

MAREK STACHOWSKI  
Jagiellonian University, Cracow

## SIBERIAN LANGUAGES IN CONTACT, 1: COLLECTIVE NUMERALS IN YAKUT, DOLGAN, TUVINIAN, TOFALAR AND SOME OTHER TURKIC LANGUAGES\*

**Keywords:** Turkic languages, Siberian languages, comparative linguistics, areal linguistics, linguistic leagues/areas

### Abstract

Morphological categories of Siberian Turkic numerals are particularly complex and therefore deemed to be especially advantageous to areal investigations. The aim of this paper is to see whether (at least some of) the suffixes of collective numerals can readily be used as isogloss connecting Yakut and Dolgan with Tuvinian and Tofalar or, maybe, also some other Turkic languages.

### 1.

One is sometimes told that Tofalar and probably also Tuvinian are the closest relatives of Yakut and, by the same token, Dolgan that in the course of time evolved from Old Yakut as Icelandic did from Old Norse. (cf. Menges 1955: 122, 131; 1958/59: passim; Schönig 1997: 155; 2001: 86). However, precise arguments have never been formulated and discussed. If this conjecture proves correct the possibility of a new evaluation of the Mongolian ethnonym *Uryangkhay* (earlier used by both Yakuts and Tuvinians as a self-designation) and the earliest history of “Tuvaic” (= Tuvinian + Tofalar) and “Sakhaic” (= Yakut + Dolgan) will probably emerge. Indeed, it is intriguing to learn whether Tuvaic and Sakhaic can be connected into one group

---

<sup>1</sup> This paper was first read at the 4<sup>th</sup> Polish Conference on Turkology, Poznań, October 7–8, 2010.

(say, an “Uryangkhaic” one), a fact that would allow us to define at least four languages of the region in a somewhat more precise way than on the basis of geography only (like e.g. “Siberian” ~ “North-East” Turkic).

Because the morphological categories of numerals are particularly numerous in Yakut and Dolgan, and since they constitute a rather complex system (see especially the diagram in Stachowski 1997: 337) it seems the most promising course is to examine to what extent this complicated system reflects a joint evolution of Sakhaic and Tuvaic.<sup>1</sup> Both the complexity of the system and the lack of preliminary studies compel us to investigate individual subsystems before more general conclusions can be drawn.

What now follows is, as far as I am aware, the first study concerning the topic of mutual connections between Tuvaic and Sakhaic. It is, for the time being, impossible to decide whether the choice of the collective numerals for the first study is reasonable. Nevertheless, it is good enough to show that these connections are by no means that simple.

## 2.

The suffix of the collective numerals is *+ya* ~ *+yan* in Yakut, but only *+yan* in Dolgan. The original form of *+ya* is certainly the Old Turkic suffix *+agu* (e.g. in OTkc. *ikegü* ‘both; a pair; zu zweit’, *onagu* ‘a group of ten’ [Erdal 1991: 93; 2004: 225]).

The suffix-final *...n* might be interpreted as another collective suffix *+(a)n*, attested, for instance, in OTkc. *oglan* ‘sons, children’ < *ogul* ‘son’, *eren* ‘men’ < *er* ‘man’ (Erdal 1991: 91), and attached to *\*+agu* formations as some type of intensifier, or the like.

The fact, however, that the derivatives with *...n* usually have the adverbial meaning ‘two at a time, in twos, two by two; zu zweit; вдвоем’ in Yakut (cf. JaND 139) suggests that this *...n* is a reflex of the instrumental suffix *+(u)n* rather (Erdal 2004: 175).<sup>2</sup> Now, the development probably was as follows:

- (2a) Proto-Yakut has a reflex of the suffix *\*+agu*.  
 (2b) An adverbial collective suffix *\*+agu+n* is invented.<sup>3</sup>

<sup>1</sup> The Dolgan system is not fully identical with the Yakut one. However, for historical reasons its peculiar features evolved only in the 17<sup>th</sup> century, so that they are inconsequential for the reconstruction of the earliest phases of Proto-Yakut history (and, possibly, also of the cohabitation of Proto-Yakuts and Proto-Tuvinians that must have come to an end after the last Proto-Yakuts had migrated northwards, i.e. in the 16<sup>th</sup> century at the latest).

<sup>2</sup> For another example of declensional forms becoming a new morphological category of Turkic numerals cf. the Yakut usage of possessive dative or accusative forms with the meaning of iterative numerals, as in: Yak. *ikki* ‘two’ → *ikki+s* ‘second’ → *ikki+s+iger* (dat. 3. sg.) ~ *ikki+s+in* (acc. 3. sg.) ‘for the second time’ (Kotwicz 1930: 208), the etymological intervocalic *...s...* being pronounced and (apart from Pekarskij’s dictionary and some attestations in Sieroszewski’s ethnographical study [Stachowski 1991: 308]) also spelt *...h...*, i.e. *ikkihiger*, *ikkihin*.

<sup>3</sup> But cf. Menges 1959/60: 105: “[...] *-ä(n)* < *-ä-gi(-n)*, [...] wahrscheinlich eine verbale Ableitung auf *-a/-ä* mit dem davon gebildeten nomen verbale auf *-yy/-gi/-yul/-gü* (mit lativisch-instrumentalischem *-n*), die noch um ein Suffix *-la/-lä* erweitert sein kann – vgl. auch die

- (2c) For some period of time, the division is clear: *\*+agu* derivatives are used (ad)nominally, *\*+agu+n* derivatives adverbially.
- (2d) As time goes on, *\*+agu+n* derivatives start being used (ad)nominally.
- (2e) In the Vilyuy dialect of Old Yakut that was the basis of the future Dolgan language *\*+agu+n* formations receive the status of the only collective numerals available, whereas the reflex of the suffix *\*+agu* passes into oblivion.

The lack of old philological sources makes it impossible to decide whether phase (2e) took place still on the Vilyuy (that is, in the first half of the 17<sup>th</sup> century, at the latest) or rather in Taimyr (that is, in the second half of the 17<sup>th</sup> century, at the earliest).

### 3.

The Tuvaic functional counterparts of the Yakut suffix *+ya(n)* are as follows:

- (3a) Tuv. *+älā(n)*, *+aldyrzy*;  
 (3b) Tdž.<sup>4</sup> *+ān*;  
 (3c) Tof. *+ālyn*, *+ān*.

Some facts can be easily observed here:

- (3d) The final *...n* can emerge and disappear in Tuv. *+älā(n)* as is also the case with Yak. *+ya(n)* but the initial parts of these suffixes differ rather considerably.
- (3e) Tuv. *+älā(n)* and Tof. *+ālyn* are probably one and the same suffix.
- (3f) Differences concerning the vowel quantity are not self-evident.
- (3g) An *...l...* can be observed in some variants; its function and origin are unclear.
- (3h) The suffixes without *...l...* (in Yakut, Tuvinian and Todža) probably constitute a separate group.
- (3i) The initial syllable in Tuv. *+aldyrzy* is reminiscent of the initial syllable in Tuv. *+älā(n)* and Tof. *+ālyn* but the subsequent syllables are strikingly different from each other.

The crucial question is now whether the Tuvinian, Todža and Tofalar suffixes are closer to those in Yakut and Dolgan than their counterparts in other Turkic languages are, to the extent that one could even draw an isogloss entitled “Collective numerals” separating the languages forming a “Tuvaic-Sakhaic Language Community”, if there was such a thing, from all the other Turkic languages.

---

mongolischen num. collectiva auf *-yula/-güla*”. – This suggestion seems to have never been accepted in Turkology.

<sup>4</sup> Tdž. = Todža ~ Toža, a north-eastern dialect of Tuvinian (Sat 1997: 384).

## 4.

If a modern morpheme (or a part of one) is a regular reflex of a protolanguage morpheme (or a part of one) it cannot be considered a regional-specific feature (unless it has perished in all the other regions). This is why we have first to reconstruct the situation in the protolanguage, and then to exclude the ordinary heritage in order to find what elements and constructions are regional innovations.

Also the Tuvinian suffix *+aldyrzy* in (3a) can be readily excluded from further considerations here, being a morpheme that does not, on the one hand, reflect the PTkc. *\*+agu*, and, on the other, is restricted to Tuvinian only which means that it can neither throw new light on the Proto-Turkic collective suffix nor create an areal isogloss.<sup>5</sup>

The long vowel in the first syllable of the suffix *+ālā(n)* seems to be original, i.e. *+alā(n)* < *+ālā(n)* because this change occurs more frequently than the opposite (which is easily understandable since the *ā - ā > a - ā* change reduces the articulatory effort) and, besides, this vowel is always long in Tofalar (the position before a syllable with a high vowel will probably have supported the original vowel length since this is a general tendency in the Siberian Turkic languages).

Now, the suffixes in (3a–c) may, apart from *+aldyrzy*, be divided into two groups:

(4a) those with ...*l...*, i.e. Tuv. *+ālā(n)*, Tof. *+ālyn*;

(4b) those without ...*l...*, i.e. Tof. Tdž. *+ān*.

One of the imaginable explanations of this situation is that both groups are genetically connected with each other, and (4b) < (4a). Another possibility is that both groups of suffixes are genetically not connected to each other at all and their phonetic similarity is fully accidental. Yet another solution is that some parts of the suffixes are genetically identical whereas the rest are accidentally similar. Even if this last possibility looks the most complex and least probable it presumably best expresses the morphological and genetic reality.

The explanation of Tdž. and Tuv. *+ān* is relatively easy: < PTkc. *\*+agu* (collective numerals) + *\*+(u)n* (instrumental case), and nothing prevents us from assuming that the suffixes *+ālā(n)* and *+ālyn* possibly include the same instrumental marker, as well.

Three etymologies of the Tuvinian suffix *+ālā(n)* have been suggested so far:

The first possibility is that gerunds in *-yp* were shortened, e.g. PTkc. *\*üč+egü* ‘three at a time’ > *\*üč+egü+len-* ‘to become a group of three persons’ > *\*üč+egü+len-ip* ‘having become a group of three persons’ > Tuv. *üžēlēn* ‘three together, (as) a group of three persons, three at a time’ (F.G. Ischakov, cited after MTof. 122 sq.). – The semantic and

<sup>5</sup> Additionally, its morphological structure is not quite clear. Even if one understands *al...* in *+aldyrzy* as an element other than *āl...* in *+ālā(n)* one will not readily accept the explanation of this suffix as a composition of *\*+al-* (verb formation) + *\*-dyr-* (causative) + *\*+y+zy* (double possessive suffix), as GTuv. 211 sq. puts it, because, firstly, there seems to be no need for reconstructing a *double* possessive suffix, and, secondly, a possessive suffix cannot be attached directly to a verbal stem.

categorial change is well known in the Siberian and the Kipchak Turkic languages.<sup>6</sup> Nevertheless, some objections should be raised here. Apocope of the gerund *-p* can often be observed in Khakass and in Shor; it sometimes occurs in Tuvian, too, but it is unknown from the other Turkic languages. Meanwhile, the forms in ...*lan* are attested also in some other areals, e.g. Old Uzb. *+ala(n)*, Uyg. *+i(j)len*, Kar. and KBalk. *+awlan*, and so on (MTof. 123). Besides, Ischakov's interpretation leaves the moveable ...*n* unexplained.

Ischakov's and Pal'mbach's other etymology connects the final ...*n* with the participle suffix *-gan* (GTuv. 211), i.e. *+ālān* < *\*+agu+lā-gan*. – Here, too, the moveable character of the final ...*n* remains unclear, and their comment that *onālā* is apparently a «truncated form» (“по-видимому [sic!], является усеченной формой”) cannot be called an explanation at all. Nevertheless, the very existence of formations with the compound suffix *\*+agu+lā-gan* seems quite sure in view of attestations like Nog. *onawlagan* ‘ten at a time’ (GTuv. 211). Furthermore, formations like Nog. *onlagan*, Kzk. *ondagan* id. show that the suffix *\*+lā-gan* could have also been attached to cardinal numbers.<sup>7</sup> But then, how did the forms without ...*n* (like Tuv. *onālā* id.) emerge?

A different interpretation again comes from V. I. Rassadin: If the element ...*l*... is an old marker of collectivity,<sup>8</sup> then: *+ālān* < *\*+agu* (coll.) + *\*+l* (coll.) + *\*+(u)n* (instr.) (MTof. 123). – In this case, however, the vowel preceding the *\*+n* causes problems because it is both long and low whereas the vowel of the Turkic instrumental is short and high: *\*+un*. A merging like *\*+lā* + *\*+un* > *\*+lān* is also out of the question because, first, it is usually the final vowel of the preceding syllable that is dropped in the Turkic languages and, besides, no nominal suffix *\*+lā* is known.

In this situation, I am rather inclined to suggest a morphologically somewhat more complex structure, yet one that can explain forms both with and without ...*n*.

In view of Nog. *onlagan*, Kzk. *ondagan* ‘ten at a time’ the existence of the verb *\*ōn+lā-* may be considered certain (although its – transitive? – meaning cannot; ‘to act ten at a time?’). In view of Nog. *onawlagan* id. the verb *\*ōn+agu+lā-*, too,

<sup>6</sup> Cf. Tuv. *orusta-* ‘to speak Russian’ > *orustap* ‘in Russian; по-русски’; Chul. *čany* ‘new’ > *\*čany+la-p* ‘having made anew’ > *\*čanlap* > *čallap* ‘again’ (Pomorska 2004: 150).

<sup>7</sup> Some additional new questions can be posed in this context: If *\*+lā-gan* formations can be accepted as certain do they also make possible the existence of hypothetical *+lan-* verbs from which Ischakov's and Pal'mbach's *\*+lan-yp* gerunds could be derived? What was the semantic difference between: (a) *\*+lā-* and *\*+lan-* derivatives from numerals?; (b) *\*+lā-gan* derivatives from collective and those from cardinal numbers?

Besides, if ...*dyr*... in Tuv. *+aldyrzy* (see (3a) above) really is a causative suffix (GTuv. 211 sq.) [which does not, however, sound very convincing in view of the following syllable ...*zy*, allegedly being a possessive suffix (GTuv. 211 sq.) – attached to a verbal stem?; with ...*z*... after the stem-final consonant?] and the preceding syllable *al*... is the verbal suffix *+al-* (as in *saryg* ‘yellow’ > *sarg+al-* ‘to turn yellow’) the whole construction *\*+al-dyr-* forms transitive verbs which is more or less astonishing from the viewpoint of collective numbers with their meaning ‘being / having become (so and so many persons at a time)’. On the other hand, *+lā-* verbs (whose existence cannot be denied because of Nog. *on(aw)lagan*, see above) also display transitive senses.

<sup>8</sup> For this interpretation of the consonant *l* in *+lar* (plural), *+lyg* (adjectives of possession) and *+l̄yn* (Yakut comitative case) see Kotwicz 1936: 30. The idea was first suggested in Böhtlingk 1848.

should be regarded as sure. Now, the Tuvinian form *onālān* might be accepted as a Tuvinian phonetic reflex of *\*ōn+agu+lā-gan* (> Nog. *onawlagan*), as was suggested by Ischakov and Pal'mbach (GTuv. 211). Then, however, the form *onālā* remains unexplained.

That is why I would rather start with *onālā* which can be explained as an old substantive in *\*-g*, i.e. Tuv. *onālā* < *\*ōn* 'ten' + *\*+agu* (coll.) + *\*+lā-* (verb) + *\*-g* (deverbal substantive). Thus, its morphological meaning was roughly 'a group of those (*\*-g*) who became/acted (*\*+lā-*) all together (*\*+agu*) (in groups of / a group of) ten (*\*ōn*)', i.e. ≈ 'a group of ten acting together'.

For Tuv. *onālān* 'ten at a time' two explanations are possible: either one suggested by Ischakov and Pal'mbach (< *\*ōn+agu+lā-gan*) or one based on instrumental derivation from *onālā*, i.e. *\*ōn+agu+lā-g* (> Tuv. *onālā*) + *\*+un* (instr.) > Tuv. *onālān*.

In none of the cases above can *onālā* be viewed as a «truncated form». Moreover, this derivation avoids the collective element ...l... whose precise meaning and very existence in Proto-Turkic are highly debatable even today.

## 5.

Let us move on now to conclusions concerning areal connections:

- (5a) The basic idea of a morphological category of collective numerals exists in most of the Turkic languages. Consequently, it can be accepted for the Proto-Turkic language, too. Its marker was the suffix *\*+agu* whose direct modern reflexes are: Yak. *+ya*, Oyr. *+ū ~ +u*,<sup>9</sup> Kirg. *+ō*, Tat. *+aw*, Kar. KBalk. Uzb. *+ow*. In other words, the consistency among these data results from their common protolanguage starting point and, for this reason, does not yield any information concerning areal connections and later morphological innovations in individual Turkic languages.
- (5b) Reflexes of the instrumental suffix *\*+(u)n* attached to the collective suffix *\*+agu* can only be found in Yakut and Dolgan (*+yan*) on the one hand, as well as in Todža and Tofalar (*+ān*) on the other. Indeed, this cannot be but a regional innovation, missing everywhere else and, by the same token, suggesting that this combination (*\*+agu+n* (adv.) 'such and such a number of persons at a time') forms an isogloss connecting Yakut and Dolgan with Tofalar and Tuvinian (represented in this context, however, by one dialect only).
- (5c) Rassadin's cautious conjecture that the Kipchak suffix *+ālyn* might have been created under the influence of Uyg. *+(j)lan* (MTof. 123 sq.) – whether correct or not – cannot be automatically extended to Tuvaic and Sakhaic, i.e. the languages that attach the instrumental suffix directly to the nominal *\*+agu* stem

<sup>9</sup> For Oyrrot, only a short vowel suffix *+u* is adduced in GOjr. 84, whereas MTof. 122 has only *+ū*. The fact that the literary Oyrrot adjective suffix is *+lu ~ +lū* (its North Oyrrot dialectal correspondence being *+lyg*, cf. e.g. North Oyr. *tattyg* 'sweet' = liter. Oyr. *tattu ~ tattū* id., GOjr. 25) makes it possible to accept a parallel alternation for the collective suffix.

and therefore have modern reflexes without the consonant ...l... which would point to a verbal derivational base.

- (5d) The idea that the verb based suffixes with ...l... and the noun based ones without ...l... were created everywhere in the Turkic linguistic world in exactly the same period does not sound convincing. Rather, the noun based suffixes without ...l... but with the instrumental  $^{*}(u)n$  (Yakut, Dolgan, Todža, Tofalar) came into being first. It can even be imagined that a reflex of  $^{*}agu+n$  originally covered the whole Tuvinian territory but was, over time, replaced by a newer construction with  $^{*}agu+lā-g(+un)$  that expanded from west to east. What ensued was an asymmetrical distribution: the older form  $+ān$  is now limited to Todža only, that is, to the north-eastern edge of the Tuvinian linguistic community (and preserved in still more northern languages like Tofalar, Yakut and Dolgan) whereas the newer construction  $+ālā(n)$  has dominated all the rest of the Tuvinian territory.
- (5e) As was suggested in (5d) above, the Tuvinian (Todža)-Tofalar-Yakut-Dolgan  $^{*}agu+n$  isogloss is, as a matter of fact, not just an areal innovation, but, rather, an areal preservation of an archaism that arose originally as an areal innovation.

## References

- GOjr. = Dyrenkova N.P. 1940. *Grammatika ojrotskogo jazyka*. Moskva, Leningrad.
- GTuv. = Ischakov F.G., Pal'mbach A.A. 1961. *Grammatika tuvinskogo jazyka. Fonetika i morfoložija*. Moskva.
- JaND = Ubrjatova E.I. 1985. *Jazyk noril'skich dolgan*. Novosibirsk.
- MTof. = Rassadin V.I. 1978. *Morfoložija tofalarskogo jazyka v sravnitel'nom osveščennii*. Moskva.
- Böhtlingk O. 1848. Kritische Bemerkungen zur zweiten Ausgabe von Kasembek's türkisch-tatarischer Grammatik. – *Bulletin de la classe des sciences historiques, philologiques et politiques de l'Académie Impériale des Sciences de St. Pétersbourg* 117V: 312–336.
- Erdal M. 1991. *Old Turkic word formation. A functional approach to the lexicon*. [vol. 1]. Wiesbaden.
- Erdal M. 2004. *A grammar of Old Turkic*. Leiden, Boston.
- Kotwicz W. 1930. Contributions aux études altaïques. – *Rocznik Orientalistyczny* 7: 130–234.
- Kotwicz W. 1936. *Les pronoms dans les langues altaïques*. Kraków.
- Menges K.H. 1955. The South Siberian Turkic languages. I: General characteristics of their phonology. – *Central Asiatic Journal* 1: 107–136.
- Menges K.H. 1958/59. Die türkischen Sprachen Süd-Sibiriens, III: Tuba (Sojon und Karayās), 1. Zur Charakteristik einer einzelnen sibirisch-türkischen Gruppe. – *Central Asiatic Journal* 4: 90–129.
- Menges K.H. 1959/60. Die türkischen Sprachen Süd-Sibiriens, III: Tuba (Sojon und Karayās), 2. Zur Charakteristik einer einzelnen sibirisch-türkischen Gruppe. – *Central Asiatic Journal* 5: 97–150.
- Pomorska M. 2004. *Middle Chulym noun formation*. Kraków.

- Sat Š.Č. 1997. Tuvinskij jazyk. – Tenišev È.R. et al. (eds.) *Jazyki mira – Tjurkskie jazyki*. Moskva: 384–393.
- Schönig C. 1997. Wie entstand das Südsibirische Türkisch? – Kellner-Heinkele B., Zieme P. (eds.) *Studia Ottomanica*. [FS Gy. Hazai]. Wiesbaden: 147–163.
- Schönig C. 2001. Some basic remarks on South Siberian Turkic and its position within North East Turkic and the Turkic language family. – *Dilbilim Araştırmaları* 2001: 63–95.
- Stachowski M. 1991. Über jakutische Glossen im Werk von W. Sieroszewski. – Brendemoen B. (ed.) *Altaica Osloensia*. Oslo: 301–315.
- Stachowski M. 1997. Bemerkungen zu Zahlwörtern sowie Datums- und Altersangaben im Dolganischen und Jakutischen. – Berta Á., Horváth E. (eds.) *Historical and linguistic interaction between Inner-Asia and Europe*. Szeged: 317–339.