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 non-profit 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 peculiarities 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 Indoeuropean, IEIr Indoeuropean, Iranian, f.i.
<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>Salikur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wux</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(class)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(modern)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uyghur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kazakh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tatar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kirghiz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. Yurgh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mongol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Yurgh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baanang</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dagur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Tu3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monguor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dong-xiang</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orogen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Legend**

- **FEs**: Special Pattern. It indicates which numbers follow the pattern recorded in the line immediately below.
- **FRe**: Regular Pattern. It indicates which numbers follow the pattern recorded in the line immediately below.
- **G**: General.
- **Nu**: Number.
- **P**: Partial.
- **SA**: Substantive Adjective.
- **SIA**: Adjectival Phrase.
- **Su**: Suffix.
- **Subst**: Subtraction.
- **U**: Unit.
- **V**: Vinculation.
- **Y**: Juxtaposition.

**Language Families**

- **IEIr**: Indo-European, Iranian
- **STSi**: Sino-Tibetan, Sinitic
- **Atu**: Altaic, Turkic
- **Amo**: Altaic, Mongolian
- **Amt**: Altaic, Manchu-tungus
- **Ktt**: Kam-Tai, Tai
- **Ktt**: Kam-Tai, Kam-Sui

---

1 From 11 on, they are Chinese loan-words.
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 IndoEuropean language is used. It belongs to the Pamir subgroup of the Iranian group, inside the Indoiranian family. The wux language is spoken in the Autonomous Province of Täshikbërgantajikë 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 Tadzhikistan 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 FinnoUgric 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 wåhån, name of a valley in the North-East of Afghanistan (Gao: 1985, 108109). 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 subdialects (Central, Eastern, Western and Northern). The variant to be described is mainly spoken in the Dábûdá village, inside the P.R. of China.

The first decade is formed by:

\[ 1 \cdot 100 + 1 \cdot 10 + 1 \cdot 1 = 111 \]

\[ 12 \cdot 100 + 2 \cdot 10 + 2 \cdot 1 = 122 \]

\[ 100 \cdot 10 + 0 \cdot 10 + 0 \cdot 1 = 1000 \]

\[ 1000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 10000 \]

\[ 10000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 100000 \]

\[ 100000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 1000000 \]

\[ 1000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 10000000 \]

\[ 10000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 100000000 \]

\[ 100000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 1000000000 \]

\[ 1000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 10000000000 \]

\[ 10000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 100000000000 \]

\[ 100000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 1000000000000 \]

\[ 1000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 10000000000000 \]

\[ 10000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 100000000000000 \]

\[ 100000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 1000000000000000 \]

\[ 1000000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 10000000000000000 \]

\[ 10000000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 100000000000000000 \]

\[ 100000000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 1000000000000000000 \]

\[ 1000000000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 10000000000000000000 \]

\[ 10000000000000000000 \cdot 1 + 0 \cdot 10 + 0 \cdot 1 = 100000000000000000000 \]
<table>
<thead>
<tr>
<th>Number</th>
<th>Uyghur</th>
<th>Average Romanization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>ji(u)</td>
<td>/aš</td>
</tr>
<tr>
<td>2.</td>
<td>bu(i)</td>
<td>ɯb</td>
</tr>
<tr>
<td>3.</td>
<td>tru(i)</td>
<td>ați</td>
</tr>
<tr>
<td>4.</td>
<td>tsẇḃṁr</td>
<td>nau</td>
</tr>
<tr>
<td>5.</td>
<td>pandz</td>
<td>șas</td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td>/aš</td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td>ɯb</td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td>ați</td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td>nau</td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td>șas</td>
</tr>
</tbody>
</table>

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

11. șas jiu  
12. șas bui  
13. șas trui  
14. șas tsẇḃṁr  
15. șas pandz  
16. șas /aš  
17. șas ɯb  
18. șas ați  
19. șas 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) șas (1 x 20 + 10)  
40. bu bist (2 x 20)  
50. bu bist e(t) șas (2 x 20 + 10)  
60. tru bist (3 x 20)  
70. tru bist e(t) șas (3 x 20 + 10)  
80. tsẇḃṁr bist (4 x 20)  
90. tsẇḃṁr bist e(t) șas (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) șas trui  
(58) bu bist e(t) șas ați  
(95) tsẇḃṁr bist e(t) șas 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, PhTT, 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'.

**toquez yigirmi**, 19, 'nine twenty'.