TYPOLOGY OF NUMERALS: MINORITY LANGUAGES OF CHINA

Francisco A. Marcos-Marín
Catedrático de Lingüística General
Universidad Autónoma de Madrid (ESPAÑA)
Ap. 46348 E-28080 Madrid SPAIN

The Department of Linguistics of the Universidad Autónoma de Madrid (UAM) is conducting a research on the Typology of Numerals. This research embraces 500 languages, pertaining to the possible maximum number of families. A part of this research is included in the framework of the Institute of Studies and Exchange with the Far East, a nonprofit organization connected to several departmental activities, including the regular exchange of scholars and students between the UAM and the University of Foreign Studies at Beijing (P.R. China).

The research on numerals includes the preparation of a typological template, in which several features are considered, such as regularity in numeral formation and classification of peculiari ties according to units, tens and hundreds, presence or absence of the name of the zero (0) inside the formulation of large names, chain structure (fusion, coordination or juxtaposition) and several other parameters. The proposed pattern also offers a basic typological characterization of the language according to the typical features of Subject Verb Object order and the position of the Adjective versus the Substantive. Some diachronical peculiarities regarding the evolution of addition and subtraction have as well been considered.

In this paper the author will expose some results achieved in the domain of languages of China not belonging to the Chinese group of the Sinitic languages. In order to give a broad vision of the scope of the task, three items will be presented: the typological scheme of the languages already studied (i), the presence of an Iranian language with a vigesimal system in Chinese territory (ii), and some observations concerning a characteristic of (at least) Altaic languages: the Oberstufenzählung (iii).

Typological templates

Abbreviations:

* indicates an anomalous circumstance, f.i. in multiplication, column of tens, it indicates that they are formed irregularly (vid. sal-ikur).

A The higher number precedes.
B The lower number precedes.

BA Base of the numeration system. Remains of a former system are indicated by means of a bracketed number in the row above.

AS Adjective Substantive.
C Centena, hundred.
D Decena, decade, ten.
F Fusion.
FA linguistic Family, and branch:IE Indo-European, IEIr Indo-European, Iranian, f.i.
### Language Families

<table>
<thead>
<tr>
<th>Name</th>
<th>Familial Origin</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salikur</td>
<td>IEIr (Indoeuropean, Iranian)</td>
<td></td>
</tr>
<tr>
<td>Wux</td>
<td>IEIr (Indoeuropean, Iranian)</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>STSi (Sino-Tibetan, Sinitic)</td>
<td></td>
</tr>
<tr>
<td>Uyghur</td>
<td>Atu (Altaic, Turic)</td>
<td></td>
</tr>
<tr>
<td>Kazakh</td>
<td>Atu (Altaic, Mongolian)</td>
<td></td>
</tr>
<tr>
<td>Salar</td>
<td>Atu (Altaic, Manchu-tungus)</td>
<td></td>
</tr>
<tr>
<td>Tatar</td>
<td>Atu (Altaic, Manchu-tungus)</td>
<td></td>
</tr>
<tr>
<td>Kirghiz</td>
<td>Atu (Altaic, Manchu-tungus)</td>
<td></td>
</tr>
<tr>
<td>W. Yugu</td>
<td>Atu (Kam-Tai, Tai)</td>
<td></td>
</tr>
<tr>
<td>Mongol</td>
<td>Amo (Kam-Tai, Kam-Sui)</td>
<td></td>
</tr>
<tr>
<td>E. Yugu</td>
<td>Amo (Kam-Tai, Kam-Sui)</td>
<td></td>
</tr>
<tr>
<td>Baanang</td>
<td>Amo (Kam-Tai, Kam-Sui)</td>
<td></td>
</tr>
<tr>
<td>Dagur</td>
<td>Amo (Kam-Tai, Kam-Sui)</td>
<td></td>
</tr>
<tr>
<td>Dong-xiang</td>
<td>Amo (Kam-Tai, Kam-Sui)</td>
<td></td>
</tr>
<tr>
<td>Orogen</td>
<td>Amt (Kam-Tai, Kam-Sui)</td>
<td></td>
</tr>
</tbody>
</table>

1. From 11 on, they are Chinese loan-words.

numhai.wpt
<table>
<thead>
<tr>
<th>Name</th>
<th>Fa</th>
<th>Ba</th>
<th>Subst.</th>
<th>Addition</th>
<th>Multiplication</th>
<th>Complementary Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evenki</td>
<td>Amt</td>
<td>10</td>
<td>0 0 0 0 A 0</td>
<td>---</td>
<td>[B00] [B00]</td>
<td>[MCDU] +</td>
</tr>
<tr>
<td>(Hezhen)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanai</td>
<td>Amt</td>
<td>10</td>
<td>0 0 0 0 A 0</td>
<td>---</td>
<td>[B00] [B00]</td>
<td>[MCDU] +</td>
</tr>
<tr>
<td>(Manchu)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sibo</td>
<td>Amt</td>
<td>10</td>
<td>0 0 0 0 A 0</td>
<td>---</td>
<td>[B00] [B00]</td>
<td>[MCDU] +</td>
</tr>
<tr>
<td>Dai</td>
<td>Ktt</td>
<td>10</td>
<td>0 0 0 0 A 0</td>
<td>---</td>
<td>[B00] [B00]</td>
<td>[MCDU] +</td>
</tr>
<tr>
<td>Zhuang</td>
<td>Ktt</td>
<td>10</td>
<td>0 0 0 0 A 0</td>
<td>---</td>
<td>[B00] [B00]</td>
<td>[MCDU] +</td>
</tr>
<tr>
<td>Bouyei</td>
<td>Ktt</td>
<td>10</td>
<td>0 0 0 0 A 0</td>
<td>---</td>
<td>[B00] [B00]</td>
<td>[MCDU] +</td>
</tr>
<tr>
<td>Maonan</td>
<td>Ktt</td>
<td>10</td>
<td>0 0 0 0 A 0</td>
<td>---</td>
<td>[B00] [B00]</td>
<td>[MCDU] +</td>
</tr>
</tbody>
</table>

**Vigesimal numeration in wux**

An ample variety of languages is spoken in the autonomous region of Xinjiang, in China. Those languages belong, usually, to the Turkic family of the Altaic group. In the Western part of the region, at the border of the Soviet Republic of Tadzhikistan, a Indo-European language is used. It belongs to the Pamir subgroup of the Iranian group, inside the Indo-Iranian family. The wux language is spoken in the Autonomous Province of Tāshikbərgantajjiké and neighbouring zones (Gao: 1985, 1-2). This minority is called Tajik by the Chinese, who relate it to the one that speaks this last language. The Tajik language is spoken, out of China, in the Soviet Tadzhikistán and Uzbekistan, in Afghanistan and Pakistan, but it must not be confounded with the wux. Wux and Salikur, another language of the Iranian family, south-eastern group, are languages different from the Tajik language, which belongs to the South-Western branch of the Iranian family (Comrie: 1981, 161). Another name for the wux is wahan, (Russian vakh, not to be mistaken for a khanyt (ostyak) dialect belonging to the Finno-Ugric branch of the Uralic group called in Russian by this name, Comrie: 1981, 106, sole Uralic language showing a clear ergative construction). The wux presented a regular vigesimal pattern, exposed after the variety called in Chinese wāhān, name of a valley in the North-East of Afghanistan (Gao: 1985, 108-109). This dialect serves as an exchange language in the border of China, Afghanistan, Pakistan and the USSR, at the South of the Pamir plateau. It is fragmented in subdivided (Central, Eastern, Western and Northern). The variant to be described is mainly spoken in the Dābūdā village, inside the P.R. of China.

The first decade is formed by:

---

3 When multiplied by the unit, this follows the hundred or the thousand: thousand one = 1000.

3 See note 2.
1. ji(u)
2. bu(i)
3. tru(i)
4. tstmwir
5. pandz
6. lās
7. mb
8. at
9. nau
10. sas

The digits (11) to (20) are formed by juxtaposition of ‘ten’ and ‘unit’:

11. sas jiu
12. sas bui
13. sas trui
14. sas tstmwir
15. sas pandz
16. sas lās
17. sas mb
18. sas at
19. sas nau
20. bist

From twenty on, decades are formed as multiples of twenty, plus ten, in due case:

20. ji bist (1 x 20)
30. ji bist e(t) sas (1 x 20 + 10)
40. bu bist (2 x 20)
50. bu bist e(t) sas (2 x 20 + 10)
60. tru bist (3 x 20)
70. tru bist e(t) sas (3 x 20 + 10)
80. tstmwir bist (4 x 20)
90. tstmwir bist e(t) sas (4 x 20)
100. pandz bist (5 x 20), and also sad

The combination of the decades, from (20) on, and the units is achieved by vinculation, with the conjunction et:

(21) bist et jiu
(22) bist et bui
(23) bist et trui
(43) bu bist et trui

Note that, due to the nature of the vigesimal system, when the decade is not a pure multiple of twenty, the number is expressed by:

pure multiple + et + ten + unit

(33), so, is “twenty and thirteen”.

This observation is utmost needed for the correct typological interpretation, given the possibility of misinterpreting the result as that of a juxtaposed type; actually, it exists coordination, and accordingly to it the construction belongs to those called by vinculation.

(33) ji bist e(t) sas trui
(58) bu bist e(t) sas at
(95) tstmwir bist e(t) sas pandz

The wux language belongs to the many tongues possessing a scheme for the addition between (10) and (20) different from the one employed from (20) on.

Oberstufenzählung

We might translate this technical term as overstepcounting or, more simply, overcounting. It designates a type of construction not to be confounded with the subtractive one: one number (usually above ten) is named by the corresponding unit and the following decade. The Altaic languages offer very clear examples (Ehlers: 1983, PhTF, JbO):

bir yigirmi, in Old Turkish, 11, is, lit., ‘one twenty’, i.e. ‘the first number belonging to the decade that ends in the twenty’, as it were, instead of ‘one ten’ expected.

äki yigirmi, 12, ‘two twenty’.  

toquz yigirmi, 19, ‘nine twenty’.  

It was believed (Pritsker, *HBO*, 5,1, 1963, p. 40) that this pattern run up to 89, due to the expected ambiguity in the decade from ninety on (Clauson: 1959, 25-26). Clauson himself believes it to be an argument in favour of a system taking us to a period when the Turks did not count above fifty. Ehlers, however, has demonstrated that the system applied also to the tens between 91 and 99, account taken of the following:

92, that will become (overcounting) ékji: yü:t ‘two hundred’, being then mistaken for ‘two hundreds’, 200’, allows no such construction. It is replaced by the unit followed by the word örti, maybe related to the nominal base ör, ‘height, high’, or the verbal ör- ‘heighten’. The resulting ékki örti is, therefore, free from ambiguity. This pattern is valid for 91-99, 191-199, etc.

The system is found in Old Turkish together with the ‘regular’ constructions on toquz (19, ten nine) (Annemarie von Gabain in PhTF, I, 32).

Kaare Thomsen (PhTF, 321) documents overstepcounting in the language of the sary uyurg ("Yellow Uyurg") , pür iyiurma 11 (lit. ‘one twenty’); peşiyärma 15 (lit. ‘five twenty’); pür otus 21 (lit. ‘one thirty’). The last numeral proves how old is otus for ‘thirty’, whose modern form is uçoñ (lit. ‘three ten’). It seems that this language shows overcounting only up to 29.

We have documented overcounting in Western Uyurg, as spoken in the Sunan Uyurg Autonomous County in Gansu Province (Chen Zong Zhen et al.: 1985, 74-75):

<table>
<thead>
<tr>
<th>Numeral</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>birejyärma (1 20)</td>
</tr>
<tr>
<td>12</td>
<td>sigujejyärma (2 20)</td>
</tr>
<tr>
<td>13</td>
<td>yüşüjyärma (3 20)</td>
</tr>
<tr>
<td>14</td>
<td>diorjyärma (4 20)</td>
</tr>
<tr>
<td>15</td>
<td>beşejyärma (5 20)</td>
</tr>
<tr>
<td>16</td>
<td>ahdejyärma (6 20)</td>
</tr>
<tr>
<td>17</td>
<td>jidejyärma (7 20)</td>
</tr>
<tr>
<td>18</td>
<td>sœisejyärma (8 20)</td>
</tr>
<tr>
<td>19</td>
<td>dohisejyärma (9 20)</td>
</tr>
<tr>
<td>20</td>
<td>jiýärma and sometimes sibon</td>
</tr>
</tbody>
</table>

For 21 - 29 both possibilities exist; the old form for 30, ohdís, is used:

<table>
<thead>
<tr>
<th>Numeral</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>birohdís (1 30)</td>
</tr>
<tr>
<td>22</td>
<td>sigohdís (2 30)</td>
</tr>
<tr>
<td>23</td>
<td>hudzohdís (3 30)</td>
</tr>
<tr>
<td>24</td>
<td>diorohdís (4 30)</td>
</tr>
<tr>
<td>25</td>
<td>beşehdís (5 30)</td>
</tr>
<tr>
<td>26</td>
<td>ahldohdís (6 30)</td>
</tr>
<tr>
<td>27</td>
<td>jidohdís (7 30)</td>
</tr>
<tr>
<td>28</td>
<td>sœisohdís (8 30)</td>
</tr>
<tr>
<td>29</td>
<td>dohaisohdís (9 30)</td>
</tr>
<tr>
<td>30</td>
<td>hudžun – hudžon</td>
</tr>
</tbody>
</table>

Sometimes the regularized forms (sibon bir 21 etc.) are used as well.

These data, together with those offered by Pritsker (1955), prove that overstepcounting is a characteristic of the Altaic numerals and help us understand the seemingly complicated evolution of the naming of numbers in those languages.

Overcounting is not to be misunderstood as subtraction. The last construction is also documented in some dialects of (neo)uyurg, for some forms between 20 and 100. O. Pritsker (PhTF, 547) presents it as a continuation of Chagatai Turkish, in Aksu, ikki käm otuz, 28 (lit. ‘two for thirty’), ikki käm qirig, 38 (lit. ‘two for forty’), variants in other dialects for 39 are bir qäm qirg, bi kem q’x (lit. ‘one for fourty’).

Given the huge number of languages studied, it is very convenient to present the partial results of the research to groups of scholars more deeply concentrated in other factors of the languages studied, in order to verify, not the raw data, even if they are not so easily available, but the
connections established between the system of numerals and the general structure of language.

**REFERENCES**

**Note:** This research would not have been accomplished, had I not had access to the Libraries of Stanford University, California, University of California, Berkeley, and of the universities of Heidelberg and Kiel, in Germany. I thank the financial assistance provided by the Comité Conjunto Hispano-Norteamericano and the EUROTRA programme of the E.C. (Stanford, 1987), the Comunidad Autónoma de Madrid (California, Berkeley, summer 1989) and the Alexander von Humboldt Stiftung (Heidelberg, between 1987 and 1991, Kiel, 1991).

**Abbreviations**

**AGI** Archivio Glottologico Italiano.


**HB0** Handbuch der Orientalistik, Leiden/ Colonia.

**IJAL** International Journal of American Linguistics.

**JAOS** Journal of the American Oriental Society, New Haven, Conn.

**JJS** The Journal of Jewish Studies, Londres.

**JNES** Journal of Near Eastern Studies, Chicago.

**JRAS** Journal of the Royal Asiatic Society.

**JSS** Journal of Semitic Studies, Manchester.

**MUSJ** Mélanges de la Faculté Orientale de l’Université Saint-Joseph, Beirut.

**PhTF** Philologiae Turcicae Fundamenta. Cf. Bazin et al., Deny et al.


**SbÖAW** Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien. Phil.-Hist. Klasse.

**U-AJb** Uralo-Altische Jahrbücher.

**WPLU** Working Papers in Language Universals, Stanford.

**WZKM** Wiener Zeitschrift für die Kunde des Morgenlandes.

**ZDMG** Zeitschrift der Deutschen Morgenländischen Gesellschaft.

**ZrPh** Zeitschrift für romanische Philologie.

**TITLES**


22. Pott, August Friedrich 1868 *Die Sprachverschiedenheit in Europa. An den Zahlwörtern nachgewiesen, sowie die quinäre und vigesimale Zählmethode*, Halle: Verlag der


Note: Further references may be found in Marcos Marin: 1989b.