On Etymology of Finnic Term for 'Sky'

Jingyi Gao

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

Abstract

Using etymological methods, the present study has identified five Sinitic and Uralic shared etymologies. These five etymologies form a rhyme correspondence. This regular sound change validates the etymological connection between Sinitic and Uralic. The Finnic term for 'sky' is among these five etymologies. It is demonstrated that this word root should be aboriginal in Sino-Uralic languages.

Keywords: Rhyme correspondence, Sinitic, Uralic, Sino-Uralic, Baltic, Germanic, Celtic, Italic, Indo-Iranian.

Introduction

The Finnic term for 'sky' (Estonian *taevas*; Finnish *taivas*; Livonian *tōvaz*; Veps *taivaz*; Votic *taivas*) has no cognate in other Uralic languages, therefore it has been previously considered a loanword to Finnic from Indo-Iranian (Schott, 1849, p. 126), from Baltic (Thomsen, 1869, p. 34, 73), or from Germanic (Koivulehto, 1972). The present study finds that this Finnic word has cognates in Sinitic languages supported by a deep rhyme correspondence consisting of five etymologies; therefore this word root must be aboriginal in Sino-Uralic languages.

Gao (e.g. 2005, 2014b, 2019; Gāo, 2008) detected and identified Sinitic and Uralic shared etymologies, and has solely researched Sinitic and Uralic shared etymologies for more than a decade. We could infer a general skepticism about this approach. Several unsound language comparisons, e.g. Finnish with Basque, Finnish with Dravidian, Proto-Uralic with Proto-Indo-European, have become tedious. It was difficult to clarify how the Sino-Uralic comparison is extraordinarily significant. Gao (2014b) turned the focus to a solid demonstration of regular sound correspondences (rhyme and onset correspondences) between Sinitic and Uralic, and included long discussions on the methodology. In light of this, one can realize that the Sino-Uralic etymological studies are distinguished from those unsound language comparisons. The general direction of Sino-Uralic affinity studies should be acknowledged.

A rhyme correspondence is a strict and composite rule of interlinguistic sound correlations. A rhyme correspondence achieves that not only a single phoneme but also a composite rhyme (the -VCv part of a CVCv morpheme) is consistently correlated among related language varieties. The first rhyme correspondences between the Sinitic and Uralic languages have been demonstrated on the example of the Finnish *-ala* and *-aja* rhymes with ten etymologies (Gao, 2014b). A total of ten rhyme correspondences with 32 etymologies has been published (Gao, 2019). The present study adds another rhyme correspondence with five etymologies.

Materials and methods

The Sinitic language family is compared to Uralic language family.

The Sinitic etymologies (etyma) are guided by Chinese etyma (DOMs), which are historically attested glyphs. Their historical glosses are cited from the first two Chinese classical dictionaries (121-SW and 543-YP). Their historical phonological features are cited from the work 1161-YJ. Their other attested equivalents including contemporary forms and glosses are represented by Mandarin [Beijing Yan] (Pinyin), Cantonese [Guangzhou Yue] (Jyutping) and Minnan [Taipei Min] (Tâi-lô). English glosses are made in the present study.

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 (represented by North, Lule, Inari, Skolt, Kildin Sami; North Sami forms are adjusted according to 1989-SSS), Mordvin, Mari\Cheremis, Udmurt\Votyak, Komi\Zyrian, Khanty\Ostyak, Mansi\Vogul, Hungarian, Nenets\Yurak, Enets\Yen, Nganasan\ Tawgi, Selkup and Kamass. Some modifications within Uralic etyma (adding or deleting equivalents) are made in the present study. Non-English glosses are translated to English in the present study.

Etymological equivalents in some other languages (mainly Indo-European, Tibeto-Burman, suggested by other scholars) are checked according to the relevant etymological dictionaries 1959-IEW, 1988-UEW, 1996-CV5ST, 2001-SSA, and 2007-EDOC.

Reconstructed forms are added only for reference reasons. All the attested forms are compared instead of trusting the phonetic and semantic details of reconstructions, because the reconstructions are subject to changes depending on attested linguistic data. Two Old Chinese reconstructions, OC-W (according to Wáng, 1980) and OC-Z (according to Zhèng-zhāng, 2013), are added. Uralic reconstructions are cited 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 languages. Moreover, the Sino-Tibetan hypothesis has been successively criticized (Miller, 1974; Beckwith, 2002, 2006, 2008; Hé, 2004; Guō, 2010, p. 21; Zhāng, 2012). Besides, there are hypotheses for the multiple origins of Sinitic (Lǐ, 1990; Schuessler, 2003).

Sinitic words are given in orthographies (in italic). Non-Sinitic words are given in orthographies (in boldface [if it is found in an official language covered by ISO 639-1] and italic) or transcriptions (in italic, mainly the Uralic Phonetic Alphabets). All the given Sinitic words are monomorphemic. If a given non-Sinitic word is longer than one morpheme, the targeted morpheme is underlined (if certain). In Finnic, conditionally apocoped phonemes are given in uppercase. 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 (,).

The methods follow traditional etymology (cf. Rask, 1818) and renewed etymology (cf. Gao, 2014a, 2014b, 2017, 2019).

Results and discussion

#1)【晝】〖說文(121-SW): 日之出入與夜為界('day'); 玉篇(543-YP): 日正中('noon'); 韻鏡 (1161-YJ): 開三等(open 3rd division); Mandarin *zhòu* 'day'; Cantonese *zau3* 'day'; Minnan *tiù* 'day'; {OC rhyme 幽韻 *-wə; OC-W *tiu; OC-Z *tus}〗 (Read: The Sinitic etymon 【 】 after

〖...〗) is compared to the Uralic etymon after the equivalents: Estonian *taevas* 'sky'; Finnish *taivas* 'sky'.

This etymon has been identified in other languages: (Germanic) Danish *Tyr* 'the god Tyr'; Swedish *Tyr* 'the god Tyr'; Norwegian *Ty* 'the god Tyr'; Icelandic *Týr* 'the god Tyr'; Old Norse *Týr* 'the god Tyr'; Old English *Tīw* 'the god Tyr'; Old High German *Ziu* 'the god Tyr'; Gothic *teiws* 'the god Tyr'; (Baltic) Latvian *dìevs* 'god'; Lithuanian *diẽvas* 'god'; Old Prussian *dēiwas* 'god'; (Celtic) Irish *dia* 'god'; Old Irish *día* 'god'; Welsh *duw* 'god'; Old Welsh *duiu* 'god'; (Italic) Latin *deus*, *dīvus* 'god', *dies* 'day'; French *dieu* 'god'; Italian *dio* 'god', *dì* 'daytime'; Spanish *dios* 'god', *dia* 'day'; Portuguese *deus* 'god', *dia* 'day'; Romanian *zeu* 'god', *zi* 'day'; (Ancient) Greek Ζεύς (*Zeús*) 'the god Zeus'; Hittite *sius* 'god'; Armenian տիվ (*tiv*) 'day, daytime'; Old Armenian տիւ (*tiw*) 'day, daytime'; (Indo-Iranian) Sanskrit देव (*devá*) 'god', द्यु (*dyú*) 'sky, day'; Iranian Persian *div* 'demon'; {Proto-Indo-European **deywós* 'god', **dyéws* 'sky'} (← Sino-Uralic).[1]

This etymon must be spread from Sino-Uralic to Indo-European (not from Indo-European to Sino-Uralic as previously claimed). There are three main reasons:

(1) Its meaning is more substantial in Sino-Uralic: 'day' in Sinitic, Italic, Armenian, Indo-Iranian ← 'sky' in Finnic and Indo-Iranian → 'god' Baltic, Celtic, Italic, Hittite, Indo-Iranian → 'a God's name' in Germanic and Greek. There should be first a substantial term for 'sky' and then the unsubstantial terms for 'day' and 'god'. It is not logical to suggest that Finnic loaned a term for 'god' to name the sky.

(2) This etymology is absent from Slavic. Its existence in Baltic could be spread from Finnic. Its existence in other Indo-European groups could be spread from neighboring language groups. Its existence in Sinitic gives this etymology much deeper aboriginality in Sino-Uralic. This term is already attested in the oracle bone scripts dating to ca. 1500 B.C.E. (see Figure 1)

Figure 1. Attested form of 【晝】 in the oracle bone script.

(3) It is a certain Sino-Uralic etymology supported by a rhyme correspondence consisting of five etymologies. The other four etymologies are following:

#2)【求】〖說文(121-SW): 索也('demand'); 韻鏡(1161-YJ): 開三等(open 3rd division); Mandarin *jiū* 'investigate'; Cantonese *gau3* 'investigate'; Minnan *kiù* 'investigate'; {OC rhyme 幽韻 *-wə; OC-W *giu; OC-Z *gu}〗 is compared to the Uralic etymon after the equivalents: Estonian *kaebaS-* 'accuse'; Finnish *kaipaS-* 'yearn for'; Sami/Lappish *gáibidi-*/*kaiˈpēti-*/*käjbidi-*/*kạịв̄ē-*/-- 'demand'. This etymon has not been identified in other languages.[2]

#3)【究】〖說文(121-SW): 窮也('low'); 玉篇(543-YP): 深也('deep'); 韻鏡(1161-YJ): 開三等(open 3rd division); Mandarin *jiū* 'investigate'; Cantonese *gau3* 'investigate'; Minnan *kiù* 'investigate'; {OC rhyme 幽韻 *-wə; OC-W *kiu; OC-Z *kus}〗 is compared to the Uralic

---

[1] REFUTATION: Previously claimed (Bodman, 1980, p. 172; 2007-EDOC, p. 624) etymological equivalent Written Tibetan *gdugs* 'midday, noon' is rejected due to phonetic inconsistencies.

[2] REFUTATION: Previously claimed (Koivulehto, 1970; 2001-SSA, vol. 1, p. 279) etymological equivalents after Proto-Germanic **kaujan*; Old English *cīegan*, *cīgan* 'to call'; Old High German *(gi)kewen* 'to call' are rejected due to phonetic and semantic inconsistencies. Previously claimed (1996-CV5ST) etymological equivalents after Tibetan *sko*, *bsko* 'to choose', *go* 'to know, understand' are rejected due to phonetic and semantic inconsistencies.

etymon after the equivalents: Estonian ***kaevaS-*** 'dig'; Finnish ***kaiva-*** 'dig'; Sami\Lappish ***goaivu-***/*kåi* '*vō-*/*koajvu-*/*kŏ̆ăī̆vɐ-*/*kŭ̆aī̆vɐ-* 'dig/scoop'; Mordvin *kojmä*/*kojme* 'scoop'; Mari\ Cheremis *koe-*/*kue-* 'shovel'; Nenets\Yurak *śīwa* 'shovel'; Enets\Yen *sea* 'spade'; Nganasan *kaibu* 'spade'; Kamass *kō* 'spade'; {Proto-Uralic **kojwa-* 'dig, scoop' (1988-UEW, p. 170)}. This etymon has not been identified in other languages.[3]

#4)【流】〖說文(121-SW): 水行也('flow'); 玉篇(543-YP): 說文水行也('flow(121)'); 韻鏡 (1161-YJ): 開三等(open 3rd division); Mandarin *liú* 'flow'; Cantonese *lau4* 'flow'; Minnan *liû* 'flow'; {OC rhyme 幽韻 **-wə*; OC-W **liu*; OC-Z **ru*}〗 is compared to the Uralic etymon after the equivalents: Estonian ***laevA*** 'ship'; Finnish ***laiva*** 'ship'; North Sami ***láivi*** 'vessel, ship'.

This etymon has been identified in other languages: Latvian ***laiva*** 'ship'; Lithuanian ***laĩvas*** 'ship'; {Proto-Baltic **laiw-* 'ship'} (← Finnic).[4]

#5)【否】〖說文(121-SW): 不也('negate'); 玉篇(543-YP): 可否('negate'); 韻鏡(1161-YJ): 開三等(open 3rd division); Mandarin *fǒu* 'negate; trouble'; Cantonese *fau4*/*pei2* 'negate; trouble'; Minnan *phí* 'negate; trouble'; {OC rhyme 之韻 **-ə*; OC-W **piə*; OC-Z **pɯʔ*}〗 is compared to the Uralic etymon after the equivalents: Estonian ***vaevA*** 'trouble'; Finnish ***vaiva*** 'trouble'; Sami\Lappish ***váivi***/*vai* '*vē*/*väjvi*/*vɐ̝ī̆v̆ᵉ*/*vɐ̝ī̆vᵉ* 'trouble'. This etymon has not been identified in other languages.[5]

Overview

Aforementioned five Sino-Uralic etymologies form a regular sound change which is a rhyme correspondence: the Old Chinese rhyme 幽韻 (**-wə*) is correlated with the rhyme *-ou* in Mandarin, *-au* in Cantonese, *-iu* in Minnan, *-aeva* in Estonian, *-aiva* in Finnish, *-aivi* in North Sami (see Table 1).

Table 1. Rhyme Correspondence: Old Chinese 幽韻 **-wə*[6] : Mandarin *-ou* : Cantonese *-au* : Minnan *-iu* : Estonian *-aeva* : Finnish *-aiva* : North Sami *-aivi*

DOM	Mandarin	Cantonese	Minnan	Estonian	Finnish	North Sami
【晝】	*zhòu*	*zau3*	*tiù*	***taevas***	***taivas***	--
	說文: 日之出入與夜為界('day')			'sky'	'sky'	
【求】	*qióu*(*qiú*)	*kau4*	*kiû*	***kaebaS-***	***kaipaS-***	***gáibidi-***
	說文: 索也('demand')			'accuse'	'yearn for'	'demand'
【究】	*jiōu*(*jiū*)	*gau3*	*kiù*	***kaevaS-***	***kaiva-***	***goaivu-***
	玉篇: 深也('deep')			'dig'	'dig'	'dig'
【流】	*lióu*(*liú*)	*lau4*	*liû*	***laevA***	***laiva***	***láivi***
	說文: 水行也('flow')			'ship'	'ship'	'vessel, ship'
【否】	*fǒu*/*pǐ*	*fau4*/*pei2*	*phi*	***vaevA***	***vaiva***	***váivi***
	說文: 不也('no')			'trouble'	'trouble'	'trouble'

---

[3] REFUTATION: Previously claimed (2007-EDOC, p. 320) etymological equivalents after Proto-Lolo-Burmese **N-gu¹*, Burmese *ku* 'help', Lahu *gu* 'prepare, practice' are rejected due to phonetic and semantic inconsistencies.

[4] REFUTATION: Previously claimed (Koivulehto, 1970; 2001-SSA, vol. 2, p. 39) etymological equivalents after Proto-Germanic **flauja*; Old Norse *fley* 'ship' are rejected due to phonetic inconsistencies. Previously claimed (1996-CV5ST) etymological equivalents after Tibetan *ãkhru*, *ãkhrud* 'to wash, to bathe', *khrus* 'bath, washing'; Burmese *khjᴐwh* (Old Burmese *khuiw*) 'to wash, bathe'; Kachin *khrut2* 'to bathe, wash'; Lushai *ṭhuaʔ* 'to wash or rinse (as inside of bottle)' are rejected due to phonetic and semantic inconsistencies.

[5] REFUTATION: Previously claimed (Thomsen, 1869, p. 157; 2001-SSA, vol. 3, p. 394) etymological equivalents after Proto-Germanic **waiwan-* 'pain'; Old High German *wēwo*; English *woe* are rejected due to phonetic and semantic inconsistencies.

[6] In this rhyme correspondence, the fifth etymon belongs to Old Chinese rhyme 之韻 **-ə*.

This is a deep rhyme correspondence with 5 etyma. It is substantially evidential. Its coincidental probability is as low as 1/324,666,368. The first etymon with a certain rhyme (1) * the first etymon has comparable onsets (1/4 [There are four sorts of onsets: labial, coronal, dorsal and laryngeal.]) * the second etymon falls into the same Old Chinese rhyme group (1/29 [There are 29 rhyme groups in Old Chinese.]) * the second etymon has comparable onsets (1/4) * the third etymon falls into the same Old Chinese rhyme group (1/29) * the third etymon has comparable onsets (1/4) * the fourth etymon falls into the same Old Chinese rhyme group (1/29) * the fourth etymon has comparable onsets (1/4) * the fifth etymon falls into the same Mandarin rhyme group (1/13 [There are 13 rhyme groups in Mandarin.]) * the fifth etymon has comparable onsets = 1 * 1/4 * 1/29 * 1/4 * 1/29 * 1/4 * 1/29 * 1/4 * 1/13 * 1/4 = 1/324,666,368.

Conclusions

Using etymological methods, the present study has identified five Sinitic and Uralic shared etymologies. These five etymologies form a rhyme correspondence. This regular sound change validates the etymological connection between Sinitic and Uralic. The Finnic term for 'sky' is among these five etymologies. It is demonstrated that this word root should be aboriginal in Sino-Uralic languages.

References

121-SW – Xǔ, Shèn 許慎. *Shuō wén jiě zì* 說文解字. Luoyang 雒阳, 121.

543-YP – Gù, Yě-wáng 顧野王. *Yù piān* 玉篇. Nanjing 建康, 543.

1161-YJ – Zhāng, Lín-zhī (jiào) 张麟之校. *Yùn jìng* 韻鏡. Hangzhou 臨安, 1161.

1959-IEW – Pokorny, Julius. *Indogermanisches etymologisches Wörterbuch*. Bern: Francke, 1959.

1988-UEW – Rédei, Károly. *Uralisches etymologisches Wörterbuch* (*Band I–II*). Budapest: Akadémiai Kiadó, 1988.

1989-SSS – Sammallahti, Pekka. *Sámi-suoma sátnegirji / Saamelais-suomalainen sanakirja*. Ohcejohka (Utsjoki): Jorgaleaddji, 1989.

1996-CV5ST – Peiros, Ilia; Starostin, Sergei. *A Comparative Vocabulary of Five Sino-Tibetan Languages*. Parkville (Victoria): University of Melbourne, Department of Linguistics and Applied Linguistics, 1996.

2001-SSA – Itkonen, Erkki; Kulonen, Ulla-Maija. (päätoimittajat). *Suomen sanojen alkuperä: Etymologinen sanakirja* (1/2/3). Helsinki: Suomalaisen Kirjallisuuden Seura; Kotimaisten Kielten Tutkimuskeskus, 2001.

2007-EDOC – Schuessler, Axel. *ABC Etymological Dictionary of Old Chinese*. Honolulu: University of Hawai'i Press, 2007.

Beckwith, 2002 – Beckwith, Christopher I. The Sino-Tibetan problem. In: *Mediaeval Tibeto-Burman Languages*. Leiden et al.: Brill, 2002; pp. 113–158.

Beckwith, 2006 – Beckwith, Christopher I. Old Tibetan and the dialects and periodization of Old Chinese. In: *Medieval Tibeto-Burman Languages II*. Leiden et al.: Brill, 2006; pp. 179–200.

Beckwith, 2008 – Beckwith, Christopher I. Old Chinese loans in Tibetan and the non-uniqueness of "Sino-Tibetan". In: *Medieval Tibeto-Burman Languages III*. Leiden et al.: Brill, 2008; pp. 161–201.

Bodman, 1980 – Bodman, Nicholas C. Chinese and Sino-Tibetan: Evidence towards establishing the nature of their relationship. In: *Contributions to Historical Linguistics*. Leiden: E. J. Brill; pp. 34–199.

Gao, 2005 – Gao, Jingyi. *Comparison of Swadesh 100 Words in Finnic, Hungarian, Sinic and Tibetan*. Tallinn: Estonian Language Foundation, 2005.

Gāo, 2008 – Gāo, Jīng-yī 高晶一. *Hàn-yǔ yǔ Běi-ōu yǔ-yán* 汉语与北欧语言. Beijing 北京: China Social Sciences Press 中国社会科学出版社, 2008.

Gao, 2014a – Gao, Jingyi. Huns and Xiongnu identified by Hungarian and Yeniseian shared etymologies. *Central Asiatic Journal* 2012/2013 [2014], *56*, 41–48.

Gao, 2014b – Gao, Jingyi. Rhyme correspondences between Sinitic and Uralic languages: On the example of the Finnish -ala and -aja rhymes. *Linguistica Uralica* 2014, *50*, 2, 94–108.

Gao, 2017 – Gao, Jingyi. Xia and Ket identified by Sinitic and Yeniseian shared etymologies. *Central Asiatic Journal* 2017, *60*, 51–58.

Gao, 2019 – Gao, Jingyi. Veel hiina ja soome-ugri keelte ühiseid etümoloogiaid riimivastavuse näitel. *Idakiri: Eesti Akadeemilise Orientaalseltsi aastaraamat* 2018 [2019], 71–80.

Guō, 2010 – Guō, Xī-liáng 郭錫良. *Hàn-zì gǔ-yīn shǒu-cè* 漢字古音手册. Beijing 北京: The Commercial Press 商務印書館, 2010.

Hé, 2004 – Hé, Jiǔ-yíng 何九盈. Hàn-yǔ hé qīn-shǔ yǔ-yán bǐ-jiào yán-jiū de jī-běn yuán-zé 汉语和亲属语言比较研究的基本原则. *Yǔ-yán-xué lùn-cóng* 语言学论丛 2004, *29*, 12–66.

Koivulehto, 1970 – Koivulehto, Jorma. Suomen laiva-sanasta. *Virittäjä* 1970, *74*, 2, 178–183.

Koivulehto, 1972 – Koivulehto, Jorma. Germanisch-Finnische Lehnbeziehungen II: Wirtschaftliches und Landwirtschaftliches. *Neuphilologische Mitteilungen* 1972, *73*, 1/3, 575–628.

Lǐ, 1990 – Lǐ, Bǎo-jiā 李葆嘉. Shì-lùn yuán-shǐ huá-xià-yǔ de lì-shǐ bèi-jǐng 试论原始华夏语的历史背景. *Yǔ-yán-xué tōng-xùn* 语言学通讯 1990, 1/2, 1–2.

Miller, 1974 – Miller, Roy A. Sino-Tibetan: Inspection of a conspectus. *Journal of the American Oriental Society* 1974, *94*, 2, 195–209.

Rask, 1818 – Rask, R[asmus] K[ristian]. *Undersögelse om det gamle Nordiske eller Islandske Sprogs Oprindelse*. Kjöbenhavn: Gyldendal, 1818.

Schott, 1849 – Schott, Wilhelm. *Über das Altai'sche oder Finnisch-Tatarische Sprachengeschlecht*. Berlin: bei G. Reimer, 1849.

Schuessler, 2003 – Schuessler, Axel. Multiple origins of the Old Chinese lexicon. *Journal of Chinese Linguistics* 2003, *31*, 1, 1–71.

Thomsen, 1869 – Thomsen, Vilhelm. *Den gotiske sprogklasses indflydelse på den finske*. København: Gyldendal, 1869.

Wáng, 1980 – Wáng, Lì 王力. *Shī-jīng yùn dú* 诗经韵读. Shanghai 上海: Shanghai Press of Classics 上海古籍出版社, 1980.

Zhāng, 2012 – Zhāng, Mín-quán 张民权. Hàn-zàng tóng-yuán jiǎ-shuō yǔ gǔ-yīn yán-jiū zhōng de ruò-gān wèn-tí 汉藏同源假说与古音研究中的若干问题. *Shān-xī dà-xué xué-bào (zhé-xué-shè-huì-kē-xué bǎn)* 山西大学学报（哲学社会科学版） 2012, 9, 10–17.

Zhèng-zhāng, 2013 – Zhèng-zhāng, Shàng-fāng 郑张尚芳. *Shàng-gǔ yīn-xì* 上古音系. Shanghai 上海: Shanghai Education Press 上海教育出版社, 2013.