Sino-Uralic Etymology for 'Moon, Month'
Supported by Regular Sound Correspondences

Jingyi Gao¹, Tõnu Tender²

¹ Beijing International Studies University, China; E-mail: gao.jingyi@bisu.edu.cn
² Institute of the Estonian Language, Tallinn, Estonia; E-mail: jingyi.gao@eki.ee

Abstract

Using etymological methods, the present study has researched four Sinitic and Uralic shared etymologies (etyma). Two of them form a rhyme correspondence. Three of them form an onset correspondence. These regular sound changes validate the genetic connection between Sinitic and Uralic. The Sino-Finnic term for 'moon, month' is among these four etyma. It is demonstrated that this term should be aboriginal in Sino-Uralic languages.

Keywords: etymology, rhyme correspondence, onset correspondence, Sinitic, Uralic, Sino-Uralic, astronomical terms, moon, month

Introduction

The Uralic term for 'moon' (equivalents e.g. Finnish kuu 'moon, month'; Estonian kuu 'moon, month'; Hungarian hőhava- 'moon') has been compared to the Sinitic term for 'moon' 【月】 and suggested as a Sino-Uralic etymology (Gao, 2008, p. 231). The present study researches and supports this etymology with regular sound correspondences.

Materials and methods

The present paper is a comparative etymology study. The Sinitic language family is compared to the Uralic language family.

The Sinitic etyma are led by Chinese etyma (DOMs) which are historically attested Chinese glyphs (Sinograms). Their historical glosses are cited from the Chinese classical dictionaries (121-SW; 543-YP; 1008-GY). Their historical phonological values are cited from the work 1161-YJ (with reference to 1008-GY) and transcribed according to (2014, pp. 81–83). Their attested equivalents including forms and glosses are represented by Beijing Yan (Mandarin) (written in Hanyu Pinyin including non-simplified forms), Guangzhou Yue (Cantonese) (written in Jyutping), Taipei Min (Minnan) (written in Tâi-lô), Sino-Japanese¹ Go-on and Kan-on (written in orthography and Hepburn), Sino-Korean (written in orthography and the Revised

¹ Sino-Japanese is a linguistic term for the portion of the Japanese vocabulary that is of Chinese origin or makes use of morphemes of Chinese origin (similar to the use of Latin or Greek in English). The same applies to the terms Sino-Korean and Sino-Vietnamese. They do not mean common proto populations.
Romanization) and Sino-Vietnamese (written in orthography), in this fixed order. Their historically attested Old Chinese (OC) rhymes are given according to Wáng (1980) and reconstructively transcribed according to Gāo (2014, p. 79).

The Uralic etyma are based on the relevant etymological dictionaries 1988-UEW and 2001-SSA. Their attested equivalents including contemporary forms and glosses are represented by Estonian, Finnish, Sami\Lappish North\Lule/Inari/Skolt/Kildin (equivalents up to 1989-YSaS; North Sami forms are adjusted according to 1989-SSS), Mordvin, Mari\Cheremis, Udmurt\Votyak, Kom\Zyrian, Khanty\Ostyak, Mansi\Vogul, Hungarian, Nenets\Yurak, Enets\Yen, Nganasan\Tawgi, Selkup and Kamass, in this fixed order. Non-English glosses are translated to English in the present study. Some modifications within Uralic etyma (adding or deleting equivalents) are made and remarked in the present study. Refutations of previously suggested etymological equivalents are given in footnotes.

For the etyma in question, etymological equivalents in other languages claimed by other scholars (mainly Germanic and Tibeto-Burman) are checked according to relevant etymological or comparative works, e.g. 1988-UEW, 1996-CV5ST, 2001-SSA, 2007-EDOC and 2012-EES. Such extended equivalents are mostly cited as in references.

Language reconstructions are listed only for reference reasons. All the attested language data are compared instead of trusting the phonetic and semantic details of reconstructions, because the reconstructions are subject to changes depending on [newly compared] attested linguistic data. Two Old Chinese (OC) reconstructions, OC-W according to Wáng (1980) and OC-Z according to Zhèng-zhāng (2013), are listed. Other reconstructions are quoted from the direct references.

Proto-Sinitic, also known as Proto-Chinese, cannot be compared because it is only a theoretical notion without reconstructed results. Proto-Sino-Tibetan cannot be compared because it is a hypothetical notion without a sufficient amount of etyma representing a sufficient number of the languages in question. Many scholars are still comparing only Tibetan, Burmese or another Tibeto-Burman language to Sinitic (e.g. Shī 2000; Zhang et al. 2019). The works Benedict (1972) and Matisoff (2003) have compared more “Sino-Tibetan languages”, whereas many comparisons do not touch Sinitic. The work 1996-CV5ST compares only five “Sino-Tibetan languages”: Sinitic, Tibetan, Burmese, Jingpo\Kachin and Mizo\Lushai, whereas still many comparisons do not touch Sinitic. Etyma without Sinitic equivalents cannot be labeled as “Sino-Tibetan”. Etyma with equivalents only in one Tibeto-Burman language and Sinitic may be non-genetically diffused (loaned/borrowed) from Sinitic. There is a website called “The Sino-Tibetan Etymological Dictionary and Thesaurus” (https://stedt.berkeley.edu), whereas its content is so far rather a thesaurus (book of synonyms, collection of X-English dictionaries) than an etymological dictionary. This is the current situation of the comparative studies between Sinitic and Tibeto-Burman. Moreover, the Sino-Tibetan hypothesis has been successively criticized (Miller 1974; Beckwith 2002, 2006, 2008; Hé 2004; Guō 2010, p. 21; Zhāng 2012, 2013, 2014; Qú & Jin 2013; Qū 2019). Besides, there are hypotheses for the multiple origins of Sinitic (Lǐ 1990; Schuessler 2003). In sum, the notion Sino-Tibetan cannot be considered as a certain language family which represents Sinitic.

Etymological equivalents are given in orthographies or transcriptions. Equivalents in Western alphabets are given in boldface if they are found in official languages covered by ISO 639-1. Equivalents in Roman alphabets are given in italic. Cyrillic alphabets are transliterated into Roman alphabets according to ISO 9. If a given equivalent word is longer than one morpheme, the targeted morpheme is underlined (if certain). In successive data, dialectal and authorial
variants are separated by a slash (/); grammatical variants are separated by a backslash (\); while lexical variants are separated by a comma (,).

Ancient and fully etymological Chinese etyma (DOMs) are put in the brackets 【】. Ordinary Chinese terms are put in the brackets { } or written without brackets. Double quotation marks ("'") are added when its target is quoted but not agreed. Double arrows (⇒ or ⇐) indicate genetic diffusions ('inherited' in western linguistics; 'born' in Sino-linguistics). Single arrows (→ or ←) indicate non-genetic diffusions ('loaned/borrowed' in western linguistics; 'educated' in Sino-linguistics).

The methods follow traditional etymology (cf. 1662-ELL; Lemon 1783; Rask 1818; Gao 2008) and renewed etymology (cf. Gao 2012-3, 2014, 2017, 2018, 2019a, 2019b, 2020). This study includes also methods of Sino-grammatology (cf. 121-SW; 543-YP; 1008-GY; 1978-82-HJ; 1989-LZ) and Sino-phonology (cf. 1008-GY; 1161-YJ) which are ancient technologies.

Results and discussion

The next etymological paragraphs is:

#Number of etymon 【DOM】 【historical reference: phonetic description】
original gloss 'gloss' (transcribed from);
Mandarin form 'gloss'; Cantonese form 'gloss'; Minnan form 'gloss'; Sino-Japanese form; Sino-Korean form; Sino-Vietnamese form; {OC rhyme group; OC-W reconstruction; OC-Z reconstruction}】
(Read: The Sinitic etymon 【DOM】 with the contents 【…】 is or has been compared (reference) to the Uralic etymon after the equivalents: … (reference).

This etymon has been or not been identified in other languages (reference). This paragraph is used for other language groups compared in other directions (not Sinitic ~ Uralic but Sinitic ~ other or Uralic ~ other) by other scholars.

#1) 【月】 【波文(121-SW); 阿告大陸之精象形(moon)'; 月，月望(543-YP); 阿告大陸之精象形(moon)'; 月望(1008-GY); 阿告大陸之精象形(moon)'; 月望(1161-YJ); 阿告大陸之精象形(moon)'; 月望(1978-82-HJ); 阿告大陸之精象形(moon)'; 月望(1989-LZ); 阿告大陸之精象形(moon)';
velar initial voiced) (ŋ"eiD^); Mandarin yue (üe) 'moon, month'; Cantonese juut6 'moon, month'; Minnan guât/guèh/gêh 'moon, month'; Sino-Japanese Go-on ごち(gochi)/がち(gachi); Kan-on げつ(getsu); Sino-Korean 월(wol); Sino-Vietnamese nguyệt; {OC rhyme 月顔* -ta; OC-W *njuat; OC-Z ***nđ") has been compared (Gao, 2008, p. 231) to the Uralic etymon after the equivalents: Estonian kuu 'moon, month'; Finnish kuu 'moon, month'; Friccine kuhlku 'moon, month'; Mordvin kov 'moon, month'; Khanty/Ostyak Ӱw/лčw 'moon'; Hungarian ʰőhava- 'month', hold 'moon'; Nganasan/Tawgi kɨtada 'moon'; Kamass kɨ 'moon, month'; {Proto-Uralic **kuye'moon, month' (1988-UEW, p. 211)} {← Proto-Sino-Uralic *ŋwe'ta 'moon'}

This etymon has not been identified in other languages.3

This etymon must be aboriginal in Sino-Uralic languages. There are two main reasons:

(1) This DOM is very ancient and already attested in the Oracle Bone Script4 (Figure 1) (1989-LZ, p. 433: e.g. 1978-82-HJ, #7949). The glyph images a crescent moon.

2 Refutation: Previously claimed ←-ŋ is not justified.

3 Refutation: Previously claimed (1988-UEW, p. 211) etymological equation from these Uralic equivalents to Lule Sami kuojiti- 'rise (moon)' is rejected due to semantic inconsistencies. It was already rejected in Aikio (2012, p. 236). Previously claimed (1988-UEW, p. 212) etymological equation from these Uralic equivalents to Yukaghir ƙiŋge 'moon, month' is rejected due to phonetic inconsistencies ←-ŋ is not justified. Previously claimed (Matossif 1980, p. 20) etymological equation from Sinitic to Angami Kohima themvǎ 'star'; Chakeshem themvǎ 'star'; Konkay sha-nušvja-hu 'star'; Mao ouv 'star'; Lotha shanting 'star'; Meluri ouvchi 'star'; Ntenyi ouvchi 'star'; Maring souwng 'star'; Sangtam chimph 'star'; Lahu mɔ̀-kà 'star' is rejected due to semantic and phonetic inconsistencies. Previously claimed (LaPolla 1987, p. 25) etymological equation from Sinitic to Proto-Tibeto-Burman **-s-ngwat 'star-month'; Dulong guw55 met55 i gu31 nye55; Angami Naga themvǎ; Lahu mɔ̀-kà; Motuo Membah karmi is rejected due to phonetic inconsistencies.
(2) It is a certain Sino-Uralic etymon supported by a rhyme correspondence consisting of two etyma (see Table 1 in the next section) and an onset correspondence consisting of three etyma (see Table 2 in the next section).

The following reinforced etyma are studied in order to form regular sound correspondences with the etymon #1.

#2) 【別】《說文(121-SW): 分解也(divide); 玉篇(543-YP): 分別也(separate); 唐韻(1008-GY): 分娩切 又被切(Chi bo'切包解也(different, separate, divide)); 唐韻(1161-YJ): 外轉第二十三等三等入聲部 月辯切(outbound, final-23, labialized-., division-1, tone-D, labial initial voiced))(beat); 漢語(1161-YJ): 離也 異也 解也 解也

Mandarin biē ‘other’; Cantonese bī6 ‘other’; Minnan piât-pât ‘other’; Sino-Japanese Go-on べち (bechi); Kan-on わつ(hetsu); Sino-Korean 열(byeol); Sino-Vietnamese biệt; {OC rhyme 月韻 *-ta; OC-W *bjat; OC-Z ‘*bred’} 】 is compared to the Uralic etymon after the equivalents: Estonian muu ‘other’; Finnish mū/muu ‘other’; Mari/Cheremis mól/mölo ‘other’; Udmurt/Votyak mǐ/d/mś ‘other’; Comi/Zyrian med/mimid ‘other’; Mansi/Vogul mátmót/móm ‘second, different’; Hungarian más ‘other’; {“Proto-Finn-Ugric” *mu another, ‘that’ (this, that)” (1988-UWE, p. 281)}. ← Proto-Sino-Uralic #mbeta ‘other’

This etymology has been identified in other languages: Tibetan phjed ‘half’; Burmese phrat ‘cut in two, chop off’; Jingpho/Kachin phjat/phrat ‘cut, severe’ (1996-CV5ST; Sinitic – Tibetan, Burmese, Kachin; 2007-EDOC, p. 167; Sinitic – Burmese prat ‘be cut in two’). 5 { ← Sinitic} These equivalents should be non-genetically diffused (loaned/borrowed) from Sinitic, because their phonetic diversity is low. It is a sign of recent occurrence (without differentiation through history). In contrast to that the same etymology exhibits different onsets within Sinitic /b/ in 1161-YJ and Sino-Japanese Go-on; /p/ in Mandarin, Cantonese and Minnan; /b/ in Sino-Japanese Kan-on), but Tibetan, Burmese and Jingpho/Kachin have the same onset /p/. An etymon exhibits diversity in Germanic languages, while its equivalents borrowed from English are very similar in their target languages in the world.

#3) 【言】《說文(121-SW): 有言曰言(speech); 五音(543-YP): 言語也(speech); 唐韻(1008-GY): 漢語(1161-YJ): 外轉第二十一等三等平聲語音切(outbound, final-21, labialized-, division-3, tone-C, velar initial voiced) 》(ian); Mandarin yán (ián) ‘speech’; Cantonese jin4 ‘speech’; Minnan giân/gân ‘speech’; Sino-Japanese Go-on ごん (gon); Kan-on げん (gen); Sino-Korean 조언 (eun); Sino-Vietnamese Ngân; {OC rhyme 元韻 *-na; OC-W *iían; OC-Z ‘*iían”} 】 has been compared (Gao, 2018, p. 78) to the Uralic etymon after the equivalents: Estonian keelli(ke) ‘tongue, language’; Finnish kielik为此 ‘tongue, language’; Sami/Lappish giella/kiella/kielal

1 Oracle Bone Script is the discovered writing system used in the Shang Dynasty (ca. 1600 — 1046 B.C.E.) [So far the oldest Oracle Bone Script is excavated from the Erligang Culture (ca. 1510 — ca. 1460 B.C.E.) (Zheng, 2008, p. 80). However, the Shang Empire’s earlier remains (scripts and other materials) have not been discovered or confirmed]. It was recognized as being ancient Chinese writing by Wang Yi-rong 王禮容 in 1899. Liu Li 劉獻 compiled and published the first collection of 1,058 rubbings including some interpretations of some unearthed scripts in 1903. In English, it was introduced as “inscriptions upon bone and tortoise shell” by Frank H. Chalfant (1906, p. 30). Wang Guangwei 王國維 (1916) demonstrated that the commemorative cycle of the Shang emperors matched the list of emperors in Sima Qian’s Records of the Historian. Other important leading scholars are for example Guo Mō-nou 郭沫若 (chief editor of 1978-82-HJ), Yu Xin-wu 于省吾 (chief editor of 1996-GL), and Yao Xiao-sui 姚孝隧 (chief editor of 1989-LZ). The work 1989-LZ is to date the largest collection of oracle bones. It contains 41,956 rubbings (without graphic-eytological equations). The work 1989-LZ is a primary academic reference book. It sorts lexical terms and identifies their graphic-eytological equations to transmitted Chinese etyma (DOMs) (without interpreted glosses). The work 1989-GL is a secondary academic reference book. It collects interpreted glosses of the terms by many scholars. For recent works in English, see Takashima (2010), Keightley (2014) and Pankenier (2015).

5 REFUTATION: Previously claimed (1988-UWE, p. 281) etymological equation from these Uralic equivalents to Sami/Lappish nubbli/nubbil/nubbil/nunnli/nunnli/p the other, another is rejected due to phonetic inconsistencies.
Overview

The etyma #1 【月】 and #2 【別】 form a rhyme correspondence (Table 1).

Table 1. Rhyme correspondence (Re#2020JGaoTT-2310-2144-41): Old Chinese rhyme 月 韵 *-ta ↔ Minnan -at ← Estonian/Finnish -uu

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>Minnan</th>
<th>Cantonese</th>
<th>Minnan</th>
<th>Estonian</th>
<th>Finnish</th>
<th>North Samsi</th>
<th>Hungarian</th>
</tr>
</thead>
<tbody>
<tr>
<td>【月】</td>
<td>wè (wy)</td>
<td>jyut6</td>
<td>guài</td>
<td>kuu</td>
<td>kuu</td>
<td>--</td>
<td>hóóhava-</td>
</tr>
<tr>
<td></td>
<td>moon, month</td>
<td>month</td>
<td>month</td>
<td>month</td>
<td>month</td>
<td>--</td>
<td>month</td>
</tr>
<tr>
<td>【別】</td>
<td>biè</td>
<td>bì6</td>
<td>pièt</td>
<td>muu</td>
<td>muu</td>
<td>--</td>
<td>máš</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>other</td>
<td>other</td>
<td>other</td>
<td>other</td>
<td>--</td>
<td>other</td>
</tr>
</tbody>
</table>

The etyma #1 【月】, #3 【言】 and #4 【岸】 form an onset correspondence (Table 2).

Table 2. Onset correspondence (Oc#2020JGaoTT-2310-2144-42): Mandarin 0- ↔ Cantonese j- ↔ Minnan g- ↔ Estonian/Finnish k-

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>Mandarin</th>
<th>Cantonese</th>
<th>Minnan</th>
<th>Estonian</th>
<th>Finnish</th>
<th>North Samsi</th>
<th>Hungarian</th>
</tr>
</thead>
<tbody>
<tr>
<td>【月】</td>
<td>Qiè (xié)</td>
<td>jyut6</td>
<td>guài</td>
<td>kuu</td>
<td>kuu</td>
<td>--</td>
<td>hóóhava-</td>
</tr>
<tr>
<td></td>
<td>moon, month</td>
<td>month</td>
<td>month</td>
<td>month</td>
<td>month</td>
<td>--</td>
<td>month</td>
</tr>
<tr>
<td>【言】</td>
<td>Qián (tián)</td>
<td>jīn4</td>
<td>jiān</td>
<td>keel(e)</td>
<td>kielíkíle-</td>
<td>giella</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>speech</td>
<td>speech</td>
<td>speech</td>
<td>tongue, language</td>
<td>tongue, language</td>
<td>language</td>
<td></td>
</tr>
<tr>
<td>【岸】</td>
<td>Qán</td>
<td>ngon6</td>
<td>gān</td>
<td>kallas/kalda</td>
<td>kallas/kallda-</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>shore</td>
<td>shore</td>
<td>shore</td>
<td>shore</td>
<td>shore</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

9 Refutation: Previously claimed (1988-UEW, p. 144) etymological equation from these Uralic equivalents to Nenets je 'tongue'; Enets siodol'stora 'tongue'; Nganasan sieja 'tongue'; Selkup: kielíkíle 'tongue'; Kamass šoka 'tongue'; Mator kašte 'tongue' is rejected due to phonetic inconsistencies. These equivalents were already questioned in 1977-FUV (p. 45). Previously claimed (1996-CVST) etymological equation from Sinitic to Jinghō/Kachin gón 'be pleasant, agreeable' is rejected due to semantic inconsistencies.

10 Refutation: Previously claimed (1996-LÄGLOS, p. 20) etymological equation from Finnic to Old Norse halte 'slope, rock'; [Proto-Germanic gálpa-z 'slope; be inclined'] is rejected due to semantic inconsistencies. Previously claimed (2007-EDOC, p. 151) etymological equation from Sinitic to Tibetan dgo 'shore, bank' is rejected due to phonetic inconsistencies.
What is the essence of Sino-Uralic? Gão (2014, p. 37, p. 51) has introduced it as a proto population in Neolithic China, which should be correlated to the Yandi Shennong nation (炎帝/神農氏) in Chinese pre-history and the Human Y-Chromosome DNA haplogroup N-M231.

We hope that the present study does not only contribute to the domains of linguistics, but also to the domain of archaeoastronomy. Astronomical terms can be very ancient and widely diffused (cf. Gao, 2019a, on the term for 'sky' in Sino-Uralic with extensions to many Indo-European languages; Gao, 2020, on the term for 'Jupiter, year' in Sino-Uralic with extensions to many Indo-European languages). What is the essence of the etymological diffusion beyond Sino-Uralic? It should be discussed in the future. It can be complexly a shared genesis or simply a shared cultural heritage.

Conclusions

Using etymological methods, the present study has researched four Sinitic and Uralic shared etymologies (etyma). Two of them form a rhyme correspondence. Three of them form an onset correspondence. These regular sound changes validate the genetic connection between Sinitic and Uralic. The Sino-Uralic term for 'moon, month' is among these four etyma. It is demonstrated that this term should be aboriginal in Sino-Uralic languages.

References

121-SW – Xu, Shen [許慎. Shuō wén jiè zì |說文解字. Luoyang |雒陽, Han |漢 (China), 121.


1008-GY – Chen, Peng-niàn děng |陳彭年等. Dà sòng cóng xiū guǎng yǔn |大宋重修廣韻. Dongjing |東京 (Kaifeng), Song |宋 (China), 1008.

1161-YJ – Zhang, Lin-zhi (jiào) |张麟之校. Yùn jìng |韻鏡. Lin'an |臨安 (Hangzhou), Song |宋 (China), 1161.


