Sumerian hyphenation: an attempt at codification

Transliterated Sumerian is written with the syllable signs making up individual words hyphenated together. Decisions about the hyphenation of Sumerian words thus relate to the concept of a ‘word’ in Sumerian (see J.A. Black, ‘Sumerian lexical categories’, in Zeitschrift für Assyriologie 92 (2002), 60–77). In Sumerian studies, uncertainty about word boundary has led to maximalist (e.g. lugal-ki-en-gi-ki-uri-ke₂, nitaḥ-kalag-ga, me-sikil-bi) and minimalist (e.g. lugal ki-en-gi ki-uri-ke₂, nitaḥ kalag-ga, me sikil-bi) uses of hyphens in transliteration. There is so far no standard in the field. Accordingly ETCSL has established and implemented the procedures described below. This is an interim statement (29th February 2004), and inevitably will be subject to revision.

The usual criteria for identification of word boundary do not exist, e.g. pause, stress. Also words are frequently written on clay tablets with no space between them. Clitics are frequent in Sumerian, and it is not always clear if something is a word in its own right or only a clitic. Clitics are attached to words or clauses, and in all probability carried no stress of their own. A whole sequence of clitics can be attached to e.g. the last word of a clause. They may often have been phonologically integrated with their hosts; the writing system certainly writes most of such clitics as if they were in sandhi with the previous word (or clitic). Finally, in Sumerian a large number of compound nouns exists.

Sometimes the judgment whether something is a lexicalised compound or an ad hoc phrase is based on frequency of occurrence. Sometimes the judgment is informally based on (Akkadian, European language) translation value. Ideally the judgment would be based on syntactic grounds, e.g. can the compound itself be modified by an adjective? At all events, it must be pragmatic, since language is irregular, but decisions are needed. Sometimes this will help with lemmatisation, by disambiguating homographs. Sometimes the decision is taken on grounds of convenience. Multiword verbs exist in which the lemma is composed of two words (which, in a finite verbal form, are separated by proclitics).

- Hyphenation is used to join all written syllables of a single word together, and:

- to join clitics or affixes to the preceding base (or, with verbal proclitics, the following base)

- to join all the elements of proper names e.g. kar-zid-da ‘Right-hand Quay’, kur-ḫeš-erinkud ‘Mountain Where Cedars Are Cut’, e₂-barag-dur₂-ḡar-ra ‘House Where Daises Are Established’, amaš-kug-ga ‘Holy Sheepfold’, where it is clear that they are proper names and not e.g. metaphorical descriptions

- to join ‘compound’ nouns (i.e. believed by us to be lexicalised), in various categories:

(i) co-ordinate compounds, e.g. maš₂-anše ‘wild animals’, gud-ab₂ ‘cattle’, gud-udu cattle and sheep’, u₂-šim ‘vegetation’; NOT an ki ‘heaven and earth’
(ii) right-headed compounds of two nouns (very few in number), e.g. an-edin ‘high (part of the) desert’, sağ-men ‘(a type of) crown’

(iii) left-headed compounds (the largest category):

a) where a noun is followed and modified by:
   - another noun in the genitive, e.g. gi-dub-ba ‘stylus’ (lit. reed of tablet), lu€-inim-ma ‘witness’, e€-nu€-da / a€-nu€-da ‘bedroom’, maš₂-ud₃-da ‘buck goat’.

b) participles in -a, e.g. kaš-de₂-a ‘wedding offering, party’, di-til-la ‘completed judgment’; but note: a₂ āg₂-ğa₂ ‘command’ (see below).

c) nominals derived from finite verbal forms, e.g. ki-in-dar ‘crevice’, ki-in-du ‘earth’.

Included here are compounds where the second element is a verbal noun but in which the first element is the head.

There are also a number of compounds representing agents (professions) and instruments (tools), e.g. dub-sar ‘scribe’, igi-nu-du₇ ‘(type of) slave’, gaba-ri ‘rival’, kišib-la₂ ‘seal-bearer (also: hand, paw)’, gana₂-ur₃ ‘harrow’, which may be thought of as expressions in which the head (lu₂ or ʒiš) is missing, for instance (lu₂-)dub-sar ‘(person) writing tablets = scribe’.


or with e₂- (+N, or +Adj): e₂-gal ‘palace’, e₂-maš₂ ‘animal stall’, e₂-dub-ba-a ‘scribal school’, e₂-ği₄-a ‘bride’, e₂-nu₂-da (a₂-nu₂-da) ‘bedroom’.

Also with gu₂-: gu₂-en ‘lord’, gu₂-bar ‘naple of the neck’, gu₂-erim₂ ‘wicked, enemy’, gu₂-gal ‘pre-eminent’, gu₂- gid₂, gu₂-širi₃ ‘breach’, gu₂-kiḡ₂, gu₂-nida, gu₂-saḡ ‘?’. 


(iv) for nouns derived with nam- or nu-: nam-lugal ‘kinship’, nam-dub-sar ‘scribal skills’, nu-kiri₆ ‘gardener’, nu-banda₃ ‘overseer’.

(v) for nouns formed from finite verbal forms: u₃-na-dug₄ ‘letter’, ga-an-tuš ‘tenant’.

(vi) There is a group of compounded adjectival forms (a list of at least 37 which can be considered to be lexicalised, for syntactic or semantic reasons), compounded with the verbs -zu ‘knowing’, -tuku ‘having’, -du/di ‘doing’ and -gal₂ ‘being/having’. These are hyphenated. They will be lemmatised as single lemmata.

kug-zu ‘wise’
gal-zu ‘skilful’ (cf. gal-zu niḡ₂-nam-ma ‘wise in everything’) 
gal-an-zu ‘clever’
usu-tuku ‘powerful’
a₂-tuku ‘powerful’
ni₂-tuku ‘reverent’
šiš-tuku ‘attentive’ (cf. sipad šiš-tuku-ni-ir ‘to his attentive shepherd’)
lipiš-tuku ‘furious’
gu₂-tuku ‘perfect, splendid’
ḥul-du ‘wicked(-acting)’
em₂-du ‘hostile(-acting)’
zid-du ‘righteous(-acting)’
maḥ-di ‘illustrious’
gal-di ‘important, impressive’
a₂-bad-ĝal₂ ‘protective’;
a₂-ĝal₂ ‘strong’, also neg. a₂-nu-ĝal₂ (cf. a₂-ĝal₂ kalam-ma-me-en ‘you are the strong one of the Land’)
a₂-nun-ĝal₂ ‘impressively strong’ (cf. a₂-nun-ĝal₂ e₂-kur-ra ‘most powerful in E-kur’)
an-ta-ĝal₂ ‘exalted’
dug₂-ĝal₂ ‘well-disposed (?)’
gaba-ĝal₂ ‘forceful’
(gu₂-)erim₂-ĝal₂ ‘wicked’ (freq. also in headless use, parallel to (gu₂-)erim₂ ‘wicked, enemy’)
ḥul-ĝal₂ ‘evil’, neg. ḫul-nu-ĝal₂ (varies with ḫul ‘evil’)
igi-ĝal₂ ‘wise’, also neg. igi-nu-ĝal₂ (cf. also igi-ĝal₂-tuku; also used participially: igi-ĝal₂ ‘taking account of’; and cf. the noun igi-ĝal₂ ‘insight, wisdom’)
inim-ĝal₂ ‘eloquent’
kug-ĝal₂ ‘holy’ (but note kug-ĝal₂ = gu₂-ĝal/gal₂ ‘canal inspector’) (varies with kug ‘holy’)
la-la-ĝal₂ ‘charming’
me-te-ĝal₂ ‘seemly’
mud-ĝal₂ ‘creative (?)’
ne₂-ĝal₂ ‘mighty’ (written with ne₃(PIRĪḠ), also ne-ĝal₂)
nir-ĝal₂ ‘lordly, authoritative’, also neg. nir-nu-ĝal₂ (cf. dumu nir-ĝal₂ an-na-ke₄ ‘lordly son of An’)
ni₂-ĝal₂ ‘awe-inspiring’
peš₁₀-ĝal₂ ‘mighty (?)’ (also written peš-ĝal₂) (cf. peš₁₀-ĝal₂ e₂-kur-ra ‘mighty one of the E-kur’)
saġ-ĝal₂ ‘safe, strong’ (in the expression: ki saġ-ĝal₂)
saġ-ki-ĝal₂ ‘obstinate’
saġ-kug-ĝal₂ ‘proud’ (var. saġ-gu₂-ĝal₂/gal; also saġ-gu₂-tuku)
su₁‹-ud-ĝal₂ ‘extended, lengthy (of time); patient-hearted’ (varies with su₁‹-ud ‘long’)
zi-ĝal₂ ‘living, alive’ (cf. id₂ zi-ĝal₂ kalam-ma ‘life-bringing canal of the Land’)

Expressions with niğ₂:
The following are assumed to be lexicalised compounds with niğ₂, on the basis mainly of frequency:
iğ₂-a-erim₂ by analogy with niğ₂-a₂-zig₃ and niğ₂-erim₂ ‘violence’
iğ₂-a₂-zig₃ ‘violence’
iğ₂-du₇ ‘that which is fitting, proper, necessary’ (usually co-occurs with vb. pa ed₂ (ak))(n.b. same writing as niğ₂-ul, see below)
iğ₂-gen₉(-na) ‘justice’
iğ₂-gig(-ga) ‘taboo’
iğ₂-si-sa₂ ‘rectitude’, ġustice’
iğ₂-erim₂ ‘wickedness’
iğ₂-hul ‘evil (abstract noun)’
iğ₂-me-ĝar ‘(ominous) silence’
iğ₂-ba ‘gift’
niġ₂-de₂-a ‘wedding gift’
niġ₂-gur₁₁ ‘treasure, riches’
niġ₂-kud ‘revenue, income, produce’
niġ₂-mi₃-us₂-sa₂ ‘wedding gift’
niġ₂-šam₂-ma ‘purchase’,
niġ₂-tuku ‘rich’
niġ₂-ur₂-rum ‘acquisitions, property’
niġ₂-ul ‘distant time in the past or future’ (n.b. same writing as niġ₂-du₇)
niġ₂-nam(-ma) ‘everything’
niġ₂-ar₃-ra ‘a type of flour’
niġ₂-gu₇ ‘food’
niġ₂-na ‘incense burner’
niġ₂-silə ‘dough’

niġ₂-TUG₂, see 2.1.7:1168 anše niġ₂-TUG₂-bi il₂-a-da [hapax; probably refers to an object (‘pack-saddles’)]
niġ₂-bun₂-na ‘turtle’
niġ₂-ur₂-limmu₂ / niġ₂-ur₂-4 ‘quadruped’
niġ₂-zí-ğal₂ ‘living creatures; wild animals’

various textiles:
  tug₃-niġ₂-barag ‘quilt, coverlet’
  tug₃-niġ₂-dara₂ ‘loincloth’
  tug₃-niġ₂-la₂ ‘binding, bandage’
  tug₃-niġ₂-lam₂ ‘(a ceremonial garment)’.

N.b. some hypercompounds of the above exist:
nam-niġ₂-erim₂, nam-niġ₂-si-sa₂, both in 1.3.1

Other expressions with niġ₂ are not hyphenated.

On the other hand, hyphenation is not used for:

• any form (finite, non-finite or nominal) of a recognised compound (multi-word) verb not covered by the above. E.g. a₂ aḡ₂-ḡa₂, inim bal-e, inim sig, su-lum mar; di kud ‘making judgements’ (but di-kud ‘judge’) (see below) etc.

• ad hoc verbal phrases or nominal phrases e.g. lu₂ tur ‘small person’ (cf. lu₂-tur ‘child’), ḥar-ra-an zu ‘knowing the roads’, sun₄ la₂ ‘wearing a beard’ cf. sun₄ za-gin₃ la₂ ‘wearing a dark beard’

• demonstratives such as ne(-en), re(-en) (contrast -bi and -e)

• the various expresssions with ud ‘day’:
  ud zal-la/zal-le-da ‘(during) the course of the day’
ud te-en-še₂/ta ‘to/from evening’
ud šu₂-uš ‘sunset’
ud sa₉(-a) ‘midday’ cf. ġi₆ sa₉(-a) ‘midnight’, ġi₆ u₃-na(-k) ‘dead of night’

and with ud = ‘time’:
ud da-ni₂ ‘(for) all time’, ud ul-li₂-a-še₂/ta ‘(in/from) those (ancient) days’ ud re-a ‘(in) those (distant) days’.

The syntactic status of some of these is a little uncertain. Note, however, ud-sakar ‘(new) moon’.

- expressions with na-me ‘any; (with negative verb) no’:
  niṅg na-me ‘anything’
  lu₂ na-me ‘anyone’
  diḡir na-me ‘any god’
  ud na-me ‘any time, ever’
  ki na-me ‘anywhere’.

There will be a relatively small list of agreed exceptions not covered by the above categories, e.g. ama-ar-gi₄ ‘freedom’ (where the second element is a verbal noun), giri₁-gzal ‘joy’ (where the second element may or may not be a verbal noun).

Note also:
nam-tar (n.) ‘fate’, but nam tar (vb. phrase) ‘who determines destinies’
šu-luḥ (n.) ‘a specific ritual’, but šu luḥ (vb. phrase) ‘who washes the hands; washing the hands’
di-kud (n.) ‘judge’, but di kud (vb. phrase) ‘who judges; act of judging’
ki-ḡar (n.) ‘socle; substitute’, but ki ḡar ‘who founds’
igi-gal₂ ‘wisdom’, ‘wise’, but igi ḡal₂ (vb. phrase) ‘who attends to, takes account of’
ḡiš-ḥur ‘plan, design’ (ḡiš-ḥur kal-kal ‘precious plans’; cf. ġiš…ḥur ‘to design’)

There are some problem cases to watch out for, e.g. does the combination noun + unsuffixed non-finite verbal form derive from a multi-word (‘compound’) verb or not?

Extra areas
Note the following:
šita₂ saḡ 50 ‘50-headed mace’, gud si dili ‘one-horned bull’ (but saḡ-dili ‘bachelor’), ab₂ igitu₂ ‘bright-eyed cows’, niṅg₂-ur₂-limmu₂ ‘four-legged things’, uḡ₃ saḡ-gig₂ ‘black-headed people’
nam gal ki aḡ₂-da ‘from great love’, nam gal ki aḡ₂ 엤dumu-zi-da-ke₄ ‘from great love of Dumuzid’, nam gal ḫul₂-la-da ‘in great joy’, nam gal₄-la-na-še₃ ‘because of her genitals’.