

JTC1/SC2/WG2 N3694
2009-10-05

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Rationale for Encoding Latin Letter Middle Dot
Source: Andrew West
Status: Individual Contribution
Date: 2009-10-05

1. Background

Document N3567 "Proposal to encode a Middle Dot letter for Phags-pa transliteration" proposed the encoding of a middle dot letter for transliteration and phonetic transcription purposes, and this character was unanimously accepted for encoding at Meeting 54 (see Resolution M54.13). Recently SEI have submitted document N3678, which suggests that there is no need to encode a middle dot letter, as the proposed usage is already catered for by U+00B7 MIDDLE DOT or U+02D1 MODIFIER LETTER HALF TRIANGULAR COLON. We believe that the author of N3678 has focussed too narrowly on one particular area of usage, that of Americanists, and does not fully appreciate the semantically different usages that the middle dot has amongst different user groups, and the different requirements of these user groups. The present document analyses the different usages of middle dot, and attempts to demonstrate that neither U+00B7 MIDDLE DOT nor U+02D1 MODIFIER LETTER HALF TRIANGULAR COLON are appropriate characters to represent the use of middle dot by Orientalists as a phonetic letter representing a glottal stop.

2. Middle Dot Use Cases

It is important to clearly distinguish the different uses to which the middle dot (frequently also referred to as a "raised dot") is put to, and to determine which is the most appropriate encoded character to represent each of these usages. We suggest that there are four main use cases for the raised or middle dot within the context of the Latin script, and that these four use cases represent semantically distinct *abstract characters* which merely share the same glyph shape.

A. Interword Separation Mark

This is the oldest usage of the middle dot in the Latin script, dating back to the Roman period when early Latin inscriptions often used a middle dot, commonly referred to as an *interpunct*, to separate words. This usage is catered for by U+2E31 WORD SEPARATOR MIDDLE DOT [Po].

B. Intraword Separation Mark

The most important and widespread modern use of middle dot is as an intraword separation mark (what might be termed an *intrapunct*), for example in dictionaries to separate words into their constituent syllables (see Fig.1), or in linguistic texts to separate a word into its grammatical parts (see Fig.2). The middle dot is also used in Catalan and Occitan to indicate that two adjacent letters do not form a single grapheme, but belong to adjacent syllables (e.g. *col·lecció* "collection" in Catalan, and *des·har* "to undo" in Occitan).

Fig.1 : Humez & Humez 2008 page 201

down to the present day: The middle dot is still used by lexicographers to show syllable boundaries, as in this entry from *The American Heritage Dictionary* (4th ed.): *syl·lab·i·fy* (sĭ-lăb'ĭ-fi) or *syl·lab·i·cate* (-kăt'). Compare *The Random House Dictionary of the English Language: The Unabridged Edition* (1971): *syl·lab·i·fy* (si lab' ə fi).

Fig.2 : Gardiner 1957 page 28


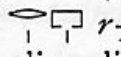
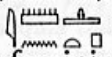

OBS. 3. Suffixes which are, or once have been, independent words are in this book preceded by a dot, ex.  *sdm.twf* 'he is heard', originally 'one (*tw*) hears (*sdm*) him (*f*)'. A hyphen is used only in compound words, exx.  *r-pr* 'temple';  *Imm-hotp* 'Amenhotpe', a man's name. Grammatical endings like the *t* of the feminine are not marked as such in transliteration, ex.  *sst* 'daughter'.⁴

Fig.3 : Froman 2002 page xxxv

Cayuga speakers requested that syllable boundaries be included in Cayuga words, and so this has been done in the English-Cayuga side of the dictionary. Syllable boundaries are indicated either by raised dots, the colon (vowel length), or the glottal stop. For example, in the word *et-só'ne:k* you will subtract, the raised dot indicates a syllable boundary between the <t> and the <s>. *Raised dots are not used in the Cayuga spelling system and should be removed when spelling Cayuga words.*

In this use case the middle dot is not a *letter*, but is simply used as a separation mark. It is therefore quite appropriate to represent such middle dots with the existing punctuation mark U+00B7 MIDDLE DOT [Po].

C. Length Marker

A middle dot is also frequently used as a length marker, especially in Amercianist usage. However, as indicated by the following examples, in such cases the middle dot should properly be written as a triangle not as a dot.

Fig.4 : Rogers 2000 page 327

- x: **Length.** In handwriting, this symbol is normally a colon; in careful transcription it is made of two triangles.
- x· **Half length.** In handwriting, this symbol is normally a raised dot; in careful transcription it is made of a triangle.

Fig.5 : Pullum & Ladusaw 1996 page 244

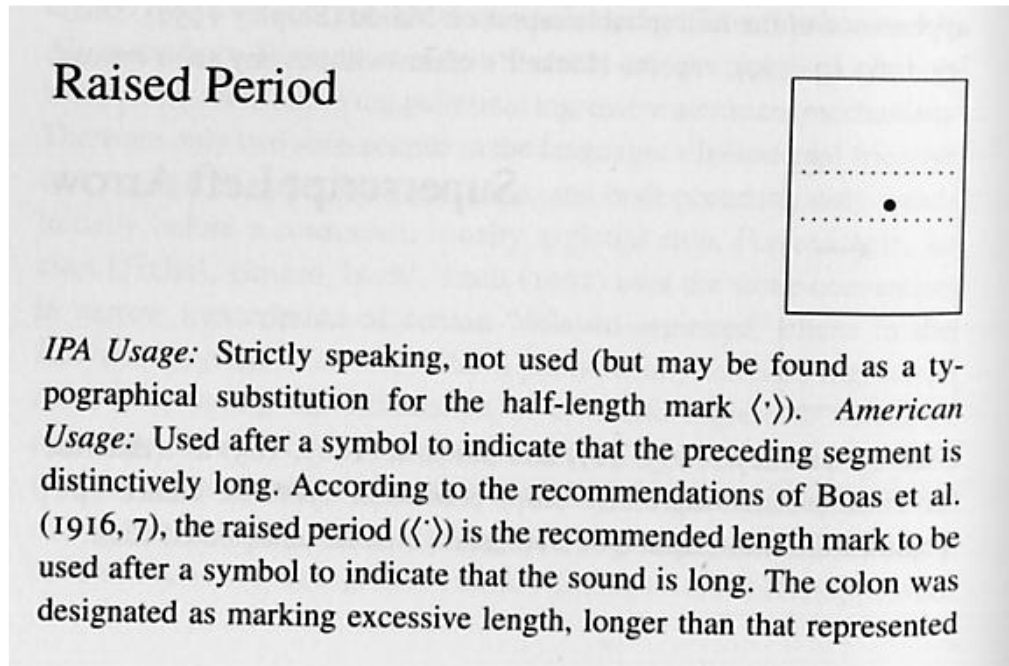
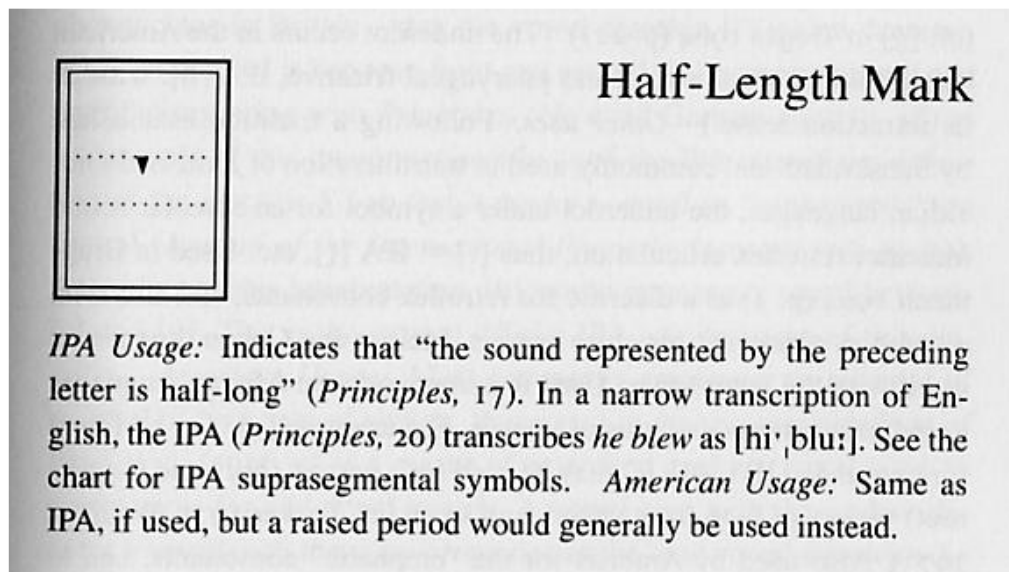


Fig.6 : Pullum & Ladusaw 1996 page 245



In this use case the middle dot is also not a *letter*, but simply a mark indicating vowel length. Written properly it should be represented by means of U+02D1 MODIFIER LETTER HALF TRIANGULAR COLON [Lm], but if the dot form is preferred or required then it would be acceptable to use U+00B7 MIDDLE DOT. The proposed LATIN LETTER MIDDLE DOT is explicitly not intended to represent this usage.

D. Phonetic Letter

The middle dot has been extensively used by scholars of East Asian languages since at least the 1930s to represent the glottal stop in phonetic notations, and in particular for the phonetic reconstructions of ancient and archaic Chinese (see Figs. 7-11), by well-known linguists such as Bernhard Karlgren (1889-1978), Aleksandr Aleksandrovich Dragunov (1900-1955), Li Fang-Kuei (1902-1987) and Edwin G. Pulleyblank (1922-). The middle dot is still used by some contemporary linguists to represent a glottal stop, for example in the reconstructed readings of Tangut characters by Gong Hwang Cherng in Li Fanwen's 2008 Tangut dictionary (see Figs. 12-13). Following on from this tradition, the middle dot is universally used by Phags-pa scholars of all nationalities to transliterate U+A856 PHAGS-PA LETTER SMALL A (which is nominally regarded as representing a glottal stop) (see Figs. 14-20). As can be seen from these examples, the use of a middle dot to represent a glottal stop has been, and continues to be, a widespread tradition amongst Orientalists, and it would be a grave disservice to this user community to pretend that this phonetic letter does not exist.

Fig.7 : Karlgren 1940 page 15

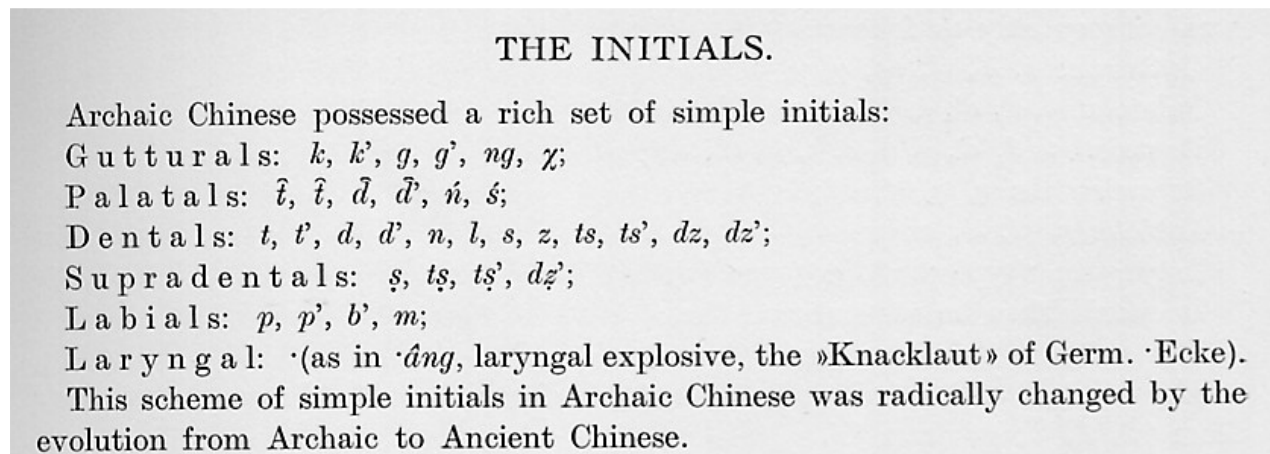


Fig.8 : Karlgren 1940 page 167

- 146 a—c.** **án / ·án / a n* — T peace, peaceful, tranquil (Shī); loan for *id.* how (Shī), where, what (Tso). b. is Yin bone (G 10: 17), c. is Chou I (inscr. 91). The graph has 'roof' and 'woman'.
- d.** **án / ·án / a n* — T repress (Shī); arrange (Ch'uts'i); examine (Li); loan for *id.* thereupon (Sün).
- e.** **án / ·án / a n* — T stool (Chouli); tray (Chouli); loan for *id.* lay hand on, seize (Sün); limit, frontier (Kuoyü); initial particle (Sün).
- f.** **an / ·an / y e n* — T, L, and **án / ·án / a n* — T late (Lunyü); loan for *id.* fine-looking, splendid (Shī); peace, rest (Li), peaceful, mild (Shī).
- g.** **an / ·an / y e n* — T, L name of a bird (Lü).
- h.** **át / ·át / o* — T, L root of the nose (Meng).
- i.** **an / ·an / y e n* — T, L name of a bird (Li, see g. above).
- Of the above, a. **án* 'peace' and f. **an* 'peace' are cognate words.

Fig.9 : Pulleyblank 1962 page 88

The transcription value of initial glottal stop

The glottal stop in Old Chinese was not a mere feature of vocalic ingress, like that of German, but a consonantal phoneme fully integrated into the series of laryngeal initials. That it was recognized as being akin to the velars is shown by its alternation with velar stops in phonetic series, e.g.: 阿 M. ·*a*, 可 M. *kha'*, 哥 M. *ka*, etc.; 謁 M. ·*iat*, 句 M. *kat*, *kai*˘. This being so, it is not surprising that in addition to its normal use to represent foreign vocalic initials, we find it representing foreign velars, or rather, in all probability back-velars or uvulars. (Similar evidence for its use in final position will be given in the continuation of this article.)

Fig.10 : Matisoff 2003 page 552

Li Fang-Kuei

p	t	ts	k	kw	·	·w
ph	th	tsh	kh	khw		
b	d	dz	g	gw		
	s		h	hw		
m	n		ng	ngw		
hm	hn		hng	hngw		
	l, r					
	hl					

The raised dot <·> represents a glottal stop /ʔ/; the digraph <ng> represents a velar nasal /ŋ/, and <w> is not a separate segment but part of a digraph indicating a labio-velar or labio-laryngeal consonant.

Fig.11 : Chao Yuen Ren 1976 page 88

A good symbol should be of a suitable size, in space if it has extension, and in time if it has duration, or in space-time if it has both, such as in gestures. Thus a small, weak red traffic light against a background of neon signs is not as good as a large, strong one. A **raised dot** "·", according to one phonetic usage, as a symbol for the glottal stop, is more likely to be overlooked by the reader or the printer than an undotted question mark "?", according to another usage. In general, the most convenient size for a symbol or complex of symbols is one that fills a substantial part of the visual field or the auditory span. But since we have more often to do with symbol complexes than with symbols in isolation, and moreover, since one of the most important uses of symbols is to examine or discern relations among parts of complexes, it is an advantage to have symbols small enough to be put together in relationship in one grasp.

Fig.12 : Li Fanwen 2008 page 19

聲		母		
p	ph	b	m	
t	th	d	n	
k	kh	g	ŋ	
ts	tsh	dz	s	z
tś	tśh	dź	ś	ź
·	(w)	j	x	y
l	lh	r		

Fig.13 : Li Fanwen 2008 page 1024

0163	𠄎	mji	2.29	I
0494	𠄎	·ji	1.16	VIII
0505	𠄎	la	2.14	IX
1616	𠄎	·o	2.42	VIII
0239	𠄎	lhji	1.30	IX
1598	𠄎	gji	1.14	V
1592	𠄎	lji	2.60	IX

Fig.14 : Poppe 1957 page 19

27	𐑏	h
28	𐑐	.
29	𐑑	y
30	𐑒	u

Fig.15 : Poppe 1957 page 46

Text I

Transliteration

1. [moŋ-k'a] -dêŋ-ri-yin-k'u-č'un-dur-
- 2. q'a-nu-su-dur-
3. [ɣ] oŋ-jhi-ŋan-[s]i-'uŋ-ge-ma-nu-č'é-ri-'u-dun-no-yad-da-č'é-rig-ha-ra-na-ba-la-qa-du[n]-
4. šil-d[e]-dun-da-ru-qas-da-no-yad-da-yor-č'i-qun-ya-bu-qun-Él-č'i-ne-d'ul-qa-qué-
5. bi-č'i[g]-
6. jŋ - [gis] - qa-nu-ba-q'a-nu-ba-ŋar-liq-dur-do-yid-Ér-k'e-ud-sén-šhi-ŋud-daš-mad-
7. č'a[ŋ-t'am-qa-da]-č'a-bu-ši-'li-ba-'l-ba-qub-č'i-ri-'eu-lu-'eu-ŋen-
8. [dêŋ]-ri-yi-ŋal-ba-ri-ju-hi-ru-'er-'eo-gun-'t'u-qayi-g'ek'-deg-sed-'ju-
9. 'uè-É-du-'e-ber-beo-'e-su-U-ri-da-nu-
10. ŋar-li-qun-yo-su-'ar-č'aŋ-t'am-qa-da-č'a-bu-ši-'li-[ba]-'l-ba-qub-č'i-ri-'eu-lu-'eu-

Fig.16 : Junast & Yang 1987 page 7

27	石	晓	·h
28	因	匣(合)	ɣ
29	石	匣	h
30	𠃉	影	.

Fig.17 : Junast & Yang page 150

四 八思巴字字头转写表

原书页码，一页分为上下两面，今以a代表上一面，b代表下一面。

八思巴字字头原误附后圆括弧内，其中误为其它字母者标出相应的字母，不成字者以⊗表示。校正的情况详见下面《校勘记》。

	š ₂ uŋ	zéuŋ	ziŋ
上八^a	š ₁ uŋ	zi g ^e iŋ 二庚	š ₂ iŋ
ji duŋ 一东		giŋ	š ₁ iŋ
guŋ	上九^b	k'iŋ	ɣiŋ
k'uŋ	huŋ	kiŋ	→ 'iŋ
duŋ	ɣuŋ	giŋ	jiŋ
t'uŋ	→ 'uŋ	diŋ	jiŋ

Fig.18 : Murata 1957 page 149

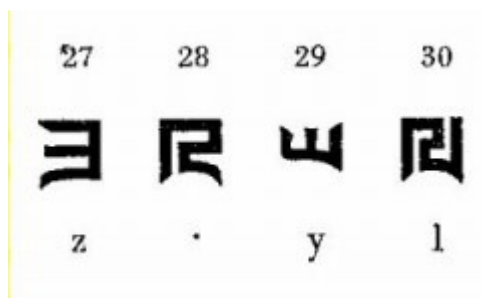


Fig.19 : Murata 1957 page 155

パクバ大字刻文 東壁 佛頂尊勝陀羅尼

Uṣṇīṣavijaya-dhāraṇī written in 'Phags-pa Script, East Wall

[Line 1] ↓ ↓ ↓

'Phags-pa: bh·rum su·a h·a // om̐ na mo bha ga va × // sar va tray lo kya //

Restored Sanskrit: om̐ [1] bhrūṃ svāhā. om̐ [5] namo bhagavate. sarva- trailokya-

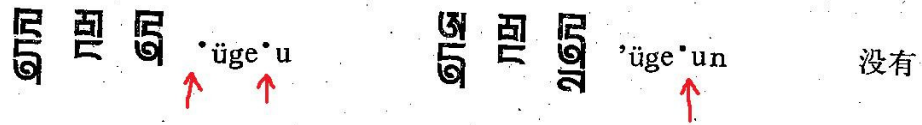
[Line 2]

bra ti × śiṣ ṭ'a ya // bud dh·aya tē na ma × t'·a om̐ bh·rum bh·rum bh·rum śo dha ya

prativīśiṣṭāya buddhāya [10] te namaḥ. tad yathā, om̐ [15] bhrūṃ bhrūṃ bhrūṃ. śodhaya

Fig.20 : Junast 1989 page 31

2.在词首元音前，“·”与零声母“”和字头符（实际上也是一种零声母符号）交替并存，这方面各有一例：



The usage of middle dot in these examples is very different from the other three usages discussed above. The interword separator (interpunct), the intraword separator (intrapunct) and the length mark are all **marks**, the first two used exclusively between letters, and the last one used exclusively to modify a preceding letter. In contrast, the phonetic middle dot is an independent **letter** with a distinct phonetic value that can occur initially, medially or finally in a word just like any other letter. It should be processed as a letter, not as a punctuation mark, which means that it is inappropriate to use U+00B7 MIDDLE DOT for this particular usage. We further note that the middle dot letter is always written as a circular dot, never as a triangular mark, and is thus unrelated to the Americanist usage of middle dot as a substitute for U+02D1 MODIFIER LETTER HALF TRIANGULAR COLON.

We strongly believe that whether or not U+00B7 MIDDLE DOT can be used as a hack substitute for the middle dot letter is irrelevant to the question of whether LATIN LETTER MIDDLE DOT should be encoded or not, and that encoding decisions should be based objectively on the facts of usage, not on the basis of irrational prejudice against encoding "yet another middle dot character". It is our contention that the above examples prove that Orientalists do use the middle dot as a *letter*, and that it should therefore be encoded as a letter for their convenience.

The encoding of such a letter should have no effect on the existing use of U+00B7 MIDDLE DOT to represent the middle dot when it is used as a separation mark or as a length mark. Although it should be obvious from its name that the new LATIN LETTER MIDDLE DOT character is not appropriate for use as a separation mark or a length mark, it should be explicitly stated in the code chart annotations for the new character that it is used by Orientalists to represent a glottal stop, and that the preferred character for other usages is U+00B7 MIDDLE DOT or U+2E31 WORD SEPARATOR MIDDLE DOT.

Finally, we would like to point out that the fact that LATIN LETTER MIDDLE DOT has the same or a similar glyph shape as U+00B7 MIDDLE DOT should not be an impediment to encoding it. There is ample precedent for separately encoding as letters punctuation marks that are used orthographically or in phonetic notation as letters:

Non-Letter Character

U+0021 EXCLAMATION MARK [Po]
 U+0027 APOSTROPHE [Po]
 U+003A COLON [Po]
 U+2018 LEFT SINGLE QUOTATION MARK [Pi]
 U+2019 RIGHT SINGLE QUOTATION MARK [Pf]
 U+201B SINGLE HIGH-REVERSED-9 QUOTATION MARK [Pi]

Corresponding Letter

U+01C3 LATIN LETTER RETROFLEX CLICK [Lo]
 U+A78B LATIN CAPITAL LETTER SALTILLO [Lu]
 U+A78C LATIN SMALL LETTER SALTILLO [Ll]
 U+A789 MODIFIER LETTER COLON [Sk]
 U+02BB MODIFIER LETTER TURNED COMMA [Lm]
 U+02BC MODIFIER LETTER APOSTROPHE [Lm]
 U+02BD MODIFIER LETTER REVERSED COMMA [Lm]

The proposed LATIN LETTER MIDDLE DOT is in this respect no different to the existing LATIN LETTER RETROFLEX CLICK (named LATIN LETTER EXCLAMATION MARK in Unicode 1), and the same rationales used to justify the encoding of this and other punctuation-like letters apply equally to LATIN LETTER MIDDLE DOT.

3. Bibliography

- Chao Yuen Ren 1976.** Chao Yuen Ren, *Aspects of Chinese Sociolinguistics : Essays by Yuen Ren Chao*. Stanford, 1976.
- Froman 2002.** Frances Froman, Alfred Keye, Carrie Joan Dyck and Lottie Keye, *English-Cayuga/Cayuga-English Dictionary*. Toronto. 2002.
- Gardiner 1957.** Alan Gardiner, *Egyptian Grammar* (3rd ed.). Oxford, 1957.
- Humez & Humez 2008.** Alexander Humez and Nicholas Humez, *on the dot : the speck that changed the world*. Oxford, 2008.
- Junast 1989.** Junast 照那斯图, "Bāsībāzì zhōng de líng shēngmǔ fúhào" 八思巴字中的零声母符号; in *Mínzú Yǔwén* 民族语文 1989.2 : 29-36.
- Junast & Yang 1987.** Junast 照那斯图 and Yáng Nǎisī 杨耐思, *Měnggǔ Zìyùn Jiàoběn* 蒙古字韵校本. Beijing, 1987.
- Karlgren 1940.** Bernhard Karlgren, *Grammata Serica*; in *The Bulletin of the Museum of Far Eastern Antiquities* No.12. Stockholm, 1940.
- Li Fanwen 2008.** Lǐ Fànwén 李範文, *Xià-Hàn Zìdiàn* 夏漢字典. Beijing, 2008.
- Matisoff 2003.** James A. Matisoff, *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley & Los Angeles, 2003.
- Murata 1957.** Murata Jiro 村田治郎 (ed.), *Chū-Yung-Kuan* 居庸關 (The Buddhist Arch of the Fourteenth Century A.D. at the Pass of the Great Wall Northwest of Peking). Kyoto, 1957.
- Poppe 1957.** Nicholas Poppe (John R. Krueger trans.), *The Mongolian Monuments in ᠬᠢᠯᠠᠭᠰᠢᠯᠠ Script*. Wiesbaden, 1957.
- Pulleyblank 1962.** Edwin G. Pulleyblank, "The Consonantal System of Old Chinese"; in *Asia Major (New Series)* Vol. IX part 1 pp. 58-144.
- Pullum & Ladusaw 1996.** Geoffrey K. Pullum and William A. Ladusaw, *Phonetic Symbol Guide* (2nd ed.). Chicago & London, 1996.
- Rogers 2000.** Henry Rogers, *The Sounds of Language: an Introduction to Phonetics*. Harlow, 2000.