. The question arises whether we can go beyond a phonemic transcription by assigning to the signs ultimate values that reflect the pronunciation, and have the syllabary also serve as a pronunciation guide; in von Soden's formulation: "Wieweit können die Lautwerte der Silbenzeichen richtig bestimmt werden?" (Syllabar, p. xvii–xxvi). I believe it is appropriate to stress here the only basically true answer to this question: to the extent that the writing system permits.

This answer has to be given on methodological grounds: The information that is not inherent in the writing system cannot be retrieved from it, since only by introducing information from outside the system can the information content of a given system increase. In other words, any information necessary to read a cuneiform text that is not contained in the writing must be supplied from outside sources. This was shown in connection with the reading process described on pp. 7ff., where the outside sources were represented by the grammar and the dictionary. These were found to provide information about the morphophonemes, and to some extent about the phonemes of the language. This state of affairs represents nothing unusual, and applies in fact to any writing system, whether it is used for a dead language or a living one. There exists no writing system, apart from some artificially devised ones, whose main purpose would be to approximate the spoken word. In fact, while most writing systems are phonemic, that is, they differentiate in the spelling the speech sounds that carry a difference in meaning as, in English, the initial consonants in *b*ore, *p*ore, *m*ore, they do not normally make purely phonetic, non-phonemic distinctions as, in English, between the unaspirated [p] of spot and the aspirated [p^c] of pot, both being spelled *p*. Therefore, the assumption that spelling reflects subphonemic

features of the spoken language is unnecessary. Neither is the assumption necessary or correct that difference in spelling represents difference in pronunciation, and that fluctuation of spelling indicates an attempt to reproduce a sound which is not equivalent to either of the two — or more — between which the spelling fluctuates,[49] since the tacit premise underlying this assumption too is the belief that orthography strives at all times to be phonetic, i.e., to approximate the spoken word. In fact, it is well known and can easily be illustrated that difference in spelling does not necessarily reflect difference in pronunciation, either on the phonemic or on the phonetic level. For example, a distinction made in the spelling may express a non-phonetic *and* non-phonemic difference, such as a morphological or lexical difference; the distinction may have been introduced to resolve an ambiguity, or it may preserve a former difference in pronunciation. Such is the case, e.g., in French *a* 'he has' and *à* 'to', Italian *da* 'from' and *dà* 'gives',[50] French *ou* 'or' and *où* 'where', *votre* 'your' and *vôtre* 'yours' on the one hand, where the diacritic mark reflects neither a different vowel nor a historical spelling, and on the other hand French *eut* and *eût* which preserve the distinction of the historical forms *habuit* and *habuisset*, and of the earlier spellings and pronunciations *eut* and *eust* respectively.

Just as there is no necessary correlation between spelling differences and phonetic or phonemic differences, the converse is also true, namely identical spellings may correspond to different pronunciations; there is thus no biunique relationship between spelling and pronunciation. To take an English example, the two morphemes *bow*$_1$ 'to bend' and *bow*$_2$ 'weapon used to shoot arrows' are spelled alike but not pronounced alike: bow$_1$ is pronounced [bau] and bow$_2$ is pronounced [bou]. It so happens in this case that the antecedents of these two spelled-alike morphemes vary with a variation corresponding to their pronunciation (covariation); this is equivalent to saying that the two pronunciations reflect two different etymologies: in this case *bow*$_1$ can be translated by German *beugen* and *bow*$_2$ by German *Bogen*, and earlier spellings of the two were *beuw* and *bow* respectively. The coalescence of the two spellings, in spite of their different etymology and pronunciation, is a historical accident. The difference in pronunciation between the two English homographs cannot be deduced from the orthography, since there are no spelling variants, only from information extraneous to the writing system, such as comparative or historical evidence. In another instance, the spelling *either* is realized in one of two pronunciations: [i:ðer] or

49. Excepting the rendering of foreign words for which no orthographic tradition exists.

50. Ambiguous, e.g., in a historical grammar of Italian: "*t* dà *d*" '*t* gives *d*' and "*t* da *d*" '*t* from *d*', the two sentences describing the opposite sound-law.

[ayðer], depending on the speaker's dialect, without the pronunciations reflecting a difference either in meaning or in etymology.

The converse case is also attested: etymologically connected items, if their pronunciation is not alike, do not necessarily retain in their spelling an indication of their etymological connection, but may be spelled in a way which does not reflect their relationship. Thus, German *Mass* and *messen*, though etymologically related, are spelled phonologically, with *a* and *e* respectively, although other similarly related pairs preserve the orthographical etymological relationship by using for the same two vowels the morphophonemic spelling with the graphs *a* and *ä*, as in *Mast* vs. *mästen*, etc.

To restore an etymological spelling, e.g. *beuw* and *bow* in the case of homographs, and the etymological spelling, e.g. *mässen*, in the case of differentiated spelling for etymologically related words would introduce into the transcription matters not relevant to phonology and phonetics, and in fact often obscuring these. A syllabary for reading cuneiform texts that included values based on the knowledge of language history, without indicating what constraints apply, could be compared to a grammar that would give illustrations for the pronunciation of all Germanic languages, old and modern, simultaneously.[51]

Furthermore, illustrations can be found for the lack of covariation in spelling in paradigmatically related groups of items. Such is the group to which the well-known English example *read* belongs. Just as for the spelling *bow* we had to posit two homographs bow_1 and bow_2, we have to posit two homographs $read_1$ (present tense) and $read_2$ (past tense). When comparing this pair with the pairs $beat_1$ (present tense) and $beat_2$ (past tense) on the one hand, and $spread_1$ (present tense) and $spread_2$ (past tense) on the other, we would assume that there is a similar correlation in pronunciation, and that $beat_1$ is pronounced [bi:t] and $beat_2$ [bet], $spread_1$ [spri:d], and $spread_2$ [spred]. We may even attempt to correlate the different quality of the vowel with the final consonant *t* in the past tense, as in the model *dream* [dri:m] (present tense), *dreamt* [dremt] (past tense) in which *t* causes a change of *i:* to *e*, contrasted

51. In fact, since modern Assyriologists are knowledgeable not only about all dialects of Akkadian, but also about its historical grammar and its relationship to members of the same language family, they incorporate their knowledge into the compilation of the syllabary. Taking the written string *atta unsar θu in himinam weihnai namo θein*, we might give a value */d/* to the grapheme *θ*, based on the correspondence *θu* = *du*, and a value */e/* before */r/* to the grapheme *a* in *unsar* compared with *unser*, etc., if we know to what Modern German words this string corresponds; or alternately, we might give the pronunciation [*i:*] to the grapheme cluster *ei*, and the pronunciation [*θ*] to initial */d/* in Modern German *dein* compared with *θein* [*θi:n*], etc., if we know the history of Germanic Languages and comparative Indo-European. For any particular period of Akkadian, spelling variations can be listed simply as such, since the difference they may have corresponded to in some previous dialect has obviously been neutralized.

with *dreamed* [dri:md] (past tense), in which the final consonant, *d*, causes no change of *i*:. The ideal spelling that would express the morphological tense difference for the cited verbs would be: present tense *read*, *bead*, *spread*, past tense *reat*, *beat*, *spreat*. Since, however, there is a difference in pronunciation in the two tenses of *read*, but none in the two tenses of *beat* and *spread*, the spelling that would reflect the phonological identity of the vowel in the present stem, its phonological difference in the past stem, and the morphophonemic identity of the past and present stems would be: present tense *read* [ri:d], *beat* [bi:t], *spreadt* [spred]; past tense *readt* [red], *beatd* [bi:t], *spreadtt* [spred].

Similarly, in Akkadian the spelling may fail to express a distinction obtaining in morphology, whether such is correlated or not with a distinction in pronunciation, or may reflect a distinction which applies not to the item so spelled but to its paradigmatic correlates. For example, the spelling *ú-še-rib* represents both the preterite *ú-še-rib* and the present *ú-še-reb*; *be-li-ia* both the singular *be-li-ia* and the plural *be-le-ia*; in both cases the substitution of *e* for *i* in our transcription carries a grammatical difference which *we* have chosen to express, not the writing system. On the other hand, of the two coexisting spellings *šu-pa-at* and *šu-ba-at* only the latter preserves for the eye[52] the morphological relationship with a group of words that all have *b* and not *p*, e.g., *(w)ašābu*, *(w)ašibu* (the latter forming a minimal pair with *(w)āšipu*); and leads us to construct a nominative *šubtu*, while the spelling *šu-pa-at* presupposes a nominative *šuptu*, that is, *šubtu* with automatic devoicing of *b* before voiceless stop, similarly to the devoicing reflected in the Latin spelling *actum* vs. *agere*, and reflects the pronunciation [šupat], with *p*, reinterpreted instead of etymological *b*, from forms where *b* and *t* are in contact position.

When the Assyriologist chooses an ultimate value that expresses the paradigmatic form and the etymology of a word, he has chosen to make a 'grammatical' transcription. That is to say, he has incorporated his morphological analysis (above pp. 12ff.) into his transcription, and hence provided a more directly informative transcription for other readers, for whom the text so presented will be more easily understandable. Still, we must not confuse ease of comprehension with some kind of methodological requirement for such a transcription. A person who wishes to use the typewriter — or the keypunch — instead of pen and ink to reproduce the cuneiform text but for whom the marks made by the typewriter are images of the cuneiform signs which he reads with equal ease, may choose, for reasons of simplicity, to select always the same value for a sign, and let us say uses alphabetic symbols instead of the

52. The eye of the reader of the Romanization, naturally, and not of that of cuneiform writing.

sign's serial number for mnemotechnic reasons only. To him, and to other trained cuneiformists, a text thus 'transliterated' will be equally comprehensible.[53]

While a 'grammatical', that is, morphophonemic transcription, although not methodologically required, is desirable, let us ask whether a transcription more closely approximating presumed actual pronunciation is desirable and/or feasible.

In aiming to arrive at the 'actual pronunciation' or 'approximation of the spoken word' we expect the spelling to give us information about, or clues to, more than one type of fact. One type would be the existence of sounds of the spoken language that have hitherto not been recognized, for example, the vowels [o], [ü], and the like, or the consonants [*h*], [*f*], etc. The other type would be the existence of a conditioned allophone or diaphone of a known phoneme, such as the spirantization of stops in intervocalic position, or the replacement of *s* by *š* before *t*.

I have purposely avoided the term 'phonetic' in the above description, since not all the phenomena referred to are covered by this term. When we describe allophonic variation, that is, variation that can never differentiate meaning, such as the distribution of unaspirated *vs.* aspirated *p* in English *spot* and *pot*, we describe in reality the phonemic system, namely the phoneme */p/*. Thus, the question is, rather, how exactly can the allophonic variations of phonemes be determined, and its corollary, how far should the allophonic variation be reflected in the transcription, i.e., in the romanized spelling of cuneiform texts.

However, as I have indicated above, p. 39, in no language is there a biunique relation between spelling and pronunciation, even in terms of phonemes. Furthermore, the fact that we deal with written languages makes even the establishment of the phonological system and the identification of phonemes an illusory pursuit. Just as phonemes can be established in a spoken language only through the distribution of allophones, so in a written language the units we endeavor to establish can be defined only in terms of the distribution of alternating spellings. These alternants, as all other parts of the language inventory, are reflected in the writing alone; since there is no possibility of eliciting any phonetic realization of the written signs, the most we can do is to state the alternations of the writing. This procedure will yield not the phonemes of the written language, but its morphophonemes, or rather its morphographemes. While a morphophoneme, as stated on p. 34, is the unit abstracted from the variants that give information about the morphological structure of a language, a morphographeme is the unit abstracted from the spelling variants in a written language. Any phonetic label

53. This is the point of view of Gelb, Memorandum on Transliteration and Transcription of Cuneiform (mimeographed: Chicago, 1948).

given to such spelling variations is, as labels generally are, a convenient short reference to the variations themselves, and thus remains on the level of name-giving and should not be confused with any kind of linguistic reality.

For example, let us assume that in certain words the initial syllable is written now with a sign whose first consonant has been otherwise established as /*t*/[54] and again with a sign whose initial has been otherwise established as /*š*/.[55] We can thus state (and in optimal cases also state the environmental conditions of) a variation *t* ~ *š* for a particular word. It is possible to give to this variation the name of the sound t̲ (i.e., [θ]); this symbol may be chosen in preference to say *č* or *ʃ* because (a) the sound [θ] is considered to be somehow 'halfway between' *t* and *š*, (it is in fact rather 'halfway between' *t* and *s*) and (b) cognate languages indeed do have a phoneme /θ/. The fact remains, however, that it can never be proved, or for that matter disproved, on such internal evidence alone, that Akkadian possessed a [θ];[56] the statement admits of no proof, nor can it be "falsified" in the logical sense, and thus its scientific value is nil.

Another example: in a writing system where it is apparent that otherwise identical strings end in different signs, a set of (written) suffixes can be isolated.[57] If these endings appear after strings whose last sign shows an alternation depending on the ending that follows it, a certain morphophonemic or morphographemic alternation of the string, i.e., the stem, can be established. However, no precise phonetic label can be given to the alternation other than the spelling alternation itself. In other words, it is impossible to tell in the cuneiform syllabic writing

54. Letters enclosed in slants are customarily representations of phonemes; in Akkadian, they are only abstract representations, as /*t*/ for something that appears to be a dental stop which is different from two other dental stops, customarily denoted as /*d*/ (voiced) and /*ṭ*/ (emphatic), see note 28.

55. This is in fact the case in Akkadian; a suggestion for handling these alternations is proposed below, pp. 48ff.

56. A similar orthographic alternation, between *t* and *s* (and correspondingly between *d* and *z*), in ancient Greek can, on the other hand, be interpreted because we know the historical antecedents that gave rise to the alternation: the iotation of *t* resulted in a sound for which no letter existed in the Greek alphabet, and this sound was written, dialectally, *tt* or *ss*, and pronounced [*čč*] or [*šš*] or something else, different from both *t* and *s*. The assumption of such pronunciation is based on the observation that in Greek *t* alternates orthographically with *tt* and *ss* where iotation is expected, and that the iotation of *t* may result in *čč* or *šš* in other Indo-European languages, and not on the orthographic alternation alone. See C.-J. N. Bailey, "The Pronunciation of *Zeta* in Ancient Greek," in Papers from the Fourth Regional Meeting Chicago Linguistic Society, ed. by B. J. Darden, et al., Department of Linguistics, University of Chicago, Chicago, 1968, pp. 177–93; id., Language 45 (1969), 642f.

57. This is usually a first step in deciphering an unknown script, e.g., Minoan Linear B.

system whether the end of a stem alternates phonologically or merely graphically. Examples have been given above for such morpheme boundary phenomena of partial or complete assimilation for a dental followed by a sibilant (*amat+su*, see note 15), a voiced consonant followed by a voiceless (*šub+tu*, p. 42), and partial or complete assimilation to a nasal (*ittadin+šu*, p. 35).

I would like to examine briefly a few of the phonetic labels given to spelling variations or alternations, first labels in the form of vowels, then labels in the form of consonants.[58]

1. *The vowel inventory of Akkadian*

When Akkadian was first deciphered, readings of cuneiform signs gained by combinatory methods and internal evidence (paradigmatics or syntagmatics) were adjusted to fit the deciphered Akkadian words' counterparts in other Semitic languages, as soon as it became apparent that Akkadian belonged to the Semitic language family. For the vowel inventory, therefore, only the vowels known in Classical Arabic, namely *a*, *i*, *u*, were accepted. In those early days, CV signs that are nowadays romanized C*i* and C*e* were both romanized as C*i*, and the vowel sign *e* was, similarly, romanized as *í* (i.e., *i* with an acute accent to distinguish it from another vowel sign *i*), in order to avoid introducing a fourth vowel into the Semitic triangular vowel system. Later developments, among them the discovery of Sumerian texts written in the same cuneiform script, but in a different language in which C*i* and C*e* syllables were shown to represent different morphemes, and comparison with Greek and Hebrew transcriptions had led to the admission of *e* first into the phonetic, then into the phonemic inventory of Akkadian.[59] There is still little clarity on this point, and the status of *e* is variously labeled phonetic or phonemic.[60] In fact, Edzard has shown that the opposition, if any, in Akkadian is not between two vowels *e* and *i*, but "Akkadisch stehen *a/e — e/i* zueinander in Opposition, wobei *e* im ersten Glied etymologisch *a*, *e* im zweiten Glied als Phonem = *i* ist."[61]

58. These labels as a rule do not appear in values of the Syllabar, nor in the keywords of AHw, but in the latter appear in connected transcriptions of words, see pp. 47f.

59. Paul Haupt, "The Assyrian E-Vowel," American Journal of Philology VIII/3 (1887), 265–91.

60. For the phonetic interpretation, note "letzteres (i.e., das aus *a* oder *i* sekundär entwickelte *e*) hat nie Phonemcharakter" GAG § 8b; *e* is considered a phoneme by Gelb, BiOr 12 (1955) 97 ad GAG § 8b. Note however that Deller, Lautlehre des Neuassyrischen, § 25g, does not consider the endings of the Assyrian subjunctive *-ūni* and ventive *-ūne* a minimal pair: "Ob der Subjunktiv wirklich *-ūni* (und damit vom Ventiv *-ūne* deutlich abgehoben) ausgesprochen wurde, soll mit dieser (zunächst nur graphischen) Distinktion [namely, *-u-ni* vs. *-u-né*] jedoch nicht behauptet werden."

61. D. O. Edzard, ZA 53 (1959) 305.

It might be preferable to speak of the opposition between an orthographic and phonological alternation *a* vs. *non-a* (*a* vs. *ă*) and an orthographic variation *i* vs. *non-i* (*i* vs. *ĭ*), disregarding the vowel *u* which does not participate in such alternations.

The alternation *a* vs. *ă* refers to discontinuous vocalic morphemes, e.g., the alternation *a-i* vs. *ă-i*, the *a* of the latter morpheme written with either a *Ce* sign or a *Ci* sign, e.g., *ša-pi-ir*, (i.e., *šapir*), *na-ši* (i.e., *naši*) vs. *še-bi-ir* and *ši-bi-ir* (i.e., *šăbir*), *ne-si* and *ni-si* (i.e., *năsi*), all of which belong to the morphological category represented by *paris*.

The correspondence of the two vowels *a* and *non-a* in the first syllable is usually referred to in historical terms, namely as the change of *a* to *non-a* in a particular environment (e.g., in the same morpheme with a laryngeal). For the resulting vowel, the label *e* is chosen, on grounds which are not orthographic (for instance, comparison with other Semitic languages, cf. Arabic *imāla*). The label *i* has not been chosen for it because it could always be misinterpreted for morphological *i*, such as the second vowel in the above quoted examples. The label expresses the fact that there had been present a particular environment which occasioned the change; the environment had disappeared, and only the change, i.e., the writing with *non-a*, remained as a testimony of a former environment. Instead of the label *e* — a label available due to the existence in the orthographic system of certain signs elicited from Sumerian — some other label could be chosen such as a_i, to be read as '*a* appearing as *non-a* and *non-u*' or, to reflect not the change but the environment, following Indo-European practice, *aH* (see note 35), to be read as '*a* accompanied by a feature *H* in one of the radicals of the root', and this feature may be assigned morphophonemically to one of the radicals.

The orthographic variation *i* vs. *ĭ* refers to the superficially similar case of orthographic variation affecting the *i* of a discontinuous vocalic morpheme, such as in the second syllable of the above cited examples, which is written with a *Ce* or *Ci* sign, e.g., *pa-te-ir* and *pa-ti-il*. The orthographic correspondence between *i* and *non-i* is referred to in terms of allophonic change due to a synchronic environment (e.g., contact position with *r*). To use for this correspondence the same label, *e*, as for the other correspondence, leads to confusion of the terms of the opposition; it would indeed be preferable to stay with the label *i* of the early Assyriologists, but the best would be to leave the vowel marked as *i*, and disregard orthographic practices of particular language periods, because allophonic changes are as a rule not reflected by the orthographies. Orthographic opposition between *i* and *e* could be shown only on the basis of orthographic consistency or at least statistical frequency in such groups as *e-en*, *e-im*, *i-en*, *i-im*, or *te-i*, *ti-i*, *te-e*, *ti-e*; such proof has however yet to be adduced.

In fact, as the cited syllable combinations show, every CV or VC

sign in which V is neither *a* nor *u* has two readings (ultimate values): C*e* and C*i*, respectively *e*C and *i*C, as can be seen from the Syllabar under each C*e* and the corresponding C*i*, respectively *e*C and *i*C values, and if the inventory includes two such signs, their values are multiplied by two, that is, there will be four equivalent values distributed among two signs, e.g., *ni*, ni_x, *ne*, *né*, or *el*, el_x, *il*, il_5.

In order to preserve the morphological oppositions in our transcription, we must choose two different labels for the morphophonemic alternation *a*/*ā* and the variation *i*/*ī*. It is possible to mark the morphophoneme *a* — when it is realized phonologically as *i* — with the letter *e*, although it obviously is not necessary, since the writing dispenses with the distinction (as in the example *in-ni-im-mi-id* quoted by Edzard, for morphophonemic *innammid* or in the spelling *ú-ši-ri-ib* [a variation on more common *ú-še-ri-ib*], to be transcribed */uširrib/*, and morphophonemically |*ušaHrab*|), but it is not proper to use the same letter for the second member of the opposition, the morphophoneme *i*, for which a spelling with e.g. C*e* is, again according to Edzard loc. cit., equivalent to phonemic C*i*.

When it was found that morphemes in which the spelling with C*a* values would be expected in the same morphological environment were also spelled with C*i* values which could not be defined by the phonological environment, e.g., <*ša-ki-in*> in infinitives where <*ša-ka-an*> is expected, they have been interpreted as attempts to render spoken [ɛ] (also labeled as *ä*). This postulated vowel[62] is in fact morphophonemically identical with the one labeled a_i, etc., namely it labels the alternation *a*/*i* (or *a*/*e*), and thus is also identical with the *e* used in the transcriptions of such spellings as *in-né-em-mi-id* for *in-NI-im-mi-id*, etc., quoted by Edzard, loc. cit.

While there is a morphophonemic opposition between the alternation *a*/*e* and the variation *e*/*i*, no such opposition obtains for other vowels whose existence has also been assumed[63] such as *o*, deduced from variation in the spelling between *a* and *u*, and *ü*, from variation in the spelling between *i* and *u*. Correspondingly, while the vowel *e* has achieved full citizen's status in syllabary, grammar, and dictionary, and is the only vowel letter apart from *a*, *i*, *u* to appear in initials of values and morphemes, the letters *ü* and *ä* (i.e., the symbols for the presumed sounds denoted in the alphabet of the International Phonetic Association by [y] and [ɛ]) appear only in discussions on the phonetics of particular dialects of Akkadian. The vowel letter *o*, however, is on its way to establishing itself at least as a letter occurring in the transcription of the

62. See J. Aro, Die akkadischen Infinitivkonstruktionen, pp. 14f. and K. Deller, Or. NS 31 226.

63. von Soden, "Vokalfärbungen im Akkadischen," JCS 2 (1948) 291ff.

Neo-Assyrian dialect. It appears in the Syllabar, although only in the single value *kot* (illustrated by a foreign geographic name). It appears more frequently in the dictionary (AHw); even there, however, it begins no lexical entry (the letter N is followed by the letter P), nor has it appeared there so far in a canonical entry printed in bold face (unlike some consonants, for which see p. 49). Its occurrences are restricted to interpretive transcriptions of morphemes cited in their non-paradigmatic spelling, from which citations with *o* the corresponding *o*-containing variants are listed after the canonical form in the lexical entry. The transcriptions of non-paradigmatic spellings are based on morphophonemic considerations: they 'disambiguate' forms in which a morphologically predictable *u* is replaced by *a*, e.g., the transcription *ašpor* indicates that expected <*ašpur*> is spelled <*aš-par*>, and the reverse, e.g., the transcription *agdoṣṣoṣ* indicates that expected <*agdaṣṣaṣ*> is spelled <*ag-du-uṣ-ṣu-uṣ*>. On the other hand, the sign PAR is not given the value *por*, nor are the signs *du*, *uṣ*, *ṣu* given the respective values *do*, *oṣ*, *ṣo*. Thus the introduction of a new vowel letter has not increased the inventory of the syllabary nor consequently the inventory of phonemes of Akkadian. The appearance of the letter *o* in the transcription of Akkadian words or syllables thus takes the place of a footnote which might state that sign A of the text stands for sign B (e.g., *par* for *pur*), or, to state it in different terms, the letter *o* indicates that there exists a spelling variation between *a* and *u*, whether in a morpheme or morpheme string such as *ašpor* and *agdoṣṣoṣ*, or in a 'meaningless' syllable such as *kot*. In this sense, the letter *o* is a notation for a diaphoneme (comparable to the notation used by some linguists for the pronunciation of English, as in *sɔrij* (for the word spelled *sorry*), which means that some speakers pronounce *sorij* and others *sarij*. As such notation, the use of the letter *o* — and other letters that may be eventually introduced in the transcription of Akkadian words — is proper and acceptable, as long as it is kept in mind that it is the notation for a spelling variation and does not predicate anything about the phonological reality that is assumed for the phoneme corresponding to the notational letter; in other words, the use of the letter *o* does not correspond to the phonological reality of an *o* phoneme, or the existence of a free variant *o* of some phoneme (*o* can not be an allophone because it is not predictable from the phonological environment); such reality has to be proven by other means.

2. *Spirantized consonants*

Spelling variations observable in certain words have been interpreted as an indication that the consonant inventory of Akkadian included some consonants that have hitherto been regarded as lost from the Proto-Semitic inventory. For instance, the variation between initial

vowel- and initial *ḫ*V-signs in the word *adannum* is thought to render an initial pharyngeal stop (Hebrew ʿayin); the by-form *ʿadannum* of this word appears in the heading, after the form *adannum*, in AHw. However, I would like to use here as an illustration only the set of consonants that participate in the so-called spirantization.

Spirant instead of stop articulation has been deduced by von Soden from variation in spelling between a stop and the corresponding spirant (i.e., a pair of consonants sharing the same distinctive features, one of which is obstruent, the other non-obstruent). Detailed arguments are given in the introduction to the Syllabar (pp. xix f.) and in JNES 27 214–20 under the title "Die Spirantisierung von Verschlusslauten im Akkadischen." The method followed is essentially that which led to the postulation of vowels: the phoneticization of spelling variation. The term spirantization itself is not without ambiguity: it can be interpreted as a term for the historical process of spirantization of stops in certain environments (e.g., the historic change customarily called the first German Lautverschiebung) and also as a term for allophonic variation (i.e., fricative articulation of stops in certain environments, as in Spanish intervocalically, or as in Hebrew and Aramaic (the so-called *bgdkpt*-phenomenon).[64] The two usages of the term lead to different conclusions: if the change is a historical one, it is not likely to be reversed within one historical period; once a stop has become a spirant in a given environment, its articulation will stay spirantic, and is not expected to change back to a stop. If the variation is allophonic in a certain period of the language, the occurrence of fricative allophones of stops must be predictable. If their occurrence is not predictable — and so far no environmental or distributional rules could be established to make such occurrences predictable — then the continuants in question must be distinct phonemes.

Neither solution has been suggested so far, mainly because of the small number of variations, and their restriction to a small group of lexical items.[65] Before assessing the conclusions that may be drawn from the data, I would like to review the evidence for spirantization. I will restrict myself to the spirantization of dentals and velars, since evidence for a spirantized pronunciation of labials ([*w*] or [*v*] for /*b*/ and [*f*] for

64. It was observed by Speiser, "The Spirantic Transcription of Ugaritic [ḫ] and [ḥ]," BASOR 175 (October 1964), that in the "occasional interchange between written *ḫ* and *k* . . . no form of spirantization (analogous to the treatment of *bgdkpt* in Hebrew and Aramaic) can be involved, because (a) most of the affected sounds are found in initial position, and (b) in the medial instance *maḫ/kir* it is the spirant that has been changed to a stop ([ḫ] > [k])" (p. 44).

65. Although phonemes have been posited for a restricted segment of the lexicon, such as Arabic emphatic *l* which occurs only in the word *Allah*, see Ferguson, Language 32 (1956) 446–52.

/p/) is much more tenuous, and it seems was prompted more by the Systemzwang of a symmetrical matrix than by compelling evidence to this effect.

The two series of consonants for which a spirantized pronunciation has been more fully documented than for the labials are the palatal and dental series. In the palatal (or rather palato-velar) series, the phoneme inventory consists of the stops /k/, /g/ and /q/ and of the fricative /ḫ/. Spirantized pronunciation has been claimed for /k/ only; a spirantization of /g/ to a voiced palatal fricative is considered as operating in Aramaic only. The claim is based on spelling variations, which can be usefully tabulated as follows:[66]

a) etymological *k* is written with <*ḫ*V(C)> signs (initially, intervocalically, and before /*m*/);
b) etymological *ḫ* is written with <*k*V(C)> signs (initially, intervocalically, and after /*m*/);
c) there is a spelling variation <*ḫš*> ~ <*šk*>, also written <*ša-ak*>, for final <*šk*> or <*ška*>;[67]
d) there is a dialectal variation between <*q*VC> and <*ḫ*VC> or <V*q*> and <V*ḫ*>.

The phonological interpretation of these spellings leads, if taken seriously, to the extreme conclusion that the phonemes /k/ and /ḫ/ are neutralized in all dialects and in practically all environments. Alternative interpretations, such as the existence of lexical doublets, may be put forward to avoid a phonological chaos, especially since the lexical items for which the spelling variation is attested are few. However, the range of the hitherto attested spelling variations is such that further attestations in other lexical items may be expected.

The case of the dental series is only superficially similar to the palatal series. The phoneme inventory consists of the stops /t/, /d/ and /ṭ/, and of the sibilants /s/, /z/ and /ṣ/. The spelling variation, however, is not so much between dental stops and dental spirants as between the (voiceless) dental stop /t/ and the palatal spirant /š/. Just as the case was within the palatal series, there is no claim for the spirantization of the (voiced) stop /d/; spelling variation of /d/ and /z/ concerns only one lexical item which admits of the historical solution *ziḫḫu* > *diḫ*(*ḫ*)*u*; attested variations or possibly morphophonemic alternations between /d/ and /z/ or /s/ affect Sumerian only.

The claim for the spirantization *t* > θ is based on spelling variations which, again, may be usefully grouped as follows (for examples see below):

66. See E. E. Knudsen, AOAT 1 147–155.
67. In the word *laḫaruššu/laḫaruška/laḫaruḫšu*.

A. There are *no* spelling variations <*t*> vs. <*š*> for etymological *t* or *š* as part of a root except one (see 1a*α*).

B. Spelling variation <*t*> vs. <*š*> is attested in foreign words (1a*γ* and 1b) and grammatical morphemes (see 1a*β* and 2).

The cases for which this variation is attested can be classified as (1) non-morphophonemic and (2) morphophonemic.

(1) Non-morphophonemic variation occurs (a) in initial position (<*š*V> varies with <*t*V>), and (b) before consonant (<V*š*> varies with <V*t*>)
   (a) applies to
      α) the spelling <*tisi*> for etymological <*šisi*>
      β) the variation <*ša*> vs. <*ta*> in a derivational prefix
      γ) other initial variations <*š*V> vs. <*t*V>
   (b) applies to <V*š*> vs. <V*t*> before consonant.

(2) Morphophonemic variation occurs in
   (a) <V*š*(*i*)> vs. <V*t*(*i*)> in pronominal suffixes and pronouns,
   (b) <*tt*> vs. <*št*> in a class of participles,
   (c) initial <*š*V*t*> vs. <*t*V*š*> in a class of paradigmatic forms of the verb.

These types admit of different interpretations; 1a*γ* and 1b) affect foreign words and can be accounted for either as sporadic change or as phonologically conditioned dissimilation, respectively assimilation; 1a*α*, the only case affecting a root consonant, can be accounted for as dissimilation; 1a*β* is most likely a scribal error; for 2a, b and c a morphophonemic interpretation is available.

1a*α*) The variation obtains for the lexical items *šisi* and *šisītum* for which the spellings <*ti-si*> and <*ti-si-tum*> occur. This may be interpreted as */tisi/* and */tisītum/* resulting from the dissimilation *#š-š* ⟶ *#t-s* (i.e., initially and in non-contact position). The similar item *šiṣītu* vs. *tiṣītu* cited sub 1a*γ* because its etymology cannot be established, is also explained by the above rule if *s* = {*s*, *ṣ*}.

1a*β*) There are two examples of the derivational prefix *ša* written as ***ta***. both in texts written by Hittite scribes.[68]

1a*γ*) Initial <*š*V> 1′ varies in Old Babylonian with initial <*t*V>,
   2′ varies in Old Babylonian with, and is written later as <*t*V>.

68. One, *taḫ-lu-uq-ti*, is cited by von Soden; the other occurs in *ta-aḫ-lu-uq-ti*, see Riemschneider, StBoT 9 p. 70. Note that the two cuneiform signs *ša* and *ta* are very similar; moreover, in Old Babylonian cursive, the two signs are usually not differentiated, hence von Soden considers as less conclusive evidence for spirantization the writing of the stem preformative */ša/* as */ta/*, the only example cited being *utaškin* for *ušaškin*. Conceivably this spelling represents not a scribal error but a dissimilation — although not elsewhere attested — of *š-š* to *t-š*.

There are four words to which the above applies: *š/tupšikku*, *š/tabsūtu*, *š/tiṭ²āru*, and *š/tiṣītu*; three of these contain, as von Soden has already pointed out, a sibilant in the syllable next to the one for which the variation is attested; they have, however, been assigned to different classes by von Soden on the basis of their etymology: the first allegedly comes from Sumerian *dusu*,[69] the second from Sumerian *šabzu*, and the third and fourth are of unknown etymology.

3′ varies, in later periods, with <*t*V> (examples: *šer²azu* vs. *tur²azu*, *šasqû* vs. *tasqû*, but also *sasqû*). Note, again, the presence of a sibilant in the following syllable.

1b) before consonant, <Vš> varies with <V*t*> (examples: *atgigi* vs. *ašgigi* [OB], *ḫaṭḫūru* vs. *ḫašḫūru* [LB], *atkuppu* vs. *aškuppu* [SB]).

2a) The two sets of pronominal suffixes and pronouns */šunūt/*, */šināt/*, */šunūti/*, */šināti/*, */šuāti/*, */kuāti/*, */šiāti/* and */šunūš/*, */šināš/*, */šunūši/*, etc. have a distribution according to their function as (genitive-) accusative *vs.* dative. Members of the first set — forms with final <*t*(*i*)> — also occur in the function of dative instead of the expected members of the second set in Old Akkadian, Old Babylonian and Old Assyrian; in later periods, members of one set freely vary with members of the other in both grammatical functions, with a preference, first, for forms with */š/* and later for forms with */t/*. The phonetic interpretation as spirantization of final or intervocalic */t/* is not necessarily the only possible solution: the variations may also be interpreted as syntactic variants (the dative case replaced by the essentially oblique genitive-accusative case in the earlier periods)[70] or as a merger of the two object cases into one, resulting in free variation of the two sets.

Moreover, if the phonetic interpretation were the favored solution (i.e., <*š*> and <*t*> would be written as an approximation of [θ]), one would expect spellings <*tunūt*> and <*tunūš*>, etc.

2b) <*tt*> varies with <*št*> in participles of derived stems with */n/* preformative, namely participles of the form *muttaprisu* from an underlying *muntaprisu* also occur with the spelling *muštaprisu*. The phonetic interpretation of this spelling variation considers a pronunciation [*muθtaprisu*], i.e., written <*tt*> and <*št*> both as

69. The word is written both <*šu-up-ši-ik-kum*> and <*tu-up-ši-ik-kum*> and comes from a Sumerian word for which the reading *dusu/i* is only a late gloss, and thus the priority of the dental in this word cannot actually be assumed. I owe this information to M. Civil.

70. Parallel to the replacement of the (2.sg.) dative suffix */ku*(*m*)*/* by the accusative suffix */ka/*, for which no phonological interpretation has been sought.

approximations of a spirantized pronunciation [θt]. However, since participles of the form *muštaprisu* do exist as paradigmatic derivations of the stem with /š/ preformative, in this case too, as in 2a, a morphological interpretation, such as the replacement of one derived stem by another, or a grammatical rather than a phonological merger, may be considered.

2c) Metathesis of initial <šVt> and <tVš> is sporadically attested in forms derived with the infix /Vt/ from stems with initial /š/ (example: <*tišamme*> and <*šitamme*>). A phonetic interpretation of the spelling variation as approximating a pronunciation [θVθ] of the initial is a priori unlikely on phonetic grounds (previously listed instances show a trend of dissimilation rather than assimilation); moreover, the sound change assumed can be shown to belong to a different level from those exemplified by the other types. This is so because the previously discussed types of variation, which may or may not be interpreted on the phonetic level alone, apply to consonants within one morph, and may represent historical change or sporadic change, while the <šVt> vs. <tVš> variation straddles two morphs (the stem and the infix), and thus is to be considered in morphophonemic terms, as a phenomenon contingent on a particular morpheme boundary. The /Vt/ segment is infixed into a base after its first consonantal radical; the phonetic change thus affects the /t/ of the /Vt/ infix in conjunction with the first radical under two conditions: (1) if the latter is an /š/, and (2) if it stands in absolute initial position. Thus, on morphological grounds too, a phonetic interpretation of the variation as [θVθ], i.e., the leveling of the difference between the consonant of the infix and the consonant of the root into one phoneme (or allophone) [θ], is extremely unlikely, since the resulting form would show neither the lexical item (radical š → θ), nor the morphological category of which the infix is the exponent (morpheme /Vt/ → Vθ). Such an interpretation could be considered only if this were the only acceptable one to explain the spelling variation. On the contrary: the range of the variation observable speaks against such an interpretation. If the label [θ] indicates that the spelling may show either <t> or <š>, then it can be used for this type only with an added restriction rule, to exclude the nonoccurring spellings <tVt> and <šVš>. In fact, the lexical item /*šitamme*/ spelled <*ši-ta(m)-me*> is also spelled <*ti-ša(m)-me*>, but never <*ti-ta(m)-me*> nor <*ši-ša(m)-me*>. The requirement of an additional rule is against the criterion of simplicity of the description.

A further reason for rejecting a purely phonological interpretation is the fact that such variation does not occur in phonologically

similar sequences which are morphologically different: e.g., */tiši/* 'nine', */tišab/* 'sit down', */šiti/* 'drink', */šiṭirtu/* 'writing' occur, but not **<šiti>* 'nine', **<šitab>* 'sit down', **<tiši>* 'drink', **<ṭiširtu>*, etc. Thus, information on the phoneme sequence is not sufficient to predict the variation, but morphological information (on the presence of an infix */Vt/*) is necessary; the rule has to be stated in morphophonemic terms. There is indeed a morphophonemic rule, stated in LAA 6.1.4.3.b as Z + V*t* = *t*VZ (this is the rule that formalizes the metathesis of the initial sibilant of a stem with the *t* of a *t*-infixed form). In this rule, the members of the set Z are {*s*, *z*, *ṣ*}; to include the variation *<šVt>* vs. *<t*Vš*>* in this rule (as it was done but, as an optional phenomenon, not formalized in LAA 6.1.4.3.c), the set Z may be redefined as Z = {*s*, *z*, *ṣ*, (š)}, where the parentheses express the optionality of the metathesis with reference to the parenthesized š. Consequently, the notation in pointed brackets (symbolizing spelling) can be replaced by slants (symbolizing phonemic transcription) for this phenomenon, such as */tišamme/*.

It should be pointed out in connection with a phonetic interpretation as 'Zwischenlaute' of spelling variation, that data have been collected in such a systematic fashion only for variation between stops and spirants, that is, exactly for the phenomenon of spirantization that is known from other Semitic languages, namely of Hebrew and Aramaic *bgdkpt*. Counterexamples have not been collected, such as spelling variations for other pairs of consonants, for example, a labial *vs.* velar, interpretable as a labiovelar 'Zwischenlaut' (an alternation known in Sumerian between the main and the Emesal dialects, as Professor M. Civil informs me), or a stop *vs.* (homorganic) nasal, standing for a prenasalized stop, e.g., *ᵐb*, *ⁿd* (a phoneme known from many languages, and as contact assimilation in Akkadian, see LAA 6.1.1.1 and 6.2.1.4); even less for the pairs */t/* vs. */l/*, or */k/* vs. */p/*, since the languages of the area do not seem to have the phonemes */ƛ/* and */k̂p/*, etc. Indeed, variation between voiced *vs.* voiceless consonant might also have been interpreted as expressing such a 'Zwischenlaut' as voiceless media.

I do not contend that evidence for such variations could be found in a significant number and distribution to suggest any phonetic or phonological law; I rather suggest that variations in such spellings, not to speak of those for which one would be hard put to find a phonetical analog in some language of the world, would not be more significant in number than and would be just as randomly distributed as, some of the variations that might suggest 'spirantization', and thus the stop *vs.* spirant variations that cannot be reasonably interpreted either as morphophonemic change or as historical change or as 'sporadic' change, may well belong to the extensive collection of scribal errors.

Let us assume, however, that the existence of such sounds as the mentioned spirants is deducible from spelling variation, and consider the consequences. We have (1) a (voiceless) velar spirant, i.e., the consonant *ḫ* that is customarily included in the phoneme inventory of Akkadian, and (2) a (voiceless) dental spirant, i.e., a hitherto not inventoried θ (Semitic transcription: *ṯ*). If (1) is a variant of another phoneme (*/k/*), then we cannot speak any more of two phonemes, only of one phoneme (say, */k/*), with two allophones (*k*, *ḫ*), (the label is as a rule given to the phoneme for etymological reasons) or we must say that the two phonemes */k/*, */ḫ/* are neutralized in some particular environment or some particular word or words. The latter assumption, that two phonemes are neutralized in some words only, is not acceptable on the basis of language universals, since phonetic change operates across the entire system (see Martinet, Economie des changements phonétiques [Bibliotheca Romanica, Series Prima X, Bern: A. Francke S.A., 2nd ed., 1964], pp. 26f.). To say that neutralization occurs in a given environment is equivalent to saying that there is only one phoneme; and since the one that is given as basic or the original of the two has been elicited on the basis of etymology or morphology, we should call it a morphophoneme. Thus, the discussions that use the terms phonemes and (phonetic) allophones would better be conducted in terms of morphophonemes and phonemes. As it is anyhow to be expected when dealing with a dead language, phonetics need not, indeed cannot, enter the picture at all.[71]

V. THE VALUE-TO-SIGN LIST

It has been mentioned (p. 17) that this list is an index to the sign-to-value list. Were one to use it as sign-selection rules, in addition to a set of writing rules, it could be rearranged in a more economic and at the same time more directly accessible form. If the sign-to-value list were abbreviated along the lines suggested, the value-to-sign list as it stands now would serve as expansion of the abbreviations in conjunction with the reading rules. If the sign-to-value list is left as it stands now, as illustration of orthography, the value-to-sign list could give in condensed form the sign-selection rules. This form could be a table, so that, instead of the alphabetical list, e.g.

71. It should be mentioned here that the phenomenon of spirantization has been duly noted by Thureau-Dangin, in his Supplément au Syllabaire accadien, published in Les homophones sumériens (Paris 1929), Appendice (pp. 41–52). However, Thureau-Dangin refrained from creating new values for a "spirantized pronunciation" such as the value *ḫaṣ* to the sign PA (basic values *pa*, *ḫaṭ*) and the value *baḫ* to the sign ḪU (basic values *ḫu*, *paq*).

dub^{p}	DUB	101
dúb^{p}	DÚB	201
dug^{k,q}	DUG	164
dúk	TUK	309
duḫ	GAB	117

etc., the table might look like this:

dup	DUB/101	DÚB/201
duq	DUG/164	TUK/309
duḫ	GAB/117	

etc., and the selection of one of the two or more equivalent columns, i.e., of the two or more possible spellings for one value, might be indicated by some type of environmental constraints that now are not included in the index, but which become apparent if one uses the list purely as a cross reference index and checks the prevailing constraints (such as historical period), under the serial number given for the sign.

It should be noted that the table, with the value rewritten in line with the notation suggested pp. 7f., and p. 24, actually conveys more information than the alphabetic list: it predicts a reading <*dug*> and <*duk*> for the sign TUK (no. 309), although no illustration for this reading is provided. The proposed notation also eliminates the necessity for writing (twice!) *dub*^{p} and (once only!) *dug*^{k,q}. If the notation proposed on p. 24 were used, the alphabetic entries

tub^{p}	DUB	101
túb^{p}	DÚB	201
tùk^{q}	DUG	164
tug^{k,q}	TUK	309
tuḫ	GAB	117
ṭup	DUB	101
ṭúb	DÚB	201
ṭuḫ	GAB	117

could all be eliminated and replaced by the revised table

ṭup	DUB/101	DÚB/201
ṭuq	DUG/164	TUK/309
ṭuḫ	GAB/117	

by this means collapsing thirteen lines of the alphabetic list into three lines of the table. Let us see, however, whether the table does not predict non-occurring values: the first and last lines do not, since the alphabetic list gives all three possibilities for the initial: voiced, voiceless, and emphatic. The second line, however, would predict the following values that are not found in the alphabetic list:

tug, *ṭug*, *ṭuk*, *ṭuq* for DUG/164
dug, *duq*, *ṭug*, *ṭuk*, *ṭuq* for TUK/309.

These values can be assigned to two groups: one would include the value *ṭuq* for both signs, whose non-occurrence is determined by general constraints (the so-called incompatibility, see below); the other would include *tug* and all initial emphatic *ṭ* values for DUG/164 whose non-occurrence is accidental, and *dug*, *duq*, *ṭug*, *ṭuk* for TUK/309. This latter gap makes us suspicious: we check the value *dúk* listed for TUK/309 and find that it is given in parentheses and that its attestations are restricted to a peripheral dialect and to one single Assyrian text. In such a case, we must rewrite our line in the table as

tuq	TUK/309	(the initial voiceless indicating that the value contains only the voiceless itself)
ṭuq	DUG/164	(where we assume that non-attestation of other values is accidental)

This may lead us to consider the gaps in the value-to-sign list and discover that they are of two kinds: systemic gaps, and accidental non-occurrences. The systemic gaps are due to the incompatibility law in Akkadian, which states that no root contains either two non-homorganic emphatics or two non-identical homorganic consonants. Thus, while the sign <*kak*> has the values *gag*, *kak*, and *qaq*, it does not have the values *gak*, *gaq*, *kag*, *kaq*, *qag*, or *qak*, since there are no Akkadian words in which two dissimilar velars occur.[72] On the other hand, to the accidental non-occurrences belong systemically predictable values, such as */kaṭ/* which is missing from the list which gives *kad*^{t} and not **kad*^{t,ṭ} as opposed to the list's *rad*^{t,ṭ} which predicts all three dentals in syllable-final position.

VI. RETROSPECT

In this review of current practices in transliterating cuneiform texts I have tried to draw attention to the fact that there is a difference between phonemic transcription and morphophonemic transcription. It is possible to use either one or the other when transliterating a cuneiform text or when transcribing it, but the underlying premises of each are different. Cuneiform orthography is a mixture of the two — as orthographies usually are — and uses both methods indiscriminately. However,

72. Exceptions are found in the list, e.g., for the sign PAP for which a value *bap* is listed (illustrated by the Sumerian loanword *bap-pi-rum*) and for the sign GAD for which a value *qàṭ* is listed (illustrated solely by the West Semitic geographical name *Qàṭ-na*).

if we want to distinguish between the various language levels we would be well advised to keep the phonemic and morphophonemic notations apart. A mixed notation, like the one used by the ancient scribes, requires constant analysis of the grammar, and thus contains more challenge for the intelligent reader and more pitfalls for the unwary one.

By implicitly containing all values that make it possible to transliterate cuneiform texts on either the phonemic or the morphophonemic levels, Das akkadische Syllabar by von Soden and Röllig provides the necessary data for the reader who is interested in explicitly keeping the various levels of linguistic analysis apart; it remains the basis for future research on Akkadian phonetics, phonemics, and morphology.