The transliteration of Sumerian
(Jeremy Black, 2004)

The transliteration of Sumerian into roman script is a matter that causes some confusion for cuneiformists who are not specialised in Sumerian, for students, and for interested non-specialists. Why does this aspect of scholarly practice seem to have changed so much over the last hundred years? Why does practice even now vary from specialist to specialist?

There is no agreed standard, and ETCSL has adopted and implemented its own system based on good practice. However, different transliteration practices will be found in other electronic resources and in printed books. The transliteration of Sumerian is still evolving. (However, the transliterated forms of Sumerian signs or words used logographically in writing Akkadian have become by and large standardised and have not kept pace with changes in our understanding of the Sumerian language. This is not really relevant here.)

Nature of the writing system
Sumerian is written in cuneiform script using a system of signs some of which stand for individual syllables (V, CV, VC or CVC), or for whole words (so-called logograms), or determinatives (signs that indicate the semantic category of a word, e.g. place name). When represented in roman script, the syllable signs that make up a Sumerian word are written joined together with hyphens (see 'Hyphenation principles').

Two important aspects of cuneiform writing are homophony and polyphony, both resulting from historical development. The same syllable sound can frequently be written in a number of different ways. These are conventionally distinguished by numerals: šu, $šu_2$, $šu_3$ etc. These are said to be homophones. Conversely one sign can have multiple sounds or 'readings', e.g. the sign DU can also in different circumstances be read g̃en, gub, tum_2 etc. This is polyphony.

Subscript numerals are written to distinguish homophonous signs from 2 onwards (i.e. number 1 is not written). It is now less common to mark numbers 2 and 3 with the acute and grave accents (éš for $eš_2$, èš for $eš_3$) in Sumerian, although it is still normally used in transliterating Akkadian.

Symbols used in transliteration
In general the symbols used to transliterate Sumerian are:
a e i u
b d g g̃ ḫ k l m n p r s š t w y z.
The symbol g̃ is used to indicate the velar nasal ('ng') consonant reconstructed for Sumerian. It was an unfortunate choice, as it exists only in special fonts; but it is widely used in the field. Some German Sumerologists use g+circumflex instead. Some scholars do not distinguish the g̃ from ordinary g, but this is potentially confusing and fails to disambiguate certain words.

Sometimes ḫ is written h (as there is no other form of aspirate in Sumerian from which it needs to be distinguished). The true phonological nature of the sound is uncertain. Occasionally the letters ' q ṣ and ṭ are used in Akkadian words or names occurring in Sumerian material, to represent the aleph (glottal stop) and the three so-called 'emphatic' consonants of Akkadian. In the ETCSL corpus ' ṣ and ṭ are encoded as entities ℵ, &s; and &t;.

Determinatives and phonetic complements (signs indicating the pronunciation of more complex signs) are written superscript.

In electronic search engines and other www material, it is sometimes quick and convenient to use j for ĝ, and c for š. This has the advantage of not requiring any diacritics. In such circumstances, the subscript numerals are written on the line, e.g. ja2 instead of ĝa$_2$, ce3 for še$_3$. This is also simpler. In some print publications, superscript determinatives or phonetic complements appear on the line separated by a point, e.g. nibru.ki for nibru^{ki}, mu-un-jar.ar for mu-un-ĝar^{ar} — again this has the advantage of not requiring any special typographical effects, although it could be confusing.

Types of variation in transliteration

The broad types of variation that are likely to be encountered in different transliteration practices are four.

(i) The transliteration is different because the editor has chosen an entirely different reading of the sign or sign-group, implying identification with a different lemma. To read the sign DU as ĝen rather than gub (which normally has the meaning 'to stand') implies an interpretation of it as representing, in this particular context, a different verb (ĝen 'to go') instead.

(ii) The transliteration is different because, on the basis of historical evidence either newly reinterpreted or not previous available, the reading of the sign has been revised (either throughout Sumerian or in this particular context). An example is the sign MUNUS, with the reading munus 'woman', previously often read as mi$_2$. The reading mi$_2$ is now restricted to the phrasal verb mi$_2$…dug$_4$ 'to treat kindly'. Many improvements of this sort have been made to transliterations in recent years, yet Sumerologists must remain familiar with the more old-fashioned transliterations encountered in many standard printed sources.

(iii) A small phonological change has been made to the reading of the sign, again because, on the basis of historical evidence either reinterpreted or not previous available, the reading of the sign has been revised. Examples are šir$_3$ (formerly sir$_3$), edin (formerly eden), iri (formerly uru), šudun (formerly šudul). Again, many changes of this sort have been made recently.

(iv) The transliteration varies between a long form of the word or syllable and a short form. The 'long' form generally has a consonant at the end. Examples are kur$_9$ ~ ku$_4$, dirig ~ diri, pad$_3$ ~ pa$_3$, til$_3$ ~ ti. It is a feature of Sumerian orthography that the final consonant of many such bases is clearly written again in any suffixed form of the word (where the suffix begins with or consists of a vowel), e.g. kur$_9$-re, dirig-ga, pad$_3$-da, til$_3$-le. This is an orthographic feature and does not imply doubling of the consonant. (In some words ending in -d, the d becomes -r- when followed by a suffix. In some words ending in -n, the n becomes -m- before a suffix.)

In almost all cases ETCSL has chosen the long form. There are two basic reasons for this. First, using the long form actually helps to disambiguate certain words. For example, the three different words zi 'life', zig$_3$ 'to rise' and zid 'right, just' are all written with the same sign ZI, which can have the readings zi, zig$_3$ or zid. By transliterating differently in each case,

the meanings are clear and the words can be lemmatised automatically. The second reason is a pedagogical one. If students learn the long readings from the first, they will be able to differentiate the lemmata and avoid errors in the identification of bases. Similarly the historical spelling *court* in French identifies the word as the adjective 'short' (fem. *courte*) and disambiguates it from *cour* 'courtyard', although the pronunciation is the same.

This has very little to do with the reality of Sumerian phonology. We will never be able to know whether Sumerian final stops were pronounced but unexploded (as in Cantonese *wok*) or not pronounced at all (as in French *chat*). It is a mistake to confuse transliteration (a regularised representation of the cuneiform signs present on the tablet) with attempts to reconstruct the exact phonology of Sumerian. The important thing is the regularity of transliteration, so that any subsequent changes can be made globally. Our experience shows that even advanced Sumerologists can adapt to new readings very quickly!

What is the basis for the transliterated values of signs?

When Sumerian was first deciphered in the second half of the nineteenth century, it was on the basis of bilingual Sumerian-Akkadian texts; the reading of Akkadian was already known. The first values assigned to Sumerian signs were those already known for Akkadian. Subsequently many of these have had to be revised. The main basis for revision was parallel variant spellings of the same word or syllable.

Gradually the ancient lists of cuneiform signs have been reconstructed. These were used by Mesopotamian scribes learning to write cuneiform, and Sumerian in particular. In these invaluable lists, signs were listed with Akkadian translations, and phonetic spellings of the Sumerian pronunciations appropriate for each Akkadian translation. However, the sign lists evolved over a period of 3000 years, and it has been realised that the most authentic Sumerian pronunciations were not those from the enormous editions of the lists dating from the Standard Babylonian literary tradition mainly of the first millennium BC, but the earlier versions that were extant in the period when Sumerian was still a living language. Assyriologists refer to the Standard Babylonian editions of some of these lists by the abbreviations Ea (short for "ea = *nâqu*", the first entry in the list) and Diri (short for diri = *atru*). The earlier versions are consequently called Proto-Ea and Proto-Diri, and date from the Old Babylonian Period (roughly 2000–1650 BC).

Thus as far as possible, the values of Sumerian signs are assigned on the basis of the pronunciation glosses in Proto-Ea and Proto-Diri. Occasionally the information in the manuscript sources of these lists is discrepant. There are some other sources which are used where Proto-Ea and Proto-Diri are not preserved. Sometimes pronunciations are given in other, later sign lists. Sometimes glosses are taken from other sources, and in certain literary contexts some words are spelled out phonetically.